

# Dissertation

submitted to the

Combined Faculty of Natural Sciences and Mathematics

of the Ruperto Carola University Heidelberg, Germany

for the degree of

## Doctor of Natural Sciences

Presented by

**M. Sc. Anne-Kathrin Herrmann**

Born in Riesa, Germany

Oral examination: September 7<sup>th</sup> 2018



Dissection and optimization of  
Adeno-associated virus (AAV)  
DNA family shuffling technology:

The journey is the reward

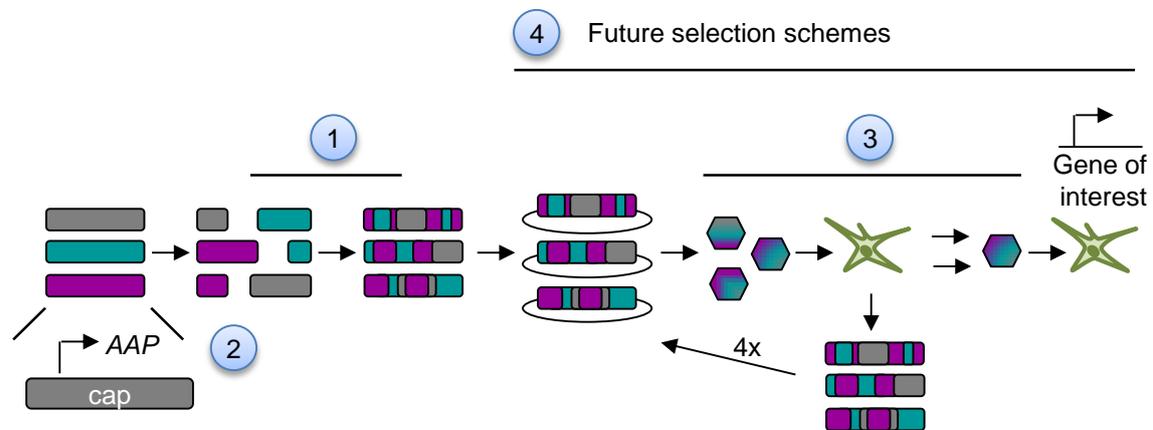
Referees: Prof. Dr. Ralf Bartenschlager  
Prof. Dr. Dirk Grimm



---

 ABSTRACT
 

---

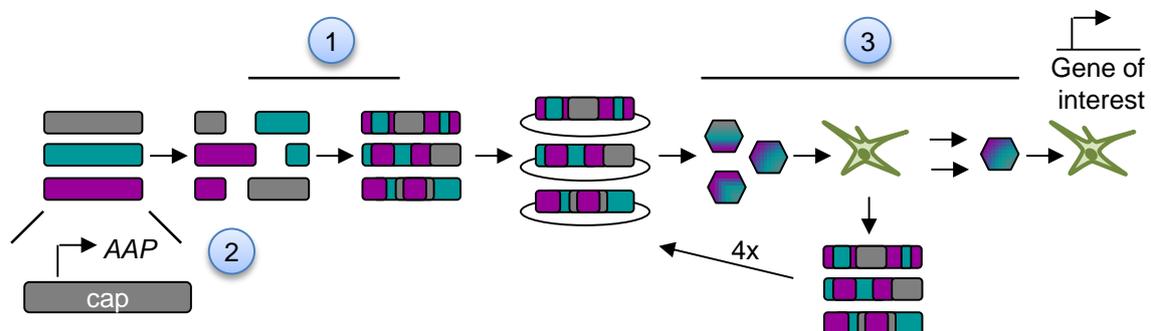


Viral vectors based on Adeno-associated viruses (AAV) have a broad application spectrum including gene therapy and basic research. However, because naturally occurring AAV capsids are rarely sufficiently efficient and/or specific for a given application, techniques were developed to broaden the existing capsid repertoire. A prototype technology is DNA family shuffling where, in a first step, homologous *cap* genes encoding capsid subunits are fragmented and recombined, yielding a viral library which can then be subjected to selection in order to enrich promising variants. The aim of the present study was to dissect and improve four critical steps along this procedure. Firstly (1), two different methodologies for production of *cap* gene fragments were compared, resulting in the identification of DNase I based fragmentation as the most robust approach. Interestingly, *cap* DNA concatamer formation during nested PCR was observed, leading to amendment of the PCR purification protocol. Next (2), we studied the impact of chimerism on the essential assembly-activating protein (AAP) that is encoded in an alternative open reading frame within *cap* and is recombined as well during DNA family shuffling. Importantly, by performing a battery of complementary experiments, we were able to show that shuffling of AAP is not impairing its function, *i.e.* the support of particle assembly. Furthermore, no influence on titers was observed for wild-type and most chimeric vector productions, altogether relieving long-standing concerns about a potential rate-limiting role of AAP for AAV vector generation and evolution. Thirdly (3), we established a pioneering *in vivo* AAV library selection strategy in which, unlike most previously reported schemes, we selected novel capsids in specific cell types within an organ instead of the organ as a whole. Specifically, we were motivated by the facts that liver disease is wide-spread in humans and that hepatic

stellate cells (HSC) are known to drive liver fibrosis, thus contributing to disease progression. Alas, tools to genetically manipulate HSCs are limited. Therefore, a library encompassing 10 capsid variants was selected in HSC by AAV injection into mice, HSC isolation and PCR rescue using purified total DNA. Following multiple selection rounds, *in vivo* bulk validation was performed based on next-generation sequencing. In total, 157 capsid variants were screened in parallel and again, the liver was segregated into the single cell types, *i.e.* hepatocytes, HSC, Kupffer cells and liver sinusoidal endothelial cells. Notably, this revealed that the selection was successful as hepatocyte-detargeted vectors were identified that showed a strong co-transduction of HSC and Kupffer cells. Intriguingly, we noted differences in vector specificity and efficiency on the DNA versus the RNA level. In order to even further restrict the new vectors to a given cell type, vector cassettes were generated bearing cell-type specific promoters and miRNA binding sites to suppress off-targeting in cells expressing these miRNAs. Testing of these constructs *in vitro* gave promising results especially for the miRNA-based detargeting strategy. Finally (4), we implemented improvements during the selection and analysis steps, including the use of PacBio/SMRT sequencing technology to monitor AAV sequence enrichments throughout the course of selection. Additionally, we managed to increase the stringency of the PCR rescue of *cap* genes, by incorporating sample-specific barcodes, *i.e.*, short, unique nucleotide stretches, into the AAV library genomes. By using these barcodes as a primer during sample recovery, we could isolate single libraries out of a complex library mixture, as validated *in vitro*. In the future, this original strategy could be exploited to track individual libraries *in vivo* upon injection of a mixture of libraries, which should in turn help to accelerate the identification of top-performing variants for validation studies. In summary, different steps along the powerful methodology of DNA family shuffling were improved advancing future vector development and the lingering concern about AAP impairment upon shuffling was dispersed.

## ZUSAMMENFASSUNG

### 4 Zukünftige Selektionstrategien



Virale Vektoren, die auf dem Adeno-assoziierten Virus (AAV) basieren, haben ein breites Anwendungsspektrum in der Gentherapie und der Grundlagenforschung. Allerdings sind die natürlich vorkommenden Kapside meist nicht ausreichend effizient und/oder spezifisch für eine bestimmte Anwendung, was die Entwicklung von Technologien gefördert hat, welche das bestehende Repertoire an Kapsiden erweitern können. Ein bekannter Vertreter ist das sogenannte DNA „family shuffling“, bei dem homologe Kapsidgene (*cap*), welche für die Kapsiduntereinheiten kodieren, zunächst fragmentiert und anschließend neu rekombiniert werden. Das Ergebnis sind virale Kapsidbibliotheken, welche für die Anreicherung vielversprechender Varianten selektioniert werden können. Die vorliegende Arbeit adressiert und verbessert vier entscheidende Aspekte des Verfahrens. Zuerst (1) wurden zwei Methoden für die Herstellung von *cap*-Fragmenten betrachtet, wobei DNase I-basierte Fragmentierung als der robustere Ansatz identifiziert wurde. Interessanterweise wurde während der nachfolgenden nested PCR die Ausbildung von DNA-Konkatameren beobachtet, woraufhin eine Verbesserung des Protokolls zur PCR-Aufreinigung erfolgte. Als zweites (2) untersuchten wir den Einfluss von Rekombination auf das essenzielle Assembly-activating protein (AAP), welches in einem alternativen Leserahmen des Kapsidgens kodiert und daher auch von DNA family shuffling betroffen ist. Mit Hilfe einer Serie komplementärer Experimente konnten wir zeigen, dass das „shuffling“ von AAP seine Funktion in der Partikelassemblierung nicht beeinträchtigt. Des Weiteren wurde kein Einfluss auf die Produktionseffizienz von AAV-Wildtypen und den meisten Chimären bestimmt. Diese Ergebnisse mindern bestehende Befürchtungen über eine potenzielle schädliche Wirkung von ungewolltem AAP-Shuffling auf AAV Vektorgeneration und – evolution. Drittens (3) haben wir eine neuartige *in vivo* AAV Selektionsstrategie

entwickelt, bei der neue Kapside in speziellen Zelltypen statt wie bisher im gesamten Organ selektioniert wurden. Wir wurden besonders durch die weite Verbreitung von Lebererkrankungen im Menschen motiviert sowie durch die Rolle bestimmter Zelltypen wie Stellatzellen, die an Leberfibrose beteiligt sind und zum Fortschreiten der Krankheit beitragen. Leider sind die bestehenden Methoden zur genetischen Manipulation dieses Zelltyps aktuell begrenzt. Deshalb wurde eine virale Kapsidbibliothek bestehend aus zehn AAV Serotypen in Stellatzellen in Mäusen selektioniert. Dafür wurde die Bibliothek in die Mäuse injiziert, bevor die Stellatzellen isoliert und die erfolgreichen Kapside nach DNA-Extraktion mittels PCR amplifiziert wurden. Nach mehreren solcher Selektionsrunden wurde für die Validierung ein Hochdurchsatzverfahren basierend auf Next-Generation Sequencing gewählt, in dem 157 Kapsidvarianten parallel in Mäusen getestet wurden. Hierbei wurde der Anteil dieser Kapside sowohl in Gesamt-DNA auch als in Gesamt-cDNA aus Stellatzellen, Hepatozyten, Kupfferzellen und Lebersinusoidendothelzellen bestimmt. Dies bestätigte eine erfolgreiche Selektion hin zu nicht-parenchymalen Zellen, da Kapside identifiziert wurden, die keine Hepatozyten mehr transduzieren. Des Weiteren wurde eine starke Ko-Transduktion von Kupfferzellen und Stellatzellen beobachtet. Interessanterweise haben wir Unterschiede in Effizienz und Spezifität der Kapside auf DNA- und RNA-Ebene festgestellt. Um die Spezifität noch weiter zu erhöhen, wurden Vektorkassetten erstellt mit zelltypspezifischen Promotoren und miRNA-Bindestellen, welche eine Vektorexpression in miRNA-exprimierenden Zellen unterdrücken. Diese generierten Konstrukte wurden einem *in vitro* Test unterzogen, welcher insbesondere die Effizienz der miRNA-basierten Strategie belegte. Zuletzt (4) wurde das Potenzial von PacBio/SMRT-Sequenzierung zur umfassenden Analyse und Verfolgung angereicherter Kapside auf DNA-Ebene verifiziert. Zusätzlich konnten wir durch die Einführung einzigartiger Barcode-Sequenzen in die viralen Kapsidbibliotheken eine stringenter PCR-Isolierung der Kapsidgene ermöglichen. Barcodes sind probenspezifische Nukleotidsequenzen, die bei Verwendung als Bindestelle für Primer die spezifische Isolation einer bestimmten viralen Bibliothek aus einer Mischung erlauben, wie *in vitro* bestätigt wurde. Darüber hinaus wurde diese innovative Strategie eingesetzt, um einzelne Bibliotheken in verschiedenen Organen in der Maus zu verfolgen. Dieser Ansatz könnte in zukünftigen Selektionen die Identifikation der besten Kandidaten für die Validierung beschleunigen. Zusammengefasst wurden verschiedene Schritte bei der potenten Methodik des DNA family shuffling verbessert und gleichzeitig die lang bestehende Befürchtung widerlegt, dass shuffling von AAV zur Funktionsminderung der zugehörigen Kapside führt. Vereint werden die Erkenntnisse und Fortschritte dieser Arbeit die zukünftige AAV Vektorentwicklung vereinfachen und beschleunigen.

---



---

## TABLE OF CONTENTS

---



---

Abstract.....	I
Zusammenfassung.....	III
Table of contents.....	V
List of tables.....	VII
List of figures.....	VIII
List of abbreviations .....	IX
<b>1. Introduction</b> .....	<b>1</b>
1.1 The adeno-associated virus (AAV).....	1
1.1.1. The biology of the virus .....	3
1.1.2. The use of AAVs as recombinant vectors .....	9
1.1.3. Gene therapy .....	10
1.1.4. Capsid modifications .....	12
1.2. The liver.....	17
1.2.1. Organization of the Liver .....	17
1.2.2. Disease development and progression.....	18
1.3. Aims of the thesis .....	19
<b>2. Results</b> .....	<b>21</b>
2.1. Optimization of DNA family shuffling .....	21
2.1.1. Fragmentation methods.....	21
2.1.2. Library amplification and subcloning.....	25
2.2. Inadvertent shuffling of AAP.....	26
2.2.1. Interchangability of AAPs between wild-type AAVs .....	27
2.2.2. Functionality of chimeric AAPs .....	31
2.3. AAV Selection on hepatic stellate cells .....	37
2.3.1. Library generation and selection.....	38
2.3.2. Validation of single chimeras .....	40
2.4. Optimization of the selection process.....	57
2.4.1. Barcode mediated PCR rescue .....	57
2.4.2. Library tracking in vivo.....	60
2.4.3. Monitoring library enrichment by SMRT sequencing.....	62
<b>3. Discussion</b> .....	<b>65</b>
3.1. Technical improvements .....	65
3.2. The shuffling of AAP is not detrimental .....	69

3.3. Selection in stellate cells.....	76
3.4. Overall conclusions .....	82
<b>4. Materials and methods .....</b>	<b>85</b>
4.1. Reagents, buffers and devices used.....	85
4.2. Methods.....	98
4.2.1. Molecular biological Methods .....	98
4.2.2. Prokaryotic methods .....	101
4.2.3. Cloning of new constructs .....	103
4.2.4. Library generation by DNA family shuffling.....	108
4.2.5. Cell culture and virus production .....	111
4.2.6. Virus validation: Transduction assays and capsid formation.....	116
4.2.7. In vivo Selection and validation .....	119
4.2.8. Proof of principle for barcoded libraries .....	122
<b>5. References.....</b>	<b>125</b>
<b>6. Supplemental information .....</b>	<b>135</b>
6.1. Supplemental figures.....	135
6.1.1. Optimization of DNA family shuffling .....	135
6.1.2. Inadvertent shuffling of AAP .....	136
6.1.3. AAV selection on stellate cells.....	138
6.1.4. Optimization of the selection process .....	141
6.2. Excel Macros applied.....	142
6.2.1. Generation of colored type assignments .....	142
6.2.2. Extraction of transduction data .....	143
6.3. Python scripts applied.....	143
6.3.1. Reverse complementing sequencing data.....	143
6.3.2. Barcode extraction from sequencing data .....	144
<b>Acknowledgments .....</b>	<b>147</b>

---

## LIST OF TABLES

---

Table 1-1: Origin of the 13 AAV serotypes .....	2
Table 1-2: Identified receptors and co-receptors for different AAV serotypes .....	5
Table 1-3: Chimeras acquired by DNA family shuffling .....	16
Table 4-1: List of devices .....	85
Table 4-2: List of chemicals, reagents and kits .....	87
Table 4-3: Buffer compositions .....	89
Table 4-4: List of materials used.....	91
Table 4-5: List of softwares applied .....	92
Table 4-6: List of oligonucleotides .....	92
Table 4-7: Plasmids used .....	94
Table 4-8: <i>Escherichia coli</i> strains used .....	101
Table 4-9: Barcode annealing .....	105
Table 4-10: Cloning strategy for cell-type specific constructs .....	107
Table 4-11: Covaris Fragmentation conditions .....	109
Table 4-12: List of cell lines used .....	111
Table 4-13: Transfection mix for one well .....	112
Table 4-14: Transfection mix for one plate .....	113
Table 4-15: Primer/ probe sets used .....	115
Table 4-16: SDS gel for Western Blot.....	118
Table 4-17: List of antibodies used.....	118
Table 6-1: Rescue of different parental AAP knockouts.....	137

## LIST OF FIGURES

Figure 1-1: Genome organization.....	3
Figure 1-2: Replication of the AAV genome .....	7
Figure 1-3: Vectorsystem .....	10
Figure 1-4: Techniques to generate AAV capsid libraries.....	14
Figure 1-5: DNA Family shuffling.....	15
Figure 1-6: Architecture of the liver .....	17
Figure 2-1: Fragmentation methods .....	22
Figure 2-2: Sequence analysis of libraries C1 and D1-3.....	23
Figure 2-3: Improvement of library purification .....	26
Figure 2-4: AAP expression and Capsid assembly monitored on the protein level .....	28
Figure 2-5: Wild-type AAP interchangeability .....	31
Figure 2-6: Rescue ability of AAP DJ .....	32
Figure 2-7: Rescue ability of unselected AAPs.....	34
Figure 2-8: Addition of AAP during rAAV production does not increase transduction efficiency.....	35
Figure 2-9: Contribution of AAPs for capsid 18 production .....	36
Figure 2-10: Generation and selection of a 1-100 library .....	39
Figure 2-11: Amino acid sequence of selected variants .....	41
Figure 2-12: Titers of chimeric capsids.....	42
Figure 2-13: Rescue ability of AAPs which underwent selection .....	43
Figure 2-14: Immunostainings of liver slices.....	44
Figure 2-15: Cell separation and staining.....	46
Figure 2-16: Bulk validation by next-generation sequencing.....	47
Figure 2-17: cDNA profiles for liver cell subtypes .....	48
Figure 2-18: gDNA profiles for liver cell subtypes .....	49
Figure 2-19: Comparison of the cDNA with the gDNA rank .....	51
Figure 2-20: Biodistribution within the liver .....	53
Figure 2-21: Details on the most promising variants.....	54
Figure 2-22: Adaptation of the vector cassette .....	56
Figure 2-23: Barcoded PCR rescue .....	59
Figure 2-24: Library tracking in vivo .....	61
Figure 2-25: Sequence analysis based on different methods .....	63
Figure 3-1: Possible phenotypes during library production .....	76
Figure 3-2: Triple strategy to mediate specificity .....	82
Figure 4-1: ccdb constructs.....	103
Figure 4-2: AAP expression construct design.....	104
Figure 4-3: Constructs for barcode selection.....	105
Figure 4-4: Cell type specific vector design.....	107
Figure 6-1: Salanto analysis libraries C1 and D1-3 .....	135
Figure 6-2: Sequence analysis of the chimeric AAPs .....	136
Figure 6-3: Homology of the capsids.....	138
Figure 6-4: Amino acid sequences sorted by selection round.....	139
Figure 6-5: The variants tested in the NGS screen.....	140
Figure 6-6: gDNA profiles without AAV5.....	141
Figure 6-7: Amplification with the common primer.....	141

---



---

## LIST OF ABBREVIATIONS

---



---

μF	Microfarrand
μg	Microgram
μl	Microliter
μM	Micromolar
AAP	Assembly-activating protein
AAV	Adeno-associated virus
AAVR	AAV receptor
Adh	Adenoviral-helper plasmid
Bp	Base pairs
BR	Basic region
Bs	Binding site
BSA	Bovine serum albumin
C	Celsius
cDNA	Complementary DNA
CMV	Cytomegalovirus
DARPIN	Designed ankyrin repeat protein
ddPCR	Droplet digital PCR
DMEM	Dulbecco's Modified Eagle Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
<i>E. coli</i>	<i>Escherichia coli</i>
<i>e.g.</i>	<i>exempli gratia</i>
EDTA	Ethylenediaminetetraacetic acid
eGFP	Enhanced GFP
eYFP	Enhanced YFP
FCS	Fetal calf serum
For	Forward
G	Gram
G	Glycine
gDNA	Genomic DNA
GFP	Green fluorescent protein
H	Hour
HA	Hemagglutinin
HEK	Human embryonic kidney cells
Hep	Hepatocyte
HSC	Hepatic stellate cells
HSPG	Heparan sulfate proteoglycan
Huh7	Human hepatoma cells
HUVEC	Human Umbilical Vein Endothelial Cells
<i>i.e.</i>	<i>id est</i>
ITR	Inverted Terminal Repeat
Kb	Kilobases
KC	Kupffer cell

kDA	Kilodalton
LSEC	Liver sinusoidal endothelial cells
mA	Milliampere
min	Minute
miRNA	microRNA
MI	Milliliter
mM	Millimolar
mRNA	Messenger RNA
Ng	Nanogram
NGS	Next-generation sequencing
nM	Nanomolar
NPC	Nuclear pore complex
O	Describes shuffling optimized serotypes
ORF	Open reading frame
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PEI	Polyethylenimine
qRT-PCR	Quantitative reverse transcription PCR
R	Arginine
rAAV	Recombinant AAV
Rev	Reverse
rh10	Rhesus 10
RNA	Ribonucleic acid
Rpm	Revolutions per minute
s	Seconds
Salanto	Shuffling-ALignment ANalysis Tool
scAAV	Self-complementary AAV
ssAAV	Single-stranded AAV
TGN	Trans-Golgi network
Trs	Terminal resolution site
UTR	Untranslated region
V	Volt
Vg	Vector genomes
VP	Viral proteins
w/v	Weight per volume
WT	Wild-type
YFP	Yellow fluorescent protein
Ω	Ohm

---

# 1. INTRODUCTION

---

Viral vectors based on Adeno-associated viruses (AAV) are nowadays a widespread tool with applications in gene therapy (section 1.1.3), vaccination strategies (Nieto and Salvetti 2014) and as a tool for genetic manipulation in basic research. However, it remains challenging to restrict genetic manipulation to the target cells of interest. This is especially the case for complex organs such as the liver, which is composed of many different subpopulations, e.g. stellate cells (section 1.2). In order to target AAVs to specific cell types, it is possible to modify the different naturally occurring variants by a variety of strategies (section 1.1.4) such as DNA family shuffling, thereby broadening or restricting the tropism further. This study aimed at the development of improved vectors to target distinct liver cell populations and to broaden our knowledge about aspects of AAV biology that might be affected by the herein applied technology of DNA family shuffling.

---

## 1.1 THE ADENO-ASSOCIATED VIRUS (AAV)

---

The Adeno-associated virus (AAV) was first described and named in 1965 when it was identified as a contaminant of an adenoviral stock (Atchison, Casto, and Hammon 1965). The virus belongs to the family of *Parvoviridae* in the genus of *Dependoparvovirus* (Adams et al. 2014). Its non-enveloped icosahedral capsid with a  $T=1$  symmetry has a diameter of approximately 22-26 nm and is formed by 60 viral protein (VP) subunits (Nonnenmacher and Weber 2012; Goncalves 2005; Madigan and Asokan 2016). They are composed of VP1, VP2 and VP3 in an estimated 1:1:10 ratio (Johnson, Ozer, and Hoggan 1971). AAVs were initially found to be dependent on adenovirus for successful progeny production (Atchison, Casto, and Hammon 1965), but over the years more viruses were identified to support the AAV life cycle including e.g. Herpes simplex virus and human papilloma viruses (Geoffroy and Salvetti 2005). Most research was performed on the prototype serotype 2 (AAV2) which was the variant initially identified. To date, 13 serotypes with various tropisms (Table 1-1) and hundreds of natural isolates were described, which originate from a wide variety of species including mammals, birds and reptiles (Farkas et al. 2004; Flotte and Berns 2005).

Although many AAV sequences have been isolated little is known about the infectious life cycle of wild-type viruses (Berns and Muzyczka 2017). An appropriate *in vivo* model system for studies is not available, therefore scientists are limited to *in vitro* studies. Open questions include the mode of transmission, the infectious dose required or if an infection could occur in the absence of a helper virus. Samples from children were

screened for AAV infection in order to learn more about its spread within humans (Chen et al. 2005). Surprisingly, none of the variants identified beared the HSPG binding motif (section 1.1.1.2). It was thus suggested that AAV2's dependency on HSPG is rather a result from *in vitro* propagation and hence, adaptation to cell culture, than resembling the natural situation (Chen et al. 2005; Srivastava 2016).

TABLE 1-1: ORIGIN OF THE 13 AAV SEROTYPES

Serotype	Species where it was isolated from	Tropism in mice
AAV1	Laboratory isolate (non-human primate) (Hoggan, Blacklow, and Rowe 1966)	Skeletal muscle (Gao et al. 2002), liver (Zincarelli et al. 2008), CNS (lesser extend) (Zhang et al. 2011)
AAV2	Laboratory isolate (human) (Hoggan, Blacklow, and Rowe 1966)	liver (Zincarelli et al. 2008) (lesser extent)
AAV3b	Laboratory isolate (non-human primate) (Rutledge, Halbert, and Russell 1998)	Liver, heart (Zincarelli et al. 2008) (lesser extend)
AAV4	Laboratory isolate (non-human primate) (Parks et al. 1967)	Lung, heart (Zincarelli et al. 2008)
AAV5	Human (Bantel-Schaal and zur Hausen 1984)	Liver (Zincarelli et al. 2008) (lesser extent)
AAV6	Laboratory isolate (Rutledge, Halbert, and Russell 1998)	Liver, heart (Zincarelli et al. 2008), CNS (lesser extent) (Zhang et al. 2011)
AAV7	Non-human primate (Gao et al. 2002)	Skeletal muscle (Gao et al. 2002), liver, heart (Zincarelli et al. 2008), CNS (Zhang et al. 2011)
AAV8	Non-human primate (Gao et al. 2002)	Liver, heart (Gao et al. 2002)
AAV9	Human (Gao et al. 2004)	ubiquitous (Zincarelli et al. 2008), CNS (very efficient) (Zhang et al. 2011)
AAVrh10	Non-human primate(Gao et al. 2003)	CNS (very efficient) (Zhang et al. 2011)
AAV11	Non-human primate (Mori et al. 2004)	Muscle, kidney, spleen, lung, heart, and stomach (Mori et al. 2004)
AAV12	Laboratory isolate (non-human primate) (Schmidt, Voutetakis, et al. 2008)	Muscle (intramuscular injection), salivary glands (injected directly) (Schmidt, Voutetakis, et al. 2008)
AAV13	Laboratory isolate (non-human primate) (Schmidt, Govindasamy, et al. 2008)	Unknown

## 1.1.1. THE BIOLOGY OF THE VIRUS

### 1.1.1.1. GENOME ORGANIZATION

AAVs package a single-stranded DNA genome with a size of 4.7 kb (Figure 1-1) (Srivastava, Lusby, and Berns 1983). The genome is flanked by inverted terminal repeats (ITR) which have a length of 145 bp each of which 125 bp are palindromic and form a hairpin structure (Lusby, Fife, and Berns 1980). The remaining single-stranded 20 bp are designated as the D-sequence (Flotte and Berns 2005). The terminal resolution site (TRS) plays an important role during replication and is situated at the junction of the hairpin and the D-sequence. The first open reading frame (ORF) *rep* encodes four transcripts resulting in the four nonstructural proteins. The longest version, Rep78, is transcribed from the p5 promoter, and the second transcript from this promoter results in Rep68 upon alternative splicing.

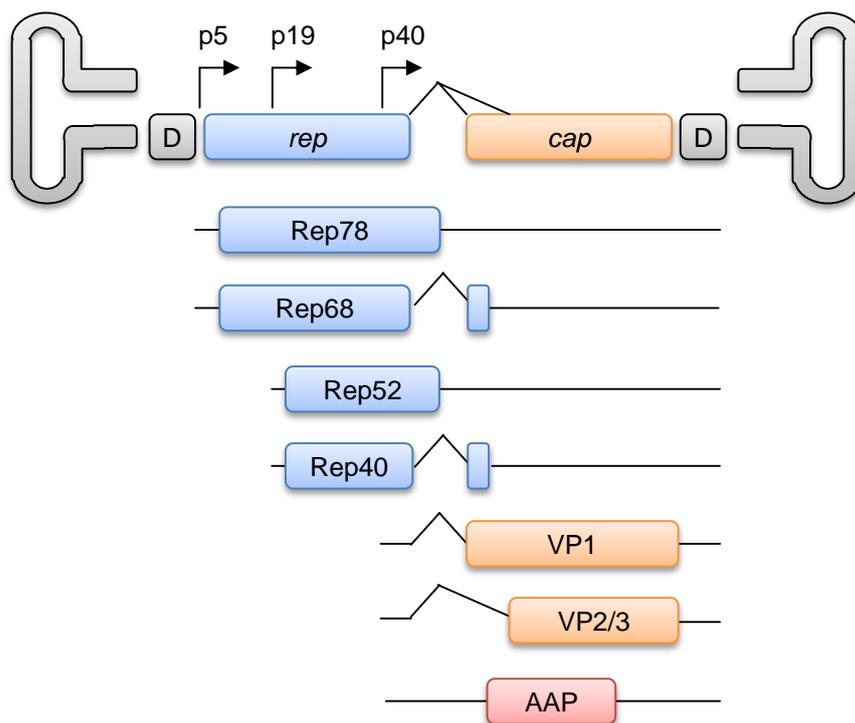


FIGURE 1-1: GENOME ORGANIZATION

This schematic depicts the genome organization of AAV2. The double-stranded loop at both ends and the single-stranded D-sequence belong to the ITR. The genome contains three promoters, namely p5, p19 and p40, and two major ORFs, *rep* and *cap*. *rep* yields the non-structural proteins Rep78, Rep68, Rep52 and Rep40. Two of these transcripts are directly generated, while the other two are the result of splicing as indicated by the diagonal lines. From *cap* two transcripts are produced by using alternative splice acceptor sites, translating into VP1, VP2 and VP3. Within *cap* there is an alternative ORF encoding AAP.

The two remaining transcripts are transcribed from the p19 promoter that is situated within *rep*. Again, there is an unspliced transcript encoding Rep52 and a shorter, spliced, version encoding Rep40. The VPs are transcribed from the second ORF, *cap*, driven by the p40 promoter. Alternative splicing of the p40 transcript, using the same 5' splice donor sites but different 3' splice acceptor sites, yields two mRNAs (Trempe and Carter 1988; Becerra et al. 1988). The mRNA encoding VP1 is generated using a weaker acceptor site resulting in fewer transcripts. The second mRNA is derived from the use of a strong acceptor site and is therefore present in higher numbers (Flotte and Berns 2005). This mRNA encodes the remaining two VPs, which are translated using different start codons. VP2 uses the alternative start codon ACG whereas VP3 translation starts with ATG (Becerra et al. 1985). In 2010, an alternative ORF was identified within *cap* encoding the assembly-activating protein (AAP) which was found to be essential in many variants for capsid formation (section 1.1.1.5) (Sonntag, Schmidt, and Kleinschmidt 2010).

### 1.1.1.2. INITIAL ATTACHMENT AND INTERNALIZATION

---

As for many viruses, initial attachment to the cell is the first major step for AAV infection. Usually this is achieved by a local increase of virus concentration at the cell surface by attachment to mostly negatively charged glycans or glycoconjugates (Huang, Halder, and Agbandje-McKenna 2014). For example, heparan sulfate proteoglycans (HSPG) were identified as a glycan moiety for AAV2 (Summerford and Samulski 1998). HSPGs are widely expressed which might explain the broad tropism observed for AAV2. Over the years, a number of additional co-receptors have been identified for AAV2 such as human fibroblast growth factor receptor 1 (FGFR1),  $\alpha V\beta 5/\alpha 5\beta 1$  integrin, hepatocyte growth factor receptor (HGFR), and laminin receptor (LamR) (Srivastava 2016). Additionally a variety of glycans and co-receptors have been identified for some of the other serotypes (Table 1-2) but there are still unidentified interaction partners. For instance, the glycan usage for serotypes 7, 8, rh10, 11 and 12 is unknown (Huang, Halder, and Agbandje-McKenna 2014). Recently, a more universal receptor named AAV receptor (AAVR) has been described (Pillay et al. 2016). Curiously, the first evidence for this receptor was found in 1996 when AAV2 particles were incubated with membrane fractions identifying a 150 kDa large protein where particles bound to (Mizukami, Young, and Brown 1996). Nearly 20 years later, AAVR was rediscovered as an essential receptor for AAV2 as a knockout prevented transduction and was confirmed to be the previously identified 150 kDa large interaction partner (Pillay et al. 2016; Pillay et al. 2017).

TABLE 1-2: IDENTIFIED RECEPTORS AND CO-RECEPTORS FOR DIFFERENT AAV SEROTYPES

Serotype	co-receptor						glycan receptor			"universal" receptor
	FGFR1	integrin	HGFR	PDGFR	EGFR	LamR	HSPG	SIA	GAL	AAVR
AAV1								$\alpha$ 2-3/ $\alpha$ 2-6 N-linked		X
AAV2	X	$\alpha$ V $\beta$ 5/ $\alpha$ 5 $\beta$ 1	X			x	x			X
AAV3	X		X			x	x			X
AAV4								$\alpha$ 2-3 O- linked		independent
AAV5				X				$\alpha$ 2-3 N- linked		X
AAV6					X		x	$\alpha$ 2-3/ $\alpha$ 2-6 N-linked		X
AAV8						x				X
AAV9		Putative				x			x	X
AAV13							X			

AAVR, Adeno-associated virus receptor; EGFR, epidermal growth factor receptor; FGFR1, fibroblast growth factor receptor-1; GAL, galactose (terminal N-linked); HGFR, hepatocyte growth factor receptor; HSPG, heparin sulfate proteoglycans; LamR, Laminin receptor; PDGFR, platelet-derived growth factor receptor; SIA, sialic acids. The table was modified from (Herrmann and Grimm 2018).

This receptor is used by a variety of serotypes and variants, e.g. AAVs 1, 3, 5, 6, 8 and 9. Yet, AAVR is not required for infection with AAV4 and the closely related variant AAVrh32.33, indicating the existence of another entry mechanism (Dudek et al. 2018).

After attachment to the cell, AAVs are internalized by a variety of mechanisms. To date, clathrin-mediated endocytosis as well as clathrin-independent pathways and macropinocytosis were described, directing the particles to either an infectious or non-infectious pathway (Berry and Asokan 2016).

### 1.1.1.3. INTRACELLULAR TRAFFICKING AND NUCLEAR ENTRY

After internalization, the virus particles traffic in vesicles presumably to early endosomes and further towards the *trans*-Golgi network (TGN) (Berry and Asokan 2016). The transfer towards the TGN was found to be essential for successful transgene expression (Nonnenmacher, Cintrat, et al. 2015). Interestingly, the previously identified AAVR receptor localizes intracellularly to the Golgi apparatus which implies a potential role of AAVR for this trafficking pathway (Pillay and Carette 2017). Additionally, a role of autophagy in intracellular trafficking in hepatocytes was discovered as yet another important mechanism (Hosel et al. 2017).

The escape into the cytosol is mediated by the phospholipase A2 (PLA2) domain found at the N-terminus of VP1 (Girod et al. 2002; Stahnke et al. 2011). Usually this domain is buried within the capsid but upon trafficking through the endosomes the N-termini of VP1 and VP2 are exposed (Sonntag et al. 2006). These structural changes of the capsid are probably partially mediated by the action of cellular proteases, as acidification of the compartments alone is required but not sufficient to prepare the capsid for subsequent trafficking steps (Nonnenmacher and Weber 2012). For example, the inhibition of the endosomal cathepsins B and L led to a reduction in transgene expression for AAVs 2 and 8 but not for AAV5 (Akache et al. 2007). The journey of the AAV via the endosomal pathway and thus, subsequent conformational changes in the capsid, is essential for infectivity as virions microinjected into the cytoplasm or nucleus failed to support infection (Sonntag et al. 2006; Nonnenmacher and Weber 2012).

Once the virions are in the cytosol, they gather in the perinuclear region (Bartlett, Wilcher, and Samulski 2000). Curiously, only a subfraction of the particles (approximately 30%) actually enter the nucleus (Xiao et al. 2012). For entry, a potential role of the nuclear pore complex (NPC) was suggested (Kelich et al. 2015; Nicolson and Samulski 2014). Furthermore, the VPs themselves play an important role. Within the capsid, four so-called basic regions (BR) are found, three of which (BR1-3) have been implied in nuclear entry. Mutations in BR1-3 reduce transduction rates and there is evidence that BR1-3 can act as nuclear localization signals (Sonntag et al. 2006; Johnson et al. 2010; Berry and Asokan 2016). Furthermore, an interaction of BR2 and BR3 with Importin-beta, a protein known to shuttle cargo through the NPC, was confirmed in varying extents across different serotypes (Nicolson and Samulski 2014). Finally, upon reaching the nucleus AAV2 virions enter the nucleoli as intact particles (Sonntag et al. 2006; Johnson and Samulski 2009).

### 1.1.1.4. CAPSID DISASSEMBLY AND REPLICATION

---

In order to release the viral genome, AAV2 needs to exit the nucleolus into the nucleoplasm (Johnson and Samulski 2009). The rate-limiting step is now the conversion of a single-stranded genome into double-stranded DNA. Once freed from the virion shell, the minus and plus strand may anneal thereby forming a double-stranded genome and thus, second strand synthesis is circumvented (McCarty 2008). Replication of the genome is initiated at the ITR on the 5' end as it already offers a free hydroxyl group which can be used for transcription initiation by the host replication machinery (Chandler et al. 2013). A second strand will be synthesized including replication of the 3' ITR

(Figure 1-2). In order to replicate the 5' ITR, a nick is generated by the Rep proteins at the terminal resolution site in the parental strand resulting in yet another 3' hydroxyl group which allows replication of the 5' ITR. Finally, the ITRs refold, leading to separation of the two strands, and during this process, the sequence of the ITRs flips. This would be a simplified model and still some aspects remain enigmatic. For instance, it seems that the capsid does not fully disassemble and a role for the capsid proteins themselves for second-strand synthesis and subsequent transcription was implied (Berry and Asokan 2016). Furthermore, the kinetics seem to differ among serotypes and cell type (Nonnenmacher and Weber 2012).

The replication of the genome and subsequent progeny production can only occur in the presence of a helper virus. If such a helper is not available, only a small quantity of the two larger Rep proteins Rep78/68 will be expressed from the p5 promoter (Weitzman and Linden 2011). The residual expression of Rep78/68 allows binding and redirecting the genome to a certain locus in the human genome to mediate integration (Surosky et al. 1997). This locus situated on chromosome 19 is called AAVS1 and is regarded as a safe harbor for genetic engineering (Kotin et al. 1990; Samulski et al. 1991; Buning and Schmidt 2015). Whereas the replication of the genome takes place in the nucleoplasm, the capsids for AAV2 will preform in the nucleolus (Goncalves 2005; Wistuba et al. 1997). Once the capsids translocate into the nucleoplasm, the minus- and plus-stranded genomes are packaged with equal efficiencies (McCarty 2008). Another essential factor for capsid assembly is the recently discovered assembly-activating protein (section 1.1.1.5).

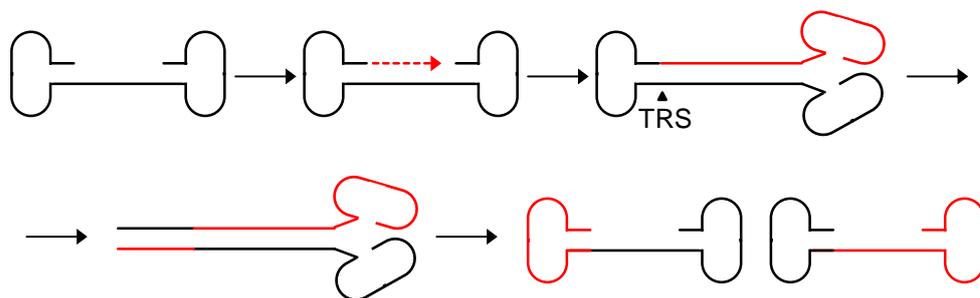


FIGURE 1-2: REPLICATION OF THE AAV GENOME

The schematic depicts the replication of the AAV genome. The free hydroxyl group of the 5' ITR is used to initiate replication (indicated by the dotted line). Replication proceeds through the 3' ITR. A nick is generated within the terminal resolution site (TRS) thereby generating yet another 3' hydroxyl which can be used to copy the 5' ITR. Finally, the strands will displace from each other. Newly synthesized DNA is shown in red.

#### 1.1.1.5. THE ASSEMBLY-ACTIVATING PROTEIN (AAP)

AAP was identified in 2010, nearly 45 years after the initial discovery of AAV (Sonntag, Schmidt, and Kleinschmidt 2010). This study shed light on contradicting findings which were present in the field back then. There were attempts to assemble AAV2 capsids with VP3 only which was pursued by two different strategies. If only VP3 was subcloned into an expression plasmid, no capsids assembled (Ruffing, Zentgraf, and Kleinschmidt 1992). However, if the full-length *cap* gene with mutated VP1 and VP2 start codons was used, capsid formation could be observed (Rabinowitz, Xiao, and Samulski 1999). Subsequently, the study identified a protein in the second ORF whose translation is initiated by a non-conventional CTG start codon (Sonntag, Schmidt, and Kleinschmidt 2010). Furthermore, the authors designed a battery of *cap* mutants and found some of them were defective in capsid assembly. This phenotype could be rescued by *trans*-complementation with this alternative protein. Hence, it was dubbed assembly-activating protein. Soon after, a role of AAP for AAV variants other than AAV2 was confirmed, namely, AAV1, AAV5, AAV8 and AAV9 (Sonntag et al. 2011).

The exact mechanism by which AAP promotes capsid assembly is still not completely understood. An interaction of the N-terminus of AAP with the C-terminus of VP was suggested (Naumer et al. 2012). AAP2 localizes to the nucleolus and a role for transporting VPs to the nucleolei was implied initially (Sonntag, Schmidt, and Kleinschmidt 2010). However, when the nuclear and nucleolar localization signals were identified at the C-terminus of AAV2, the notion was raised that localization is a coincidence (Earley et al. 2015). Bioinformatical analysis of AAP5 revealed a possible absence of nucleolar targeting and led to the hypothesis that AAV5 might be excluded from the nucleolus. Nonetheless, the necessity of nucleolar localization might be at least valid for AAV2. Recently, the localization of AAPs from different serotypes was studied and indeed, they were found in different compartments (Grosse et al. 2017).

In the same study, another assay system was presented to study AAP knockouts. Instead of using constructs solely expressing VP3, the full-length *cap* gene was mutated. The CTG start codon was changed and an early stop codon was introduced (Grosse et al. 2017). Thereby a more natural system was designed to study the role of AAP for capsid assembly. Curiously, the AAP2 mutant still showed low VP level expression which could be stabilized to wild-type levels by inhibition of the proteasome. Although VPs were protected from degradation capsid assembly was not observed. Possibly, AAP protects the VPs from degradation, by fostering capsid assembly, and a role as scaffold or chaperone was suggested.

---

### 1.1.2. THE USE OF AAVS AS RECOMBINANT VECTORS

---

The generation of plasmids containing all viral elements paved the way for the *in vitro* production of AAVs. These initial plasmids contained the ITRs, AAV2 *rep* and *cap* and gave rise to viral progeny in the presence of adenovirus (Laughlin et al. 1983; Samulski et al. 1982) (Figure 1-3 A). Soon after, the ITRs were identified to be the viral element sufficient for genome packaging, and the viral genes could be easily replaced by a gene of interest (McLaughlin et al. 1988; Samulski, Chang, and Shenk 1989). For production of such recombinant AAV (rAAV) particles, the presence of a helper virus and the expression of the AAV genes *in trans* was required. The production system was improved when adenoviral proteins required for AAV replication were identified and subcloned into a plasmid (Grimm et al. 1998; Xiao, Li, and Samulski 1998; Matsushita et al. 1998) (Figure 1-3 B). This strategy required no co-infection with the cytopathic helpervirus and AAV stocks were free of adenoviral contamination.

Within the cell, the recombinant vector genomes will persist episomally as long as the cells do not divide (Duan et al. 1998). One limitation of conventional single-stranded vectors is the requirement of second-strand synthesis for transgene expression (Ferrari et al. 1996; Fisher et al. 1996). This process requires time and is potentially inefficient (Wang et al. 2003). As a solution, self-complementary vectors were developed that were shown to mediate more efficient transduction than their single-stranded counterparts (McCarty, Monahan, and Samulski 2001). This was achieved by introducing a mutation into the terminal resolution site in one of the ITRs. As a consequence, the strands cannot be displaced from each other during vector replication (Wang et al. 2003; McCarty et al. 2003) (Figure 1-3 C). Instead replication proceeds through the mutated ITR until the intact ITR is reached again and resolution can occur. However, self-complementary vectors can only package half the size of wild-type AAV, *i.e.*, approximately 2.2 kb.

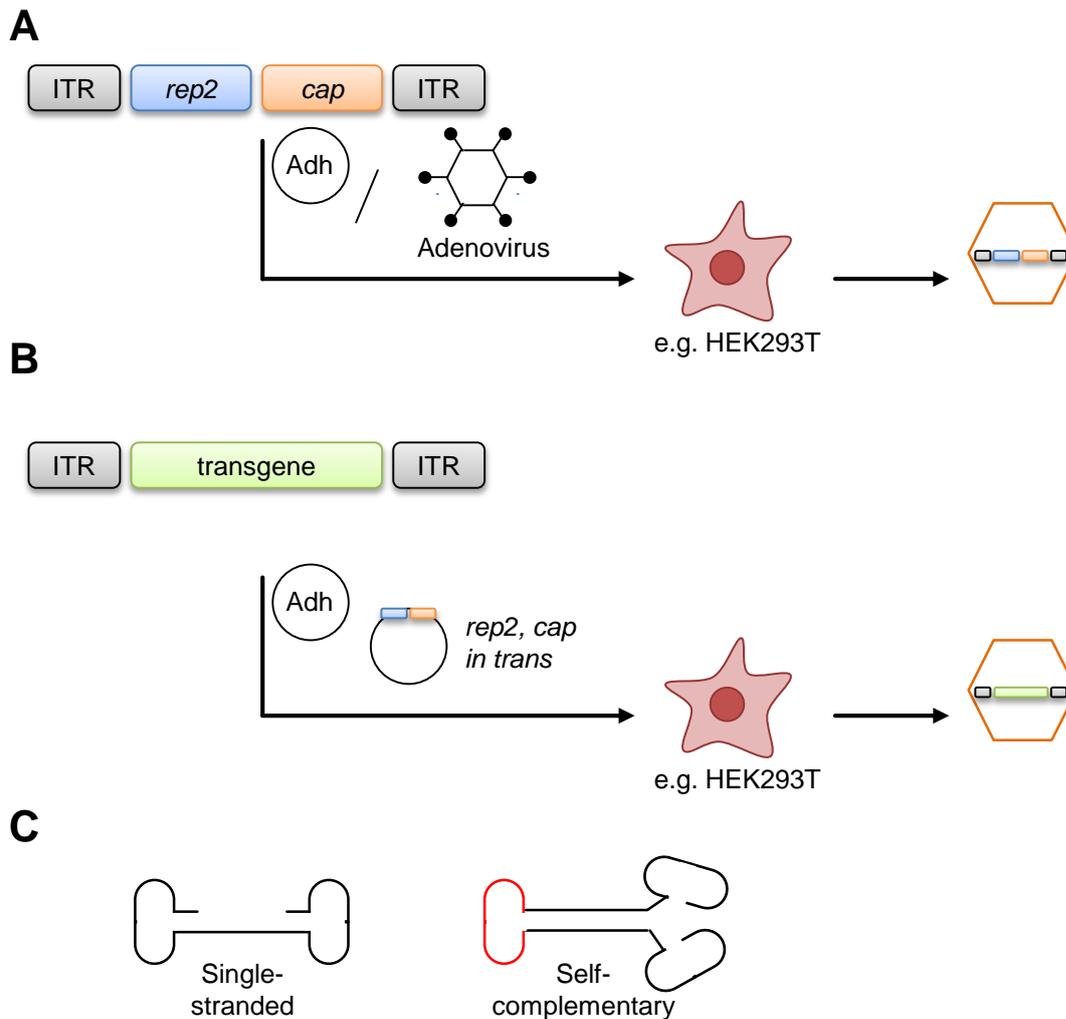


FIGURE 1-3: VECTORSYSTEM

This schematic depicts the two kinds of vector systems. (A) In case of the wild-type virus, the viral genes (*rep* and *cap*) are flanked by inverted terminal repeats (ITRs) that mediate packaging. For successful wild-type progeny production, the functions supplied by a helper are essential. Either they are supplied by the helper virus itself or encoded by a plasmid (Adh). (B) For production of recombinant vectors, the viral genes are replaced by a transgene and transfected into HEK293T cells along with the Adh plasmid and a plasmid encoding *rep* and *cap in trans*. (C) For the generation of rAAVs, two genome configurations are available. Single-stranded vector genomes package cargo resembling the size of the natural genome whereas self-complementary vectors bear a mutated ITR (in red) that prevents resolution during replication. Therefore, the vector genome will be double-stranded and, hence, half of the size can be packaged.

### 1.1.3. GENE THERAPY

If a disease is the result of a genetic disorder, usually only the symptoms can be treated without curing the patient. This constitutes a life-long burden for the patient and the health system. The possibility of targeted manipulation of patient DNA, *i.e.* gene therapy, raised the hope for a curative treatment. This is achieved by introducing nucleic acids into the human cells to correct a mutated gene, introduce a functional copy of the

dysfunctional version or to knock out a malfunctioning gene (Kay 2011). Next to the treatment of inherited diseases, gene therapy is investigated as well to cure cancer. Viruses are the vehicle of choice since they naturally evolved to infect a cell by delivering its genetic cargo.

To date, two gene therapies based on AAV vectors have been already approved in the Western world and many more are in clinical trials. In 2012, Glybera was the first gene therapy approved in Europe. An AAV1-based vector was injected intramuscularly to treat lipoprotein lipase deficiency (Bryant et al. 2013). Very recently, the second AAV based therapy was approved in the US. Luxturna treats inherited retinal disease and is delivered by AAV2 (Ameri 2018).

There are several advantages of the use of AAVs as vehicles for gene transfer in therapeutic applications compared to other viral vectors, such as integrating retro- or lentiviruses. AAV exhibits a low immunogenic profile and is not associated with any human disease (Boutin et al. 2010). On the contrary, it is even believed to have a protective effect against certain diseases (Flotte and Berns 2005; Berns and Muzyczka 2017). For example, women seropositive for AAV were less likely to develop cervix carcinoma caused by human papilloma virus (Smith et al. 2001).

In contrast to e.g. lentiviruses that integrate into the human genome, AAVs mainly persist episomally and thus have a greatly reduced carcinogenic potential (Buchholz, Friedel, and Buning 2015). When integration profiles were analyzed from patients receiving Glybera, a random integration pattern was observed, which most likely occurred during repair of double-strand breaks by non-homologous end-joining (Kaepffel et al. 2013; Miller et al. 2005). However, the integration frequency was much lower compared to vectors derived from integration-competent retroviruses (Kaepffel et al. 2013). The high risk associated with integrating viral vectors has been seen with retroviral treatment of X-linked severe combined immunodeficiency (SCID) when five out of 20 patients developed leukemia (Kay 2011). While the AAV episome persists in non-dividing cells, it will eventually be lost in dividing cells, as the vector genome will not be replicated along and be diluted out. Another advantage is their small capsid size in comparison to lentiviruses which is believed to lead to a deeper tissue penetration (Buchholz, Friedel, and Buning 2015).

One downside of the use of AAV as vectors for gene therapy is the existence of neutralizing antibodies within the human population that might limit the efficacy of the treatment. For example, the seroprevalence for AAV2 ranges from 30% in the US to 60% in Africa (Calcedo et al. 2009). Another study conducted with samples from France

observed the highest prevalence of neutralizing antibodies against AAV2 (59%) followed by AAV1 (50.5%) (Boutin et al. 2010). Surprisingly, neutralizing antibodies against AAV6 were less common (37%) although it shares 96% homology with AAV1. The other serotypes tested showed a low incidence of neutralizing antibodies, namely, AAV5 (3.2%), AAV8 (19%) and AAV9 (33.5%), making them attractive choices for gene therapy.

---

#### 1.1.4. CAPSID MODIFICATIONS

---

Nature has already provided scientists with a great repertoire of naturally occurring variants that infect a variety of different cells. However, these do not always fulfill the requirements needed in a certain situation due to a lack in cell tropism or efficiency. For successful vector transgene expression, delivery is the first important step. A variety of technologies have been developed to alter the capsid, thereby broadening the tropism further. Modification of the capsid also overcomes one drawback of natural isolates, especially for therapeutic approaches: the presence of neutralizing antibodies within the population. There are two main strategies for capsid modification, namely, rational design (section 1.1.4.1) and directed evolution (section 1.1.4.2) of randomized capsid libraries. Recently, a third strategy identified *in silico* novel ancestral AAV variants based on the reconstruction of the evolutionary lineage of present AAV sequences (Zinn et al. 2015).

---

##### 1.1.4.1. RATIONAL DESIGN

---

There are two approaches to rationally redirect the capsids to a given target: chemical modification or genetical alteration. One example for chemical modification is the coupling of receptorspecific antibodies to the capsid after production (Bartlett et al. 1999). However, thermal instability and uncertainties about *in vivo* applications make this strategy less favorable (Daya and Berns 2008). In another scenario, the ligand recognizing the receptor was coupled to streptavidin and the capsids were biotinylated (Ponnazhagan et al. 2002).

In genetic modification strategies, the *cap* gene itself is altered to encode the redirecting element. There are certain peptides known that specifically bind to a given cell type, and the AAV capsid tolerates the insertion of small peptides into the capsid. In an early attempt, the L14 peptide binding to integrin receptors was inserted into the AAV2 capsid at different positions (Girod et al. 1999). All mutants were able to produce particles, and

one of them was successfully retargeted to cell lines which were refractory to AAV2 wild-type infection but bore receptors recognized by L14.

In another study, designed ankyrin repeat proteins (DARPINs) specific for the protein Her2/neu were fused to the N-terminus of VP2 (Munch et al. 2013). Her2/neu is broadly expressed on tumor cells. AAVs were successfully targeted to these cells, and tumor growth in a mouse model was reduced by delivery of a suicide gene. One potential drawback observed was the generation of particles devoid of VP2-DARPIN incorporation resulting in unwanted off-targeting. This could be resolved by adapting the purification process for capsids with incorporated DARPINs (Munch et al. 2015).

Next to the insertion of small peptides and DARPINs, another option to redirect capsids is the specific mutation of amino acids. For example, the basic patch encompassing arginines at positions 484, 487, 585 and 588 and lysine at position 532 is necessary to mediate HSPG binding of AAV2 (Kern et al. 2003; Opie et al. 2003). Residues R484, R487 and K532 are located on one capsid subunit whereas R585 and R588 are present on another. Mutation of R585 and R588 abolishes the HSPG binding capability. *Vice versa*, this loop can also be introduced into non-HSPG-binding variants, such as AAV5, enabling heparin binding (Opie et al. 2003).

Although the strategy of rational design yielded some success stories, it has the inherent drawback that prior knowledge of capsid structures is required which is not always available. Furthermore, in some cases, redirecting peptides act differently in different serotypes, as shown with the peptide 7m8. In the context of AAV2, it was shown to mediate efficient transduction of mouse retina (Dalkara et al. 2013). However, upon engraftment into other serotypes the tropism differed in some cases (Khabou et al. 2016). This suggests that the peptide itself is not the sole mediator of this effect but that the surrounding capsid structure is important, too. Therefore, screening methods based on capsid libraries constitute a feasible alternative.

---

#### 1.1.4.2. DIRECTED EVOLUTION

---

Directed evolution is based on screening of a library containing a large amount of different AAV variants on the cell type of interest and selecting for variants that are able to transduce the target cells efficiently. The first step requires the creation of a plasmid library which then can be used for the preparation of a viral library (Grimm and Zolotukhin 2015). The viruses are then subjected to a selection pressure by *e.g.* infecting a cell of interest or a target organ. The *cap* sequences of the variants that successfully entered the target are recovered and used for subsequent production of secondary

libraries. After iterative selection on the target, single candidates are selected and further characterized. Ideally, they will have gained novel characteristics as compared to the wild-type. There are three widely used strategies to produce the viral libraries: error-prone PCR, insertion of a peptide library and DNA family shuffling. In principle, all of these techniques can be combined (Figure 1-4).

In error-prone PCR, the *cap* gene is mutated during the PCR amplification step thereby creating libraries based on one serotype (Perabo et al. 2006; Maheshri et al. 2006). However, only a few point mutations per gene will be introduced and some of them might be deleterious. Perabo et al. observed that their average rate of 5.7 mutations per clone dropped to 0.9 after production (Perabo et al. 2006). Still, this approach can be powerful as exemplified with the variant EP1.9 which originated from a library generated by error-prone PCR (Asuri et al. 2012). In this variant, a single amino acid (R459G) was mutated in comparison to AAV2. Interestingly, several libraries generated by shuffling, error prone PCR and peptide display were screened in this study in human embryonic stem cells in parallel and EP1.9 was the most efficient capsid variant.

In the case of peptide display libraries, randomized small peptides are inserted into a tolerated position within the capsid and the resulting library is selected on the target of interest (Muller et al. 2003; Perabo et al. 2003). One successful example for selecting a new variant based on peptide display libraries was the already mentioned identification of the peptide 7m8 in the AAV2 context, which transduces retinal cells (Dalkara et al. 2013).

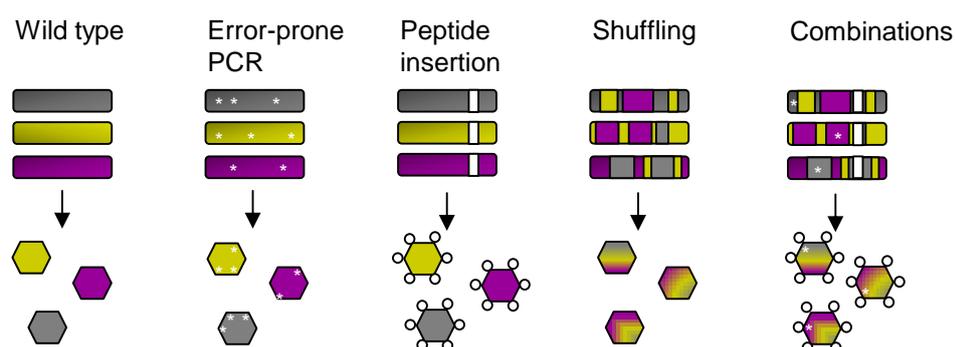


FIGURE 1-4: TECHNIQUES TO GENERATE AAV CAPSID LIBRARIES

The schematic depicts the different methods to generate viral libraries. Error-prone PCR introduces random mutations into the *cap* gene (depicted by asterisks). Additionally, randomized peptides can be integrated into tolerated regions of the capsid. Furthermore, based on homology, DNA family shuffling can yield chimeric capsids. A combination of all strategies can be applied as well.

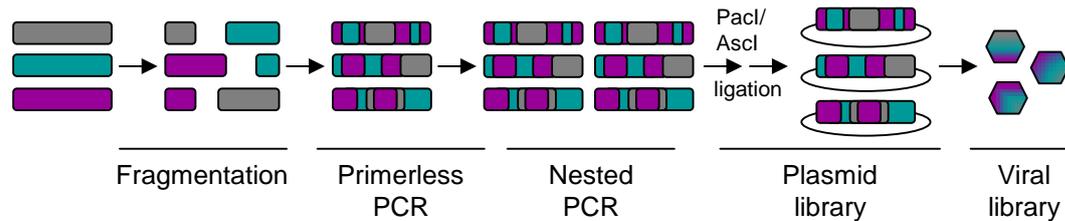


FIGURE 1-5: DNA FAMILY SHUFFLING

The schematic depicts the different steps for the generation of shuffled libraries by DNA family shuffling. Initially, full-length genes are fragmented and used as templates for a first primer-less PCR. Based on homology, fragments will self-prime forming chimeric variants. In a second PCR, the chimeras will be amplified and, in the case of AAV *cap* genes, subcloned into an acceptor plasmid by *PacI/Ascl* digestion. Upon ligation and transformation, a plasmid library is produced that can be used for virus production. The asterisks indicate steps that were dissected experimentally in this work.

The latter approach only yields a diversification at the C-terminal part of the VPs, where the peptide was inserted. However, since N-terminal elements are required for *e.g.* intracellular trafficking, it might be favorable to diversify these positions as well. Towards this aim, DNA family shuffling is an intriguing approach (Stemmer 1994b, 1994a). Here, DNA of closely related proteins can be fragmented and used in a primer-less PCR (Figure 1-5). Due to homology, complementary ends will partially anneal and will be filled up by the polymerase. This yields chimeric sequences consisting of stretches from the parental sequences. DNA family shuffling was introduced into the AAV field in 2008, resulting in the identification of the chimera DJ (Grimm et al. 2008). A library encompassing eight wild-type viruses was subjected to five rounds of selection. In one selection scheme the viral library was incubated with intravenous immunoglobulins (IVIg) prior to infection thereby eliminating immunoprevalent variants. The resultant variant DJ is a recombinant of AAV2, AAV8 and AAV9 and was found to have an immune evasion phenotype compared to its parents. DJ transduces a variety of cell lines *in vitro* but retained liver tropism in mice *in vivo* (Grimm et al. 2008; Liu and Moon 2016). Independently, two more groups published the DNA family shuffling approach soon after (Koerber, Jang, and Schaffer 2008; Li et al. 2008). Since then, a number of groups have performed shuffling on the AAV *cap* gene and have selected the resulting libraries in different targets. In the beginning, most studies focused on cell lines or organs *in vivo* of which a selection can be seen in Table 1-3.

In recent years, selection strategies no longer solely focused on the complete organ but instead attempted to enrich capsids in specific cellular subpopulations. One study made use of a Cre recombinase specifically expressed in astrocytes within the brain. At the

3' end of the *cap* gene, a sequence was inserted that was flipped upon Cre treatment. This reverted sequence was used as primer binding site during the PCR rescue for capsid variants that successfully entered Cre-positive target cells. This led to the identification of the AAV9-based peptide variant PHP.B (Deverman et al. 2016). In another recent example, a comparable strategy was applied to enrich variants for efficient gene transfer in murine neuronal stem cells in the subventricular zone of the brain (Ojala et al. 2018).

TABLE 1-3: CHIMERAS ACQUIRED BY DNA FAMILY SHUFFLING

Chimera	Derived from	Selection on	Details	Reference
DJ	AAV2, AAV8, AAV9	Human hepatoma cell line	Adenovirus coinfection	(Grimm et al. 2008)
LK03	Parental library composed of AAV1-4, AAV6, AAV8-9 (mostly AAV3b)	Human hepatocytes (xenograft mouse model)	Adenovirus coinfection	(Lisowski et al. 2014)
M41	AAV1, AAV6-8	Muscle (mouse)	PCR rescue	(Yang et al. 2009)
B1	Parental library composed of AAV1-2, AAV4-6, AAV8-9, AAVrh.8, AAVrh.10, AAVrh.39, AAVrh.43, no clear assignment possible (mostly AAV8)	Brain (mouse)	PCR rescue	(Choudhury et al. 2016)
HAE-1/ HAE-2	AAV1, AAV6, AAV9 (HAE-2 only)	Human airway epithelium	Adenovirus coinfection	(Li et al. 2009)
AAV2.5T	AAV2, AAV5 in combination with error prone PCR	Human airway epithelium	Adenovirus coinfection	(Excoffon et al. 2009)
Chimeric-1829	AAV1-2, AAV8-9	CS1 (hamster melanoma) cell line	Adenovirus coinfection	(Li et al. 2008)
AAV-U87R7-C5	AAV1-2, AAVrh.8, AAVrh.10 with additional mutations	U87 (human glioma cell line)	Adenovirus coinfection	(Maguire et al. 2010)
32/ 83	Parental library composed of AAV1-6, AAV8 (E531K) and AAV9 (no clear assignment possible at some positions)	Neuron enriched regions	PCR rescue	(Gray et al. 2010)
Olig001	AAV1-2, AAV6, AAV8-9	Striatal cells (rat)	PCR rescue	(Powell et al. 2016)

## 1.2. THE LIVER

The liver is a vital organ that fulfills a variety of functions such as glycogen storage, production of hormones, xenobiotic transformation and protein synthesis (LeCluyse et al. 2012). Additionally, at any given moment it receives approximately 25-30% of the total blood within the human body. Furthermore, immune tolerance is generated there and, due to the abundance of cells with antigen-presenting capability, the liver is important for an immune response (Doherty 2016).

### 1.2.1. ORGANIZATION OF THE LIVER

Hepatocytes are executing most functions associated with the liver and constitute 80% of the total liver mass (Godoy et al. 2013). Whereas the hepatocytes are referred to as parenchymal cells, the remaining cell fraction are the non-parenchymal cells (NPCs). Within the total cell population, 60% are the parenchymal cells and the remaining 40% the NPCs (LeCluyse et al. 2012). The NPCs only constitute 6.5% of the total liver volume and next to Kupffer cells, hepatic stellate cells and liver sinusoidal endothelial cells (LSEC), other cell types are present. For example, cholangiocytes build the biliary ducts and oval cells represent a reservoir of stem cell-like cells.

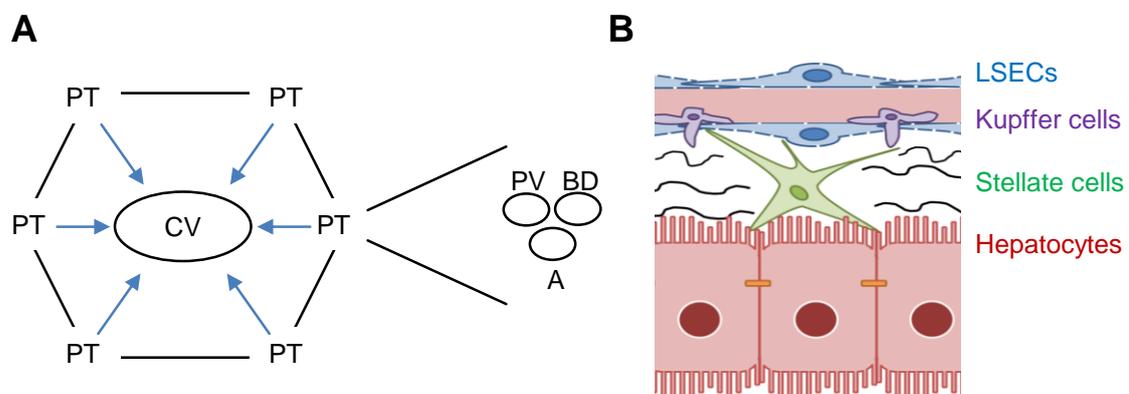


FIGURE 1-6: ARCHITECTURE OF THE LIVER

(A) The hexagonally shaped lobule is the structural unit of the liver and characterized by the central vein (CV) in the middle and the portal triad (PT) at the corners. The PT consists of the portal vein (PV), the bile duct (BD) and a hepatic artery (A). The functional unit is the acinus created by zonation of hepatocyte function due to an oxygen gradient indicated by the blue arrow. (B) The cell types of the liver encompass hepatocytes and the non-parenchymal cell fraction: liver sinusoidal endothelial cells (LSECs), Kupffer cells and stellate cells.

The liver's structural unit is the hexagonally shaped lobule (Godoy et al. 2013) (Figure 1-6 A). The portal triad is situated at the corners of the hexagon and is composed of the portal vein, a hepatic artery and a bile duct. In the center, the central vein can be found where hepatocytes radiate to the periphery of the hexagon. Usually such a cord is one hepatocyte thick, and blood is flowing from the artery and the portal vein along the sinusoids to the central vein (LeCluyse et al. 2012). As a consequence, an oxygen gradient is generated and this describes the functional unit of the lobule namely the acinus.

The major cell types within the NPC fraction are Kupffer cells, stellate cells and LSECs (Figure 1-6 B). The sinusoids are lined by LSECs and constitute approximately 20% of the liver cell population (Kmiec 2001). These endothelial cells have openings in their plasma membrane called fenestrae (Godoy et al. 2013). They allow the diffusion of small molecules such as hormones, proteins and toxicants from the blood to the hepatocytes. Furthermore, they bear scavenger receptors for the uptake of supposed antigens and can act as antigen-presenting cells to T-cells, which might play a role for immunological tolerance.

The space of Disse is located between hepatocytes and LSECs (LeCluyse et al. 2012). Stellate cells can be found there and they constitute around 6% of the liver cell population. In a quiescent state, stellate cells are a storage place for vitamin A and they participate in regeneration. Kupffer cells are liver-resident macrophages situated in the sinusoids where they screen the incoming blood derived from the digestive tract for pathogens (Godoy et al. 2013). They contribute 15% of the liver cell population. Upon recognition, they can launch an immune response by secretion of e.g. cytokines thereby involving the other cells in the liver in the inflammatory response.

---

### 1.2.2. DISEASE DEVELOPMENT AND PROGRESSION

---

Approximately one million people die per year from chronic liver disease caused by e.g. viral hepatitis and alcohol abuse (Fernandez-Iglesias and Gracia-Sancho 2017). Upon liver injury stellate cells can convert into an activated state where they replicate and deposit excess extracellular matrix (LeCluyse et al. 2012). It is thought that this contributes to liver fibrosis and disease progression possibly leading to liver cirrhosis. Usually, pathogens are cleared from the blood but there are some which can successfully establish their infection in the liver, such as the hepatitis viruses and subspecies from the malaria-causing parasite *Plasmodium*. There are two possibilities

how the pathogen reaches the hepatocytes (Protzer, Maini, and Knolle 2012). In one scenario, they leave the blood stream through the fenestrae of LSECs and diffuse through the space of Disse to their target. Alternatively, the pathogens might traverse through other cells such as LSECs and Kupffer cells. Indeed, it was shown for *Plasmodium* that the transfer through Kupffer cells actually increases infection efficiency (Prudencio, Rodriguez, and Mota 2006).

---

### 1.3. AIMS OF THE THESIS

---

Viral vectors based on AAVs have a wide range of applications including gene therapy and basic research. However, most of the naturally occurring variants have the drawback that they are not directly applicable due to e.g. pre-existing immunity or a lack of efficiency and/or specificity. Therefore, methods to broaden the AAV repertoire are both promising and urgently needed to overcome existing hurdles and to foster clinical AAV application. One of these techniques is DNA family shuffling, whereby complex viral capsid libraries are first generated out of a pool of capsid gene fragments and then selected in a given target cell. In this study, we sought to address, improve and standardize different aspects of the AAV DNA shuffling protocol, by optimizing the fragmentation, nested PCR and purification methodologies, altogether aiming to streamline the existing workflow and make it more accessible to interested researchers.

Another important question related to DNA family shuffling concerns the alternative ORF *aap* within the *cap* gene, which is essential for AAV particle assembly. The AAP ORF will be recombined as well during shuffling, raising concerns in the field that this might impair its function. We therefore wanted to study the functionality of chimeric AAPs further and assess whether providing excess AAP during production would overcome possible restrictions.

Additionally, the fate of a viral library was followed up after its generation. One aim here was to develop vectors that are specific for the different liver cell types, i.e. hepatocytes, LSECs, Kupffer cells and stellate cells. As mentioned above, stellate cells are particularly interesting for the research of liver disease but their study is hampered due to limited tools for genetic manipulation. Existing methods include carbon tetrachloride injection into mice to induce liver fibrosis and mouse models where the Cre recombinase is under the control of a stellate cell-specific promoter (Mederacke et al. 2013). As it would be highly beneficial if knockdown or overexpression studies could be conducted, we aimed to fill in this gap by trying to isolate stellate cell-specific AAV vectors. To this end, an AAV library was selected in stellate cells in mice and after four rounds of selection, single

candidates were further validated yielding promising variants. At last, improvements were implemented into the selection and analysis workflow. These include novel PacBio/SMRT sequencing technologies that allow an in-depth analysis of AAV sequence enrichments and an increase in stringency during the PCR rescue of AAV genomes by the usage of barcodes, *i.e.* unique nucleotide stretches.

---

## 2. RESULTS

---

---

### 2.1. OPTIMIZATION OF DNA FAMILY SHUFFLING

---

The generation of viral libraries by DNA family shuffling is an efficient method to diversify AAV capsid genes and a prerequisite for the selection and discovery of AAV variants with novel features, including cellular tropisms or immune evasion. This technique was in the center of the present study, whose aim it was to dissect the biology underlying selected steps in the protocol and to harness the enhanced understanding in order to optimize the library generation and selection workflows. In the following, the results will be described, while technical details are supplied in section 4.2.4.

---

#### 2.1.1. FRAGMENTATION METHODS

---

The initial steps of DNA family shuffling comprise the amplification of genes of interest, *i.e.*, AAV *cap* genes, and their subsequent fragmentation into smaller pieces for ensuing recombination into recombinant chimeric sequences. In this work, two different methods of fragmentation were compared: Covaris-based ultrasonication (physical fragmentation) versus DNase I digestion (enzymatic fragmentation) (Figure 2-1 A). The rationale behind the selection of ultrasonication was that it should allow, in principle, to standardize the fragmentation procedure as it enables the generation of fragments of a defined size range. In fact, the inability to fully control the conventional DNase I fragmentation reaction and the size of the resulting DNA pieces is one of the bottlenecks of the current DNA family shuffling protocol. In a pilot experiment, the Covaris S2 device was set to generate fragments of 150 bp, 300 bp and 800 bp in size. Indeed, when resolved on a 1% agarose gel, different fragment sizes were observed that roughly met the expectations (Figure 2-1 B). This was most pronounced for the 150 bp setting, which was the smallest setting and yielded a distinct band of a size between 100 and 200 bp. The medium setting (300 bp) generated a fragment pool rather than a specific band in the range of 100 to 300 bp. For the largest setting (800 bp) a smear between 200 bp and 800 bp was observed.

Next, all samples (highlighted by boxes in Figure 2-1) were purified from the gel and used for subsequent PCR steps. In all cases, the 800 bp sample produced the full-length *cap* genes (2.2 kb) but with varying efficiencies between the replicates (Figure 2-1 D). The samples with a smaller fragment size failed to yield full-length *cap*, with the

exception of the second replicate (#2) where a recovery of the 2.2 kb *cap* band was observed for the 300 bp input sample.

The second fragmentation method used was DNase I digestion. As seen in Figure 2-1 C, the longer the DNA is digested the smaller the fragments become. For example, digestion for 0.5 min (#2) yielded an even looking smear whereas digestion for 1.5 min (#1) produced a strong signal with a peak at approximately 150 bp. In all cases, the smear ranging from 100 bp to 1 kb was extracted and purified for subsequent PCR steps. All samples were able to recover the full-length *cap* gene to a similar extent (Figure 2-1 E).

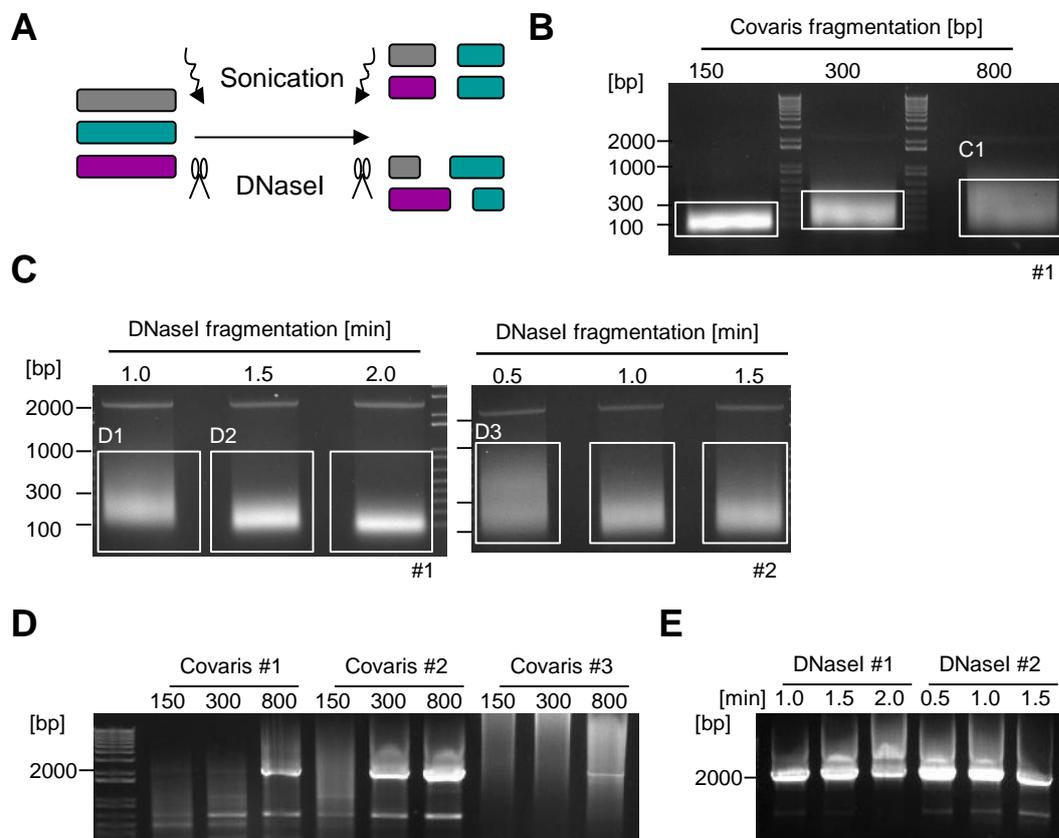


FIGURE 2-1: FRAGMENTATION METHODS

Ultrasonication and DNase I digestion were compared for the generation of capsid libraries. (A) Fragmentation is the initial step for DNA family shuffling. (B) DNA was fragmented by ultrasonication with settings generating fragments of 150 bp, 300 bp and 800 bp and resolved on a 1% agarose gel. A representative image from three independent fragmentations is shown and the boxes highlight the region which was excised and purified. The fragment pool C1 was used for further subcloning. (C) A number of different DNase I digestions were performed and the boxes indicate the regions which were used for excision and subsequent DNA purification. The digestions occurred over a course of 0.5 to 2 min and were resolved on a 1% agarose gel. Fragment pools D1, D2 and D3 were used for further subcloning. (D, E) All samples were used to perform the subsequent PCR steps to reassemble full-length *cap* genes. For comparison, equal volumes were separated on a 1% agarose gel. Adapted from (Herrmann et al. manuscript in preparation).

As shown above, the DNase I digestions produced varying results on the gels (band or smear) indicative of different size distributions. To study whether all products would yield comparable chimeric clones, selected fragment pools highlighted in Figure 2-1 (fragment pools C1, D1-3) were subcloned and single chimeras were sequenced and analyzed by the Shuffling-ALignment ANalysis Tool (Salanto) from our laboratory (Schurmann et al. 2013).

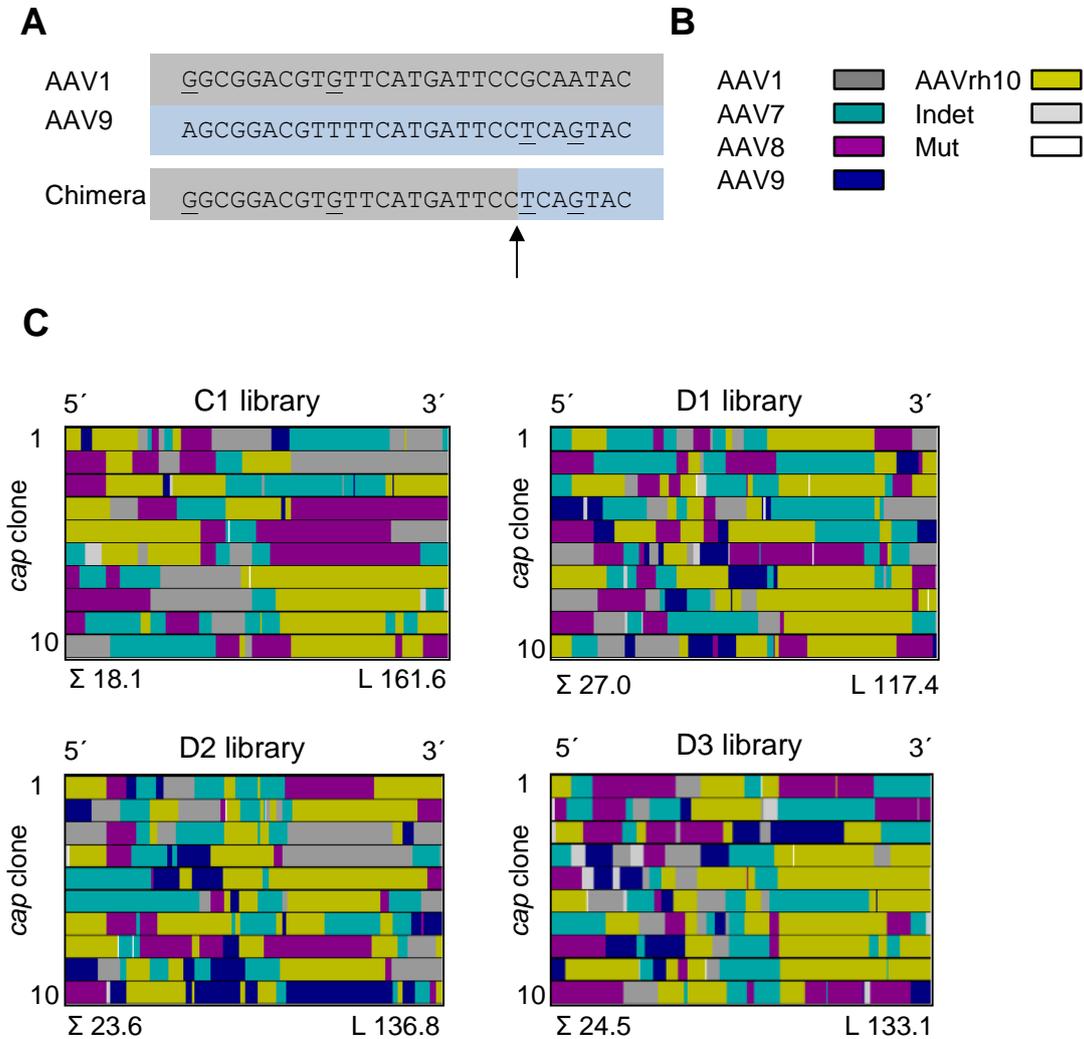


FIGURE 2-2: SEQUENCE ANALYSIS OF LIBRARIES C1 AND D1-3

(A) The in-house tool Salanto compares the sequence of a shuffled clone (chimera) to the parental references (e.g. AAV1 and AAV9). The sequence is assigned to a certain parent until there is a change in sequence. From there on, the next parental reference will be assigned. The arrow indicates the position where a crossover from AAV1 to AAV9 occurs. Nucleotides that differ between the parental serotypes are underlined. (B) A color code is implemented to visualize the parental distribution among the different chimeric clones. Mutated nucleotides are referred to as "mut". In case a sequence stretch is ambiguous, it will be assigned as "indet". (C) Ten clones from each library (C1, D1-3) were analyzed with respect to the distribution of parental serotypes and colored accordingly. One bar represents the complete *cap* gene of a single clone in 5'-3' orientation. Additionally, the average sum ( $\Sigma$ ) of fragments per serotype is depicted below each block and the average fragment length (L) in bp is shown. Adapted from (Herrmann et al. manuscript in preparation).

Of note, one sequencing reaction would not be able to cover the complete 2.2 kb *cap* gene and therefore, three primers were used (#822, #36 and #823) and the complete *cap* sequence was assembled manually. In a first step, all chimeric sequences were aligned to their parental references, *i.e.*, AAV1, AAV7, AAV8, AAV9 and AAVrh10 (Figure 2-2 B). This alignment can then be fed into the Salanto tool. After definition of the parental sequences, the software assigns now the chimeric sequences to their respective parents. The option “type assignment” was chosen for analysis. Here, a nucleotide specific of a given parent is recognized and from there on the chimera is assigned to this reference (Figure 2-2 A). As soon as there is a change in the sequence which is not in accordance with the selected parental reference, a crossover will occur to another parental sequence. This is exemplified by the change in color in Figure 2-2 C. If several options exist for a sequence stretch, it will be assigned as undetermined and labeled “indet”. Furthermore, Salanto recognizes positions where none of the parental references fit as mutations (assigned “mut”).

To compare the libraries generated by the two fragmentation methods, 10 clones each were sequenced and analyzed with Salanto with the abovementioned type assignment. For depiction, the complete nucleotide sequence of the chimeras is depicted and the respective parental serotype was colored according to the legend (Figure 2-2 B, C). Additional features of Salanto were used to assess the library composition. Firstly, the number of fragments derived from each serotype was counted and the mean sum of fragments is depicted. Similarly, the average fragment length was determined. The mean was chosen as the serotypes are not equally distributed among the libraries. For instance, AAV9 represented by dark blue seems to be less abundant in library C1. The analysis for the individual serotypes can be found in the supplementary section in Figure 6-1.

A correlation between the sum of fragments and the length can be observed, namely, the longer the fragment length becomes the number of fragments decreases. Covaris-based fragmentation yielded less (18.1) but larger fragments (161.6 bp). In contrast, libraries based on DNase I digestion, *i.e.*, D1, D2 and D3, yielded more fragments (27, 23.6 and 24.5, respectively) of a shorter average length (117.4 bp, 136.8 bp and 133.1 bp, respectively). All tested DNase I-based libraries gave comparable results despite the differences in the fragment pools resulting from the initial digestions (Figure 2-1 C). This finding in combination with the observation that all samples originating from DNase I-based fragmentation efficiently produced full-length *cap*, highlights the robustness of enzymatic fragmentation for DNA family shuffling.

---

## 2.1.2. LIBRARY AMPLIFICATION AND SUBCLONING

---

Next to the different fragmentation methods, the PCR setup was modified with the aim to improve the resulting DNA yields for more efficient subcloning (Figure 2-3 A). Fragmentation is the initial step for DNA family shuffling and the fragment pools are allowed to re-anneal in a first PCR which is primerless. Based on homology fragments will partially anneal and become filled up by the polymerase. This is then followed by a second nested PCR step which amplifies full-length *cap* genes (section 1.1.4.2, Figure 1-5). Firstly, the elongation step for the second, nested PCR was varied from 1 to 6 min and equal volumes of the PCR reactions were separated on a 1% agarose gel. This showed that an elongation time in the range of 2 to 3 min was optimal to recover full-length *cap*, which can be observed as the band running slightly above the 2 kb marker band (indicated by arrows in Figure 2-3 B). Longer time points (5 to 6 min elongation time) failed to recover *cap* efficiently. Curiously, bands with larger sizes than 2.2 kb occur e.g. between 4 and 5 kb (2-4 min elongation time) and at roughly 6 kb (3-4 min elongation time). One possible explanation is the formation of concatamers during PCR amplification that become more prominent upon an increase in elongation time. The nature of the band is highly intriguing as the standard protocol within the lab for DNA purification after PCR in the context of DNA family shuffling specifically recommends the recovery of the 2.2 kb band from the gel (Große 2016). In order to study the composition of the band between 4 and 5 kb, it was separately excised, purified and digested with *PacI*/*Ascl* restriction enzymes which flank the *cap* genes. If concatamer formation occurred, the restriction sites should be still present and hence, the concatamer should be resolved upon restriction enzyme digestion. Indeed, after *PacI*/*Ascl* digestion the extracted large band with a size between 4 and 5 kb became the same size as the full-length 2.2 kb *cap* band (Figure 2-3 C).

Based on these observations the possibility to improve DNA yields by purification method was investigated. To this end, a large-scale PCR was set up, pooled and half of the volume was loaded on a gel for subsequent excision of the 2.2 kb band and purification. The other half was directly processed for column purification. Both samples were digested according to the protocol (section 4.2.4.4) and were resolved side-by-side on an agarose gel. As seen in Figure 2-1 D the signal intensity is much stronger for the sample that was initially column-purified indicative of a higher yield. We thus conclude that it is beneficial for future library generations to initially purify the large-scale PCR over a column and to then extract the DNA from a gel after digestion with the restriction enzymes required for cloning.

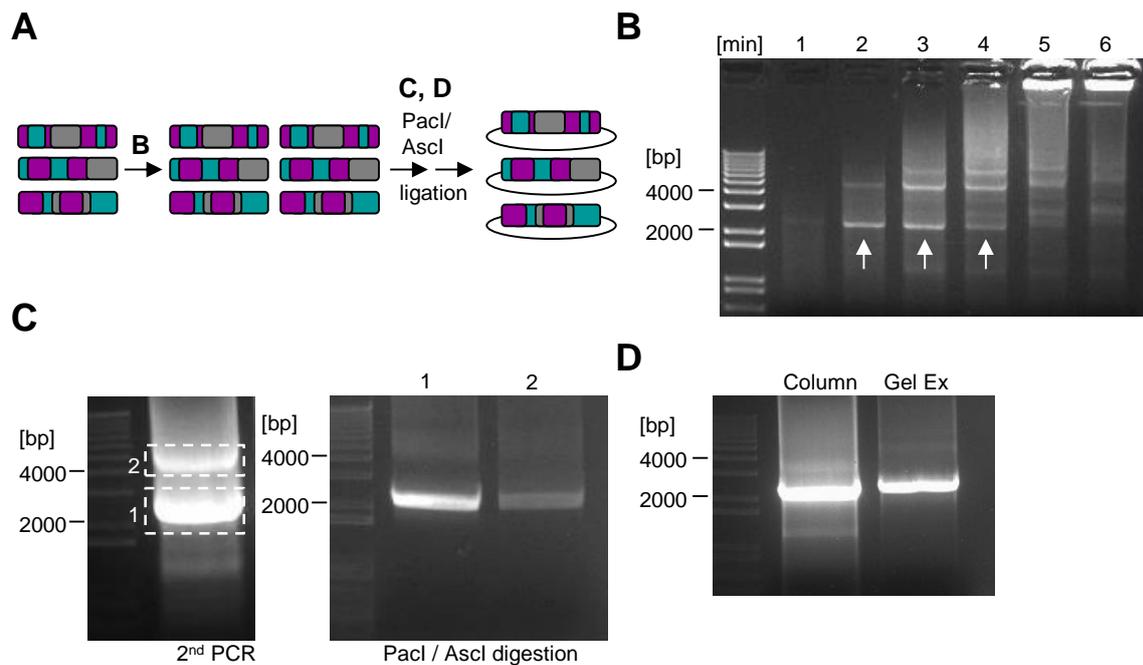


FIGURE 2-3: IMPROVEMENT OF LIBRARY PURIFICATION

(A) The schematic depicts the steps of DNA family shuffling which were studied further. (B) The elongation time of the second PCR was varied ranging from 1 to 6 min. Equal volumes of the PCR reactions were visualized on a 1% agarose gel and full-length *cap* genes are indicated by the arrows. (C) Next to the band of interest at 2.2 kb, bands with larger sizes were observed when the elongation time was increased. To address the possibility of concatamer formation, both bands, at 2.2 (1) and the one between 4 and 5 kb (2), were separately excised, purified and digested with *PacI/Ascl*. After digestion, both of them ran at a size of approximately 2.2 kb. (D) Comparison of two purification methods. After library generation by the second PCR, the DNA was either run on a gel, excised and purified or directly purified over a column. After *PacI/Ascl* digestion, the samples were again run on a gel, illustrating the higher yields from the direct column purification. Adapted from (Herrmann et al. manuscript in preparation).

## 2.2. INADVERTENT SHUFFLING OF AAP

During DNA family shuffling for the generation of novel *cap* libraries, the ORF of AAP is shuffled along. One concern that has lingered in the field since the initial publication of AAP (Sonntag, Schmidt, and Kleinschmidt 2010) was that this inadvertent shuffling of AAP might turn out to be detrimental for production of AAV particles and/or during subsequent selection steps. To address this question, we made use of a series of AAP knockout mutants that had been generated previously in our laboratory (Große 2016; Grosse et al. 2017). For most of these knockout mutants, the unconventional AAP start codon CTG was altered within the context of an AAV helper plasmid encoding AAV *rep*, *cap* and AAP. Additionally, a stop codon was introduced as early as possible in the AAP ORF without changing the overlapping VP frame. In the same study, the DNA sequences of the AAV4 and AAV5 *cap* genes were adapted to AAV2 in order to increase sequence

homology and thus shuffling efficiency. Thereby, the VP frame was kept intact but, accidentally, the AAP frame was altered, leading to the introduction of various stop codons. These versions were used as AAV4 and AAV5 AAP knockout constructs. Additionally, a trans-complementation assay had been established in which so-called crude lysates were produced by transfection of four plasmids. Usually, one plasmid encodes the transgene flanked by ITRs, one plasmid provides AAV *rep* and *cap* and the third plasmid provides helper virus functions which are required for successful AAV progeny production (compare sections 1.1.2, 4.2.5.2). Two days after transfection in HEK293T cells, the cells are harvested and subjected to iterative cycles of freezing and thawing in order to break the cells apart. Finally, the samples are spun down and the AAV containing supernatant is referred to as crude lysate. In the *trans*-complementation assay either the wild-type constructs with endogenous AAP or the AAP knockout constructs were used and were additionally supplemented with exogenous AAP or stuffer (Figure 2-5 B). For supplementation with exogenous AAP, the wild-type AAPs were subcloned under the control of the strong CMV promoter on an expression plasmid additionally bearing an HA-tag laboratory (Große 2016; Grosse et al. 2017). Previous results showed that the constructs with mutated AAP failed to produce rAAV vectors packaging an eGFP reporter as no cells expressed the transgene upon transduction (Große 2016; Grosse et al. 2017). However, supplementation with exogenous AAP could potentially recover vector production and subsequent cell transduction to wild-type levels.

---

### 2.2.1. INTERCHANGABILITY OF AAPs BETWEEN WILD-TYPE AAVS

---

The aforementioned AAP knockout constructs had been generated by a former PhD student in the laboratory (Stefanie Große) who also obtained the first evidence that AAP depletion diminishes rAAV particle yields. Here, two important related questions were addressed, *i.e.*, whether the genetic AAP knockouts would also be reflected on the protein level and whether the drop in rAAV transduction efficiency was a result of unsuccessful particle assembly. Therefore, three exemplary knockout AAP versions, namely those derived from AAV2, AAV8 and AAV9, were cloned into the expression plasmid as described in section 4.2.3.2. In order to assess the AAP expression from the resulting plasmids, they were transfected into HEK293T cells, the cells were harvested and a Western blot was performed (section 4.2.6.2).

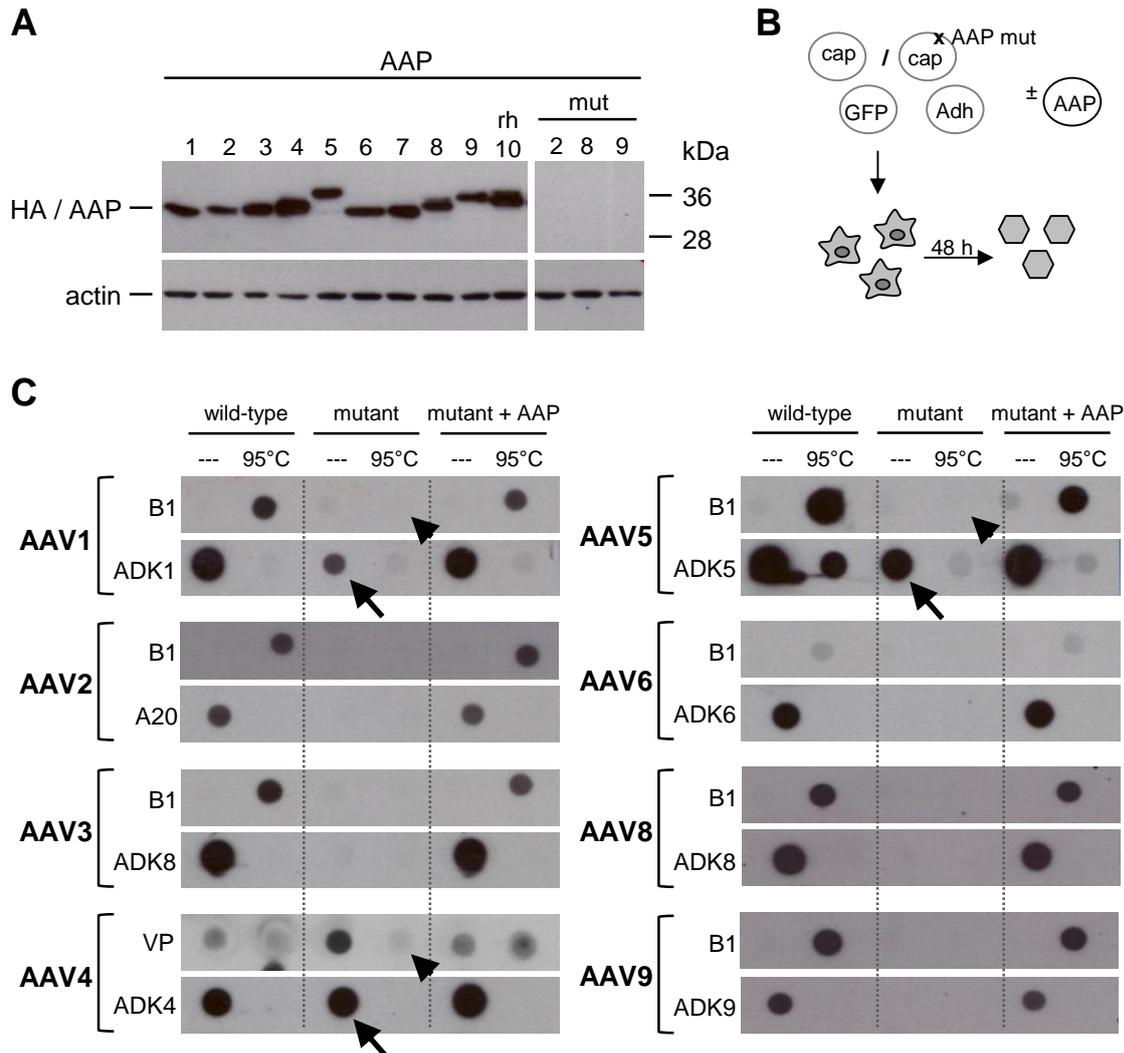


FIGURE 2-4: AAP EXPRESSION AND CAPSID ASSEMBLY MONITORED ON THE PROTEIN LEVEL

(A) AAP expression from the AAP expression plasmids was assessed by Western blot. Each plasmid was transfected into HEK293T cells and the cell pellet was lysed 48 h after transfection. Equal volumes were loaded and stained for actin (housekeeper) or HA (AAP). The label “mut” refers to the knockout versions. (B) A *trans*-complementation assay was performed during production of rAAV particles and particle assembly was assessed. The wild-type capsid, the AAP knockout mutant or the mutant co-transfected with AAP were used to package a GFP reporter-encoding vector plasmid in HEK293T cells. After 48 h the cells were harvested, cracked open by freeze/ thaw cycles and the supernatant acquired after centrifugation was used for native dot blot analysis. (C) Equal volumes of untreated or heated sample were spotted and either free VPs were detected with the monoclonal B1 antibody (or polyclonal anti-VP serum in the case for AAV4 which is not recognized by B1), or assembled particles were detected with the A20/ ADK antibodies. Structures recognized in the AAP deficient sample are indicated by the arrow (native condition) and the arrowhead (heat treatment). The experiment was performed three times and representative blots are shown. This figure was published and adapted from (Grosse et al. 2017).

Actin served as a loading control (housekeeper) and the AAPs were detected by their HA tag which is present in all expression plasmids. As expected, AAP was successfully expressed from all plasmids encoding the wild-type AAPs and the signal was absent for the knockout versions (Figure 2-4 A). Additionally, varying sizes for the different AAP variants were observed.

Once the presence or absence of expression of AAP had been successfully confirmed, the *trans*-complementation assay mentioned above was recapitulated on the protein level as well. To this end, so-called crude rAAV lysates (section 4.2.5.2) were produced using 10 wild-type capsids with endogenous AAP or AAP knockout mutants supplemented with stuffer DNA and, in the case of the mutant, with an AAP expression plasmid (Figure 2-4 B). An Adenoviral helper (Adh) plasmid provided the factors required for AAV progeny production and an eGFP expression cassette flanked by AAV ITRs was chosen as a reporter. Two days after transfection of the different components, the cells were harvested, lysed by iterative cycles of freezing and thawing and, finally, the cell debris was spun down. The supernatant containing the rAAV particles is referred to as crude lysate.

In order to assess particle assembly non-denaturing conditions are required to maintain the capsid structure. To this end a dot blot under native, non-denaturing conditions was performed by loading equal amounts of crude lysate on a membrane (section 4.2.6.3). In addition, the samples were heated at 95°C to promote disassembly of the capsids into VP subunits. Antibodies were used which specifically detect the assembled capsids, namely, A20 (AAV2), ADK1 (AAV1), ADK4-6 (AAV4-6), ADK8 (AAV3 and AAV8) and ADK9 (AAV9) (Table 4-17). In order to detect free VP, the B1 antibody was applied which recognizes VP1-3 from most serotypes except for AAV4. For the latter a polyclonal anti-VP serum was used which recognizes free VPs and possibly assembled capsids of many AAV serotypes including AAV4 (Sonntag et al. 2011). As expected for the wild-type AAV helper plasmids encoding the intact, endogenous AAP, assembled particles were detected under native conditions and VP subunits were detected when promoting disassembly by high temperature (Figure 2-4 B). Neither free VP nor assembled particles were detected for most of the AAP knockout variants, *i.e.* AAV2, AAV3, AAV6, AAV8 and AAV9. Upon *trans*-complementation of the knockout versions with AAP, however, assembly could be readily detected again comparable to the wild-types. Curiously, for the AAP knockouts in AAV1, AAV4 and AAV5 signals were detected with their respective ADK antibodies despite the absence or very weak detection of free VP (highlighted by arrows in Figure 2-4 B).

Next, *trans*-complementation assays were performed to assess the rescue potential of different wild-type AAPs for a given knockout mutant. In this assay, crude lysates are produced and the ability to mediate eGFP transduction is used as a surrogate marker of successful particle formation (Figure 2-5 A) (section 4.2.6.1). Specifically, we tested the ability of 10 different wild-type AAPs to rescue the indicated AAP knockout versions by monitoring the transduction efficiency via FACS analysis. In order to avoid saturation, a dilution of the crude lysate was chosen where no saturation was observed, *i.e.* where percentages of GFP-positive cells remained below 95%. All results were compared relative to the cognate wild-type control and allowed us to make several interesting observations (Figure 2-5 B). Generally, all serotypes tested exhibit a dependency on AAP as the knockout versions (second row) barely gave rise to infectious particles, as evidenced by the dark color indicating low to no infection. An exception were the AAV4 and AAV5 AAP mutants which exhibited transduction efficiencies of 8.9 % and 3.2 %, respectively. Upon *trans*-complementation with AAP, a high interchangeability between AAPs from different serotypes could be observed. The exceptions were again AAP4 and AAP5 which failed to support most serotypes other than their own (exemplified by a darker color). In terms of homology AAP4 and AAP5 are more divergent from the remaining wild-types (Figure 2-5 C). In contrast, the AAV4 and AAV5 knockouts were rescued by a range of other AAP variants. Curiously, the AAV3 mutant supplemented with AAP2, 3, 7 and rh10 exceeded the transduction efficiency of wild-type AAV3. Furthermore, the AAV7 and rh10 mutants are rescued more efficiently than the AAV4 and AAV5 mutants but not to the same extent as the remaining variants AAV1-3, AAV6 and AAV8-9.

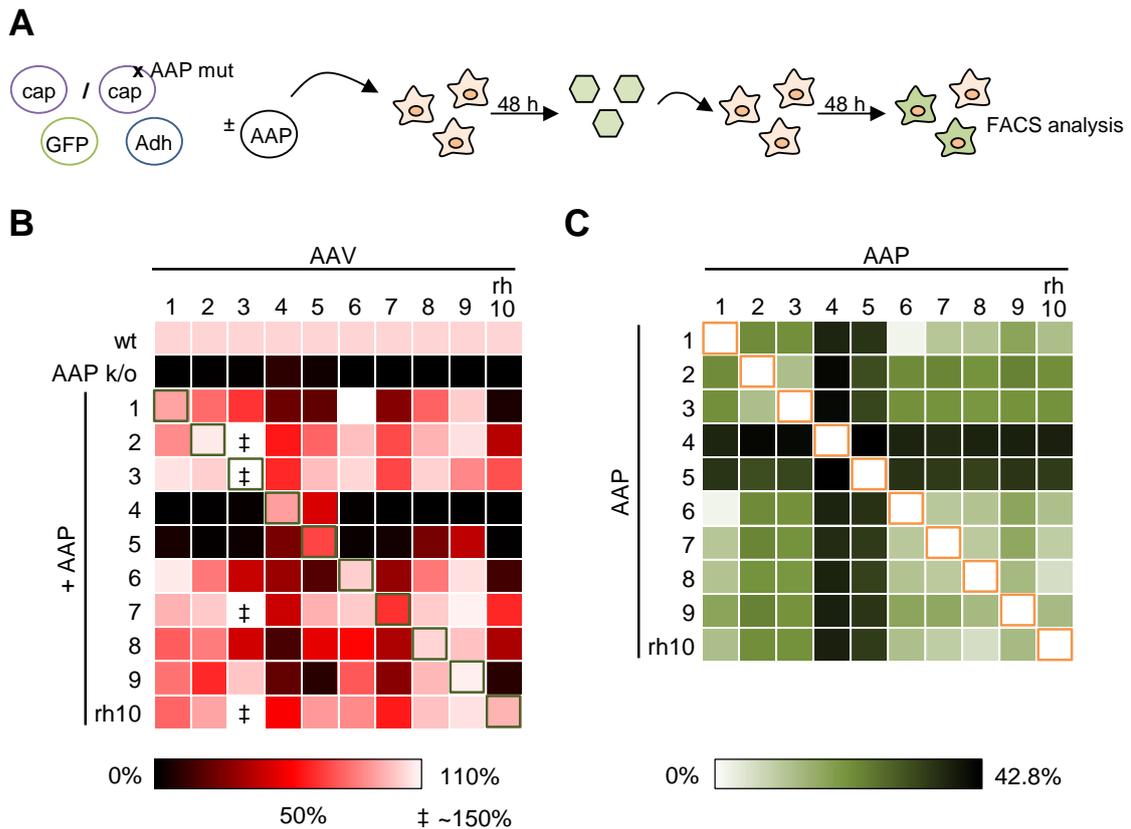


FIGURE 2-5: WILD-TYPE AAP INTERCHANGEABILITY

(A) Schematic depicting the experimental setting for the *trans*-complementation assay. Recombinant AAV particles are generated by transfection of an adenoviral helper plasmid (Adh), a transgene (GFP reporter), the capsid (wild-type or AAP mutant) and either stuffer DNA or AAP into HEK293T cells. After 48 h, the cells are harvested and exposed to freeze/thaw cycles for lysis. The cell debris is pelleted and the AAV supernatant is used to transduce cells. FACS analysis is performed 48 h later to determine percentages of GFP-positive cells. (B) Different cell lines were used for transduction and for analysis the ones were chosen where no saturation was observed, *i.e.* MCF7 (AAV2-6) and SF539 (AAV1, AAV7-rh10). All values were compared relative to the wild-type which was set to 100% and the result is depicted as a heatmap ranging from no infectivity (black) to 110% (white). In some cases, the efficiency exceeded the cognate wild-type (‡). The mean of  $n=3$  are depicted. (C) The homology of the wild-type AAPs is presented as a heatmap. The scale ranges from no difference in homology (0%, white) to a difference of maximally 42.8% (black). This figure was published and adapted from (Grosse et al. 2017).

### 2.2.2. FUNCTIONALITY OF CHIMERIC AAPS

As described above, a high degree of AAP interchangeability was observed in this work among the wild-types. This encouraged us to investigate further whether chimeric AAPs could mediate this effect as well. AAV DJ was the first chimera published when the method of DNA family shuffling was introduced into the AAV field (Grimm et al. 2008). Therefore, the AAP from the chimera AAV DJ, referred to as AAP DJ, was studied whose parental sequences are derived from AAV2, 8 and 9 (Figure 2-6 A).

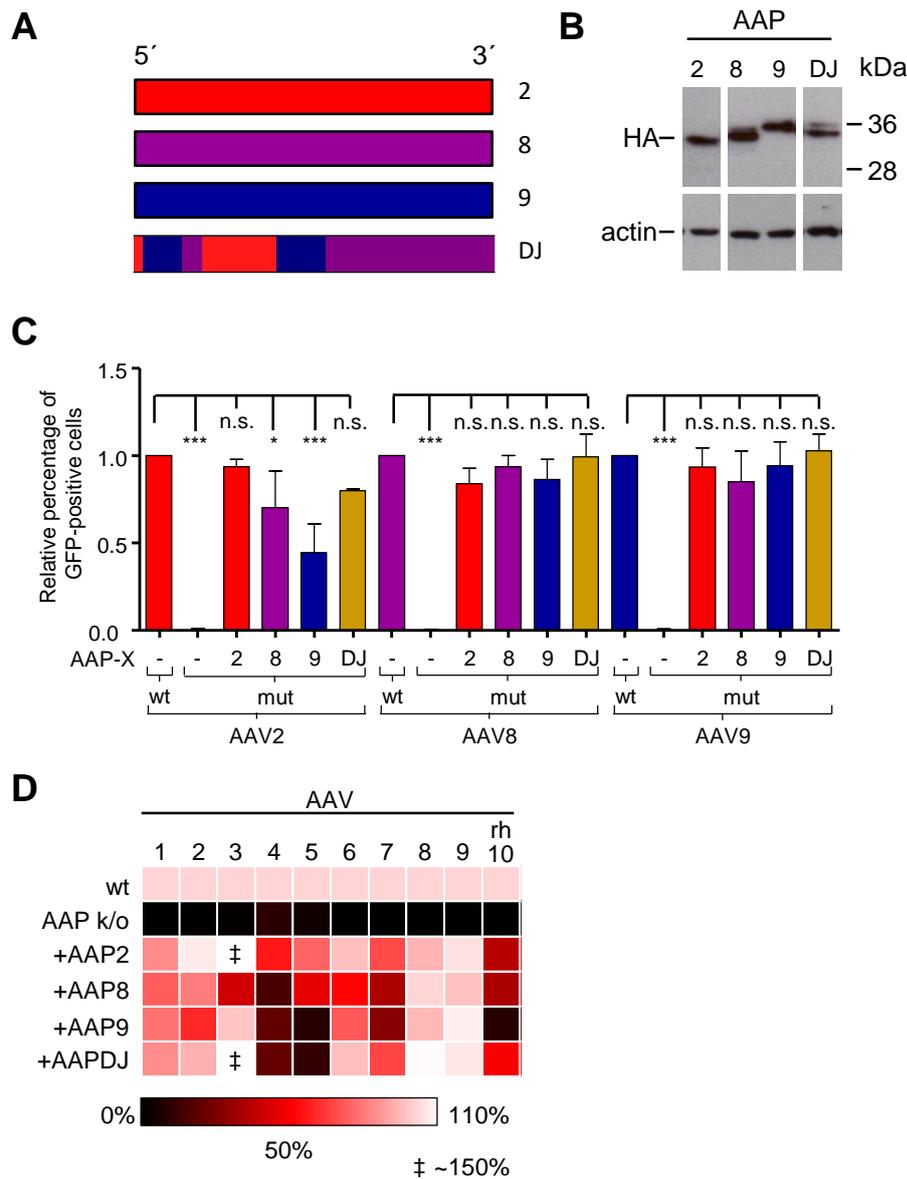


FIGURE 2-6: RESCUE ABILITY OF AAP DJ

(A) The nucleotide sequence of AAP DJ was aligned and assigned to the parental sequences of AAP2, 8 and 9. The respective parental sequence stretches were colored accordingly. (B) The Western blot analysis from Figure 2-4 A was extended to include expression from AAP DJ. (C) The rescue of AAP mutants from AAV2, 8 and 9 was tested by monitoring the percentage of GFP-positive cells relative to the respective wild-type. HEK293T cells (AAV2) and SF539 (AAV8 and AAV9) were used. Depicted are the means and standard deviation of three independent transfections. A one-way ANOVA with Bonferroni's multiple comparison test was performed and the significance is depicted above the graphs (n.s., not significant; \*,  $p < 0.01$ ; \*\*\*,  $p < 0.0001$ ). (D) The dataset from Figure 2-5 B was extended to include the results of *trans*-complementation with AAP DJ. Part of this figure (panels B-D) was adapted from (Herrmann, Grosse, et al. 2018).

Initially, successful AAP expression was confirmed by Western blotting where, similar to AAP8, AAP from AAV DJ ran as a double band (Figure 2-6 B). As before, a *trans*-

complementation assay was performed with the parental knockout versions, comparing the rescue potential of parental AAPs with chimeric AAP DJ. In the case of the AAP knockout versions for AAV8 and 9, all AAPs tested including AAP DJ were capable of restoring the efficiency back to wild-type level (Figure 2-6 C). In contrast, rescue of the AAP2 knockout mutant was only achieved with its cognate AAP and with AAP DJ, whereas *trans*-complementation with either AAP8 or AAP9 was inefficient.

Subsequently, the assay was extended to include all wild-type knockout versions tested before. Notably, AAP DJ rescued most knockouts and, comparable to AAP2, *trans*-complementation of the AAV3 mutant with AAP DJ led to transduction efficiencies even exceeding wild-type AAV3 levels (Figure 2-6 D).

Having established the functionality of one chimeric AAP originating from a successfully selected variant, the experiment was expanded to larger numbers of shuffled AAPs from unselected libraries. Therefore, a total of 60 chimeric AAPs were subcloned into expression plasmids. The AAPs originated from five different libraries and 12 arbitrarily chosen clones per library were assessed for their capability to rescue the AAV2 mutant in the *trans*-complementation assay performed before (Figure 2-5 A). The parental origin of the chimeric sequences is depicted elsewhere (Figure 6-2). The 1-100 library was generated during this study using the sequence-adapted AAV4 and AAV5 variants with repaired AAP start codons (Große 2016).

Of note, the vast majority of chimeric AAPs successfully rescued the AAV2 mutant back to wild-type levels. In total, 54 clones rescued the AAV2 mutant and only six clones failed to support infectious particle formation (Figure 2-7). These six clones (12, 16, 18, 21, 22 and 45) along with three of the variants that barely crossed the 5% threshold (19, 39 and 51) were tested again with AAP knockout mutants of their respective parental serotypes (Table 6-1). All the variants tested could rescue at least one of their parental mutants above the threshold level of 5%. Remarkably, some clones achieved a rescue efficiency of more than 80% (clones 12, 19, 21, 45).

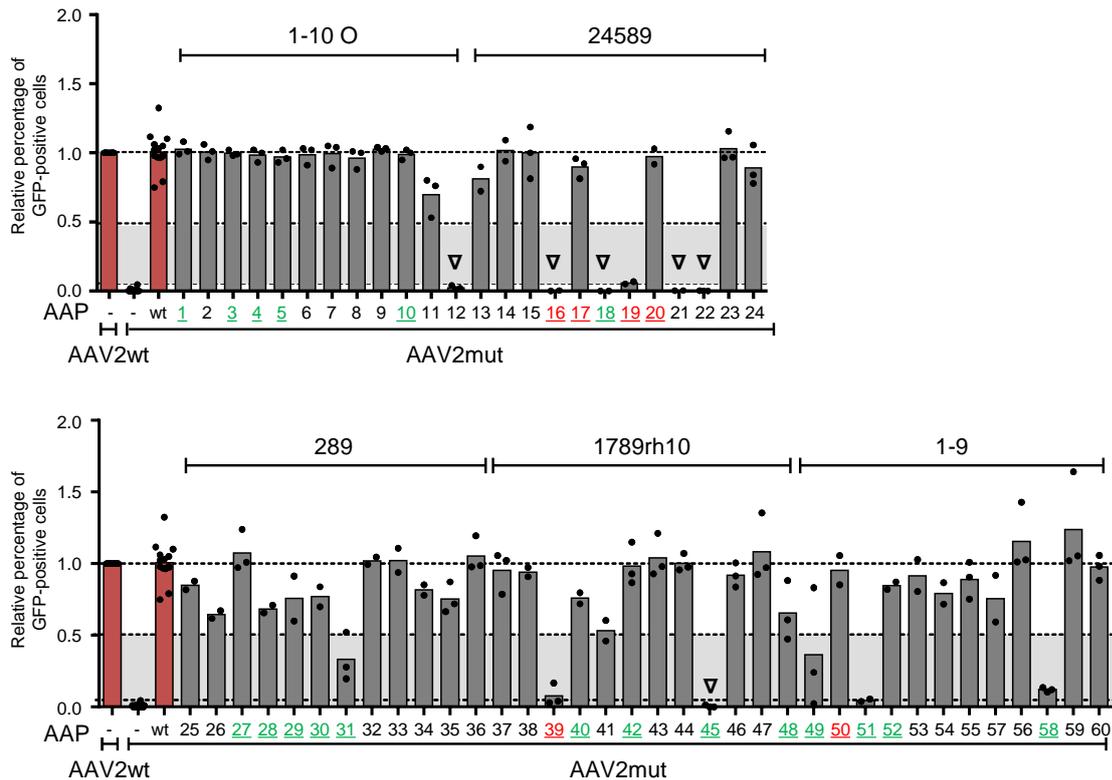


FIGURE 2-7: RESCUE ABILITY OF UNSELECTED AAPs

AAP sequences from 60 arbitrarily chosen chimeras were subcloned and used to *trans*-complement the AAV2 AAP knockout mutant. The library origin is denoted above the bars and 12 clones per library were analyzed. The amount of GFP-positive cells is depicted relative to the wild-type and the means are shown from at least two independent transductions. The light gray box indicates intermediate performers achieving rescue efficiencies of 5-50% and the triangles denote variants that failed to rescue the AAV2 mutant but succeeded with at least one parental knockout. The colored numbers describe the performance of the cognate capsid (Figure 2-8 C), *i.e.*, transduction-competent (green) or not (red). This data set is an expansion of Figure 3.10 (Große 2016) and was taken from (Herrmann, Grosse, et al. 2018).

As all chimeric AAPs tested were functional, the questions arose whether AAP addition during production has an effect on vector yield and whether the cognate capsids are functional. Therefore, the experimental setup was adapted to use wild-type capsids instead of the AAP knockout mutants and to supplement a mixture of AAPs instead of a single one (Figure 2-8 A).

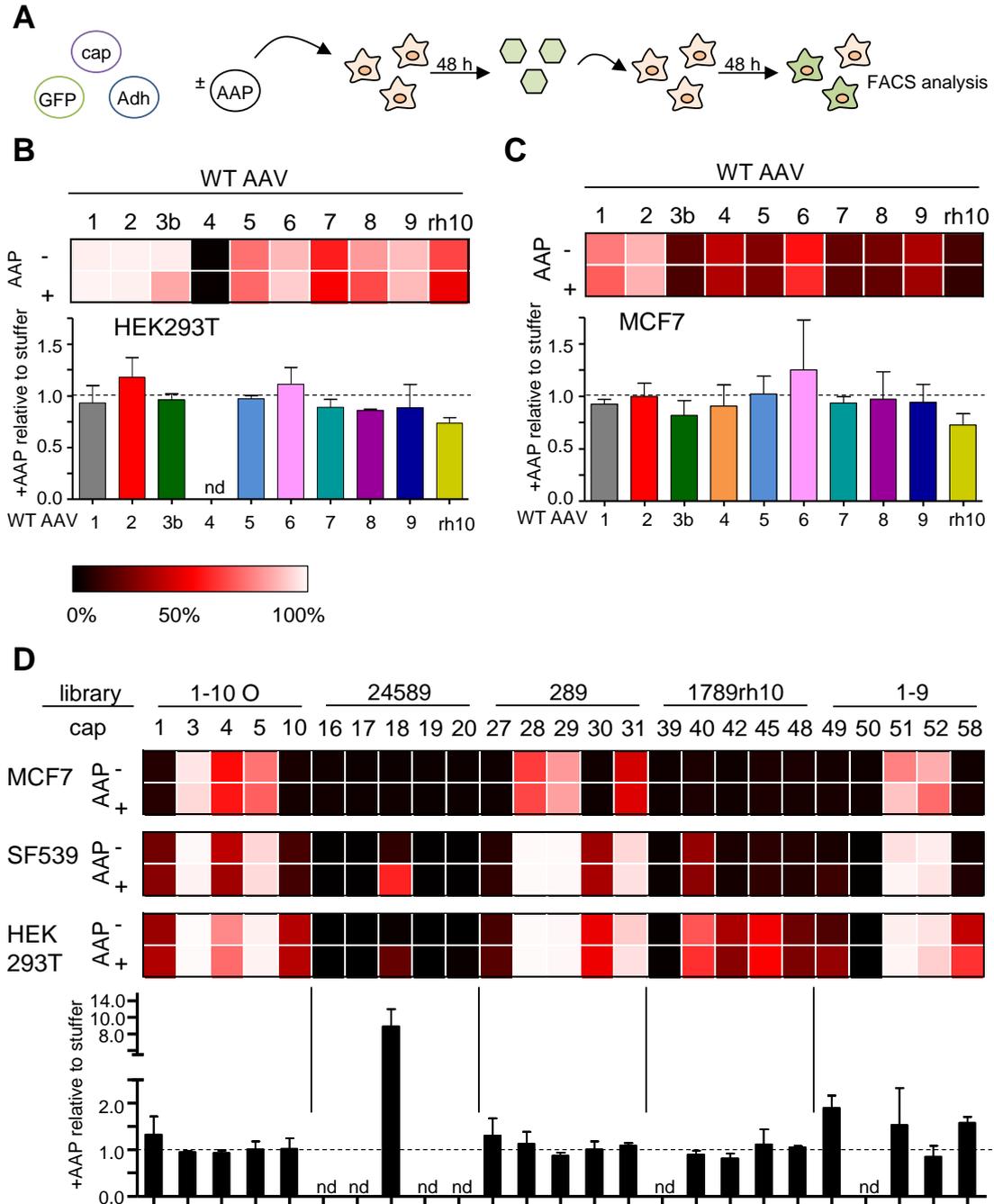


FIGURE 2-8: ADDITION OF AAP DURING RAAV PRODUCTION DOES NOT INCREASE TRANSDUCTION EFFICIENCY

(A) A *trans*-complementation assay was performed using the wild-type capsids and either a cocktail of wild-type AAPs (+) or stuffer DNA (-) was supplemented during vector production. Initially, the wild-types were tested on HEK293T (B) and MCF7 (C) cells. The mean of the transduction efficiency of three independent transfections is depicted as a heatmap. The bar graphs below show the mean and standard deviation and represent the fold difference of AAP supplementation relative to the stuffer control. The dotted line indicates no change (1). (D) The same experimental setup was expanded to five capsid chimeras per library and the indicated cell lines were tested. The bar graph below represents the same analysis as in (B, C) for HEK293T cells. nd, not determined. This data set was adapted from (Herrmann, Grosse, et al. 2018).

In this setup, the endogenous AAP will be expressed from the capsids. Firstly, the wild-types were studied and a mixture of all wild-type AAPs, *i.e.*, AAP1-10 was supplied. Two cell lines were tested, HEK293T cells (Figure 2-8 B) and MCF7 cells (Figure 2-8 C). AAV4 is unable to transduce HEK293T cells and, therefore, the two cell lines were chosen to include this AAV serotype. When comparing the results of AAP supplementation during production relative to the stuffer control, a ratio around one can be found indicating that no increase in transduction efficiency occurred. Additionally, five cognate capsids whose AAPs were tested before (Figure 2-7 B), were subcloned into a helper context to allow for their use for vector packaging. During production, an AAP mixture representing the AAPs from the parental library was supplemented, such as AAP2, AAP8 and AAP9 for capsids originating from the 289 library. Three different cell lines were screened as the tropism is unknown for these novel capsid variants. Out of 25 capsids, 19 were able to transduce at least one of the cell lines tested. In most cases, the ratio of transduction efficiencies for the AAP-supplemented to stuffer control-supplemented samples ranged around one.

However, the chimeric capsid 18 is a curious exception as its transduction efficiency increased approximately 10-fold by the addition of the AAP24589 mixture. To further dissect this interesting result, the experiment was repeated but now single AAPs were added during production. This revealed that AAP5 and AAP2 were the AAP variants predominantly contributing to this phenotype (Figure 2-9).

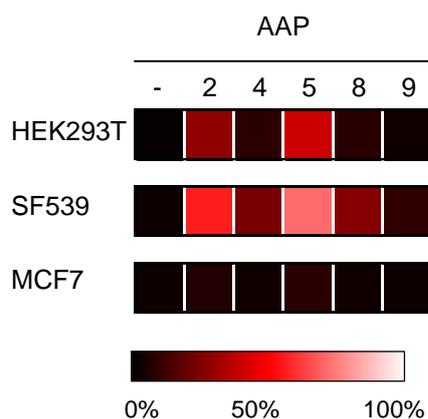


FIGURE 2-9: CONTRIBUTION OF AAPs FOR CAPSID 18 PRODUCTION

The observed increase in transduction efficiency upon AAP addition in clone 18 was further dissected. A *trans*-complementation assay was performed using the chimeric capsid from variant 18 and the indicated AAPs were supplemented. The resulting crude lysates were tested in different cell lines as before. The mean transduction efficiency from two independent transfections is shown. This data set was adapted from (Herrmann, Grosse, et al. 2018).

Finally, of note is that four out of six capsid variants that gave no transduction originated from the 24589 library, although, all six cognate AAPs were functional (compare colored numbers in Figure 2-7 B).

In summary, AAP is an essential factor for assembly of the wild-types tested and only the AAV4 and AAV5 AAP knockout mutants retained partial activity. The concern that the recombination of the AAP frame during DNA family shuffling would impair its function could be relieved as all chimeric AAPs were capable of *trans*-complementing at least one of their parental AAP knockout mutant versions.

---

### 2.3. AAV SELECTION ON HEPATIC STELLATE CELLS

---

So far, early steps of DNA family shuffling were studied with the aim to improve existing shuffling protocols and to understand an important underlying biological question, *i.e.*, the impact of recombination on the AAP frame. Next, the selection step itself was studied further by performing iterative rounds of selection on a target cell. Our laboratory has a long-standing interest in liver diseases as a variety of pathologies are associated with it. For instance, the Malaria causative agents from the *Plasmodium spp.* are transmitted by a mosquito bite and travel with the blood stream towards the liver where they establish their infection in hepatocytes. Recently, our laboratory in collaboration with the group of Ann-Kristin Mueller (Heidelberg University Hospital, Germany) showed a protective effect by over-expression of miRNA-155 against repeated challenge with the parasite (Hentzschel et al. 2014). In order to mediate sterile immunity, usually three injections of a genetically attenuated parasite were required, however, only one injection was necessary when miRNA-155 was overexpressed. The underlying biological mechanisms are not yet elucidated but a role of the non-parenchymal cell (NPC) fraction, and especially Kupffer cells, was identified. The lack of cell-type specific vectors limits further studies to *e.g.* overexpress this miRNA only in a specific cell type, such as Kupffer cells. In addition, in collaboration with the laboratory of Holger Willenbring (University of California, San Francisco [UCSF], CA, USA), we have an interest in the study of liver fibrosis and specifically, in stellate cells which contribute to disease progression. Recently, AAV6 wild-type vectors were found to target activated stellate cells in a mouse model of liver fibrosis achieved by injection of carbon tetrachloride (CCl<sub>4</sub>) (Rezvani et al. 2016). In order to study early steps of fibrosis, viral vectors targeting the quiescent stellate cells would be highly beneficial for distinct genetic manipulation in this cell population.

---

### 2.3.1. LIBRARY GENERATION AND SELECTION

---

To generate an AAV library by DNA family shuffling for subsequent selection in stellate cells, a variety of serotypes was incorporated. In detail, the serotypes 1-9 and rh10 were chosen. As noted before, AAV4 and AAV5 are less homologous than the remaining serotypes within the library. Hence, to increase their abundance the sequence optimized versions of the corresponding capsid genes were used, named 4O and 5O, respectively (sections 2.2, 2.2.2.).

All *cap* variants were mixed in equal amounts and enzymatic fragmentation was chosen as we found it to be the most robust approach for DNA family shuffling (section 2.1.1). Upon DNase I treatment an even looking smear was acquired and, as before, the range from 100 bp to 1 kb was excised, purified and used for subsequent PCRs (Figure 2-10 A). After cloning, the library was transformed into electrocompetent bacteria and an efficiency of  $1.2 \times 10^6$  colonies was achieved (section 4.2.2.2). In order to get an impression about the shuffling efficiency, 12 clones were sequenced and analyzed by Salanto as before (section 2.1.1). As hoped for, stretches of AAV4O and AAV5O were incorporated in, e.g., clones 7 and 8. Furthermore, all serotypes used for DNA family shuffling were present and on average 13.6 crossovers occurred per clone indicative of successful shuffling. At the N-terminal end, more crossover events can be observed than at the C-terminal end where the stretches tend to get larger. This is in line with the fact that AAV capsid genes display the lowest degree of homology at the C terminus that encodes the variable surface of the capsid (Figure 6-3 B, C).

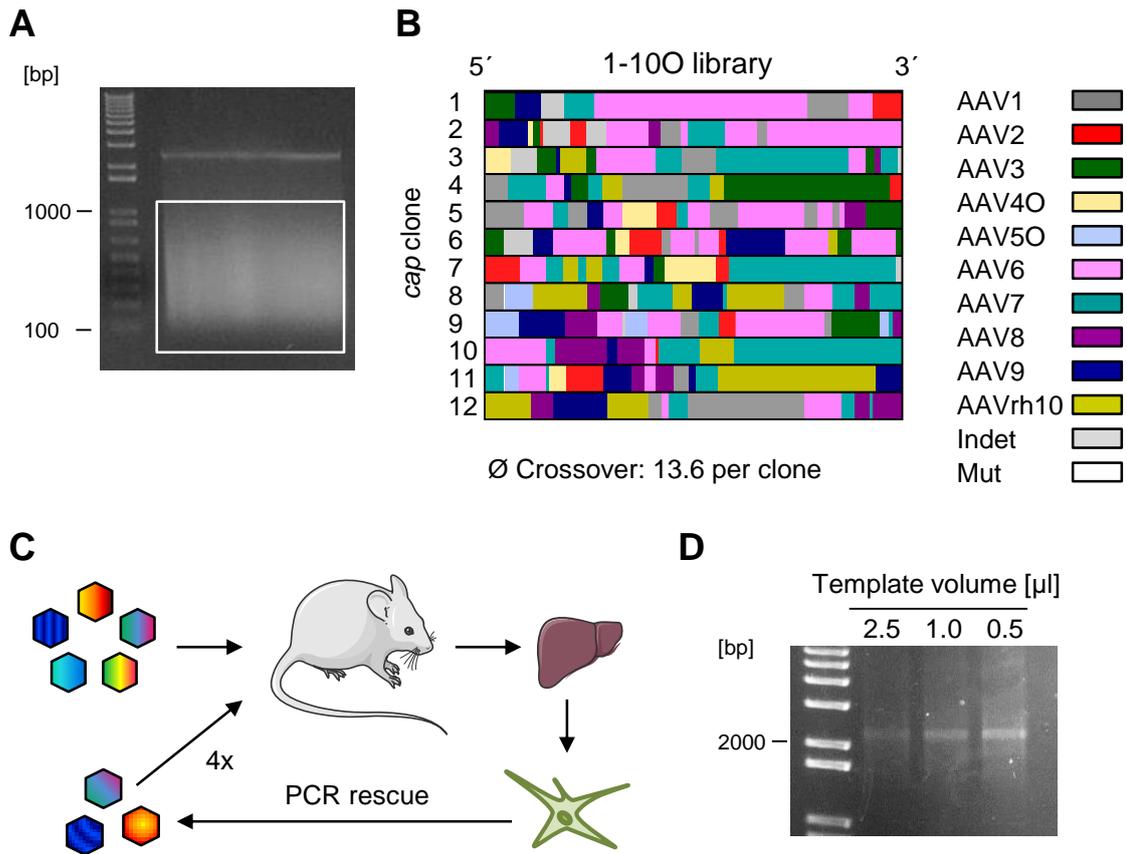


FIGURE 2-10: GENERATION AND SELECTION OF A 1-100 LIBRARY

(A) A DNase I digest was performed on serotypes 1-9 and rh10 and the box indicates the region purified for subsequent PCR steps. (B) The complete *cap* gene was sequenced from 12 clones. Each bar represents the nucleotide sequence of one clone. A type assignment using Salanto was performed and parental stretches were colored accordingly. To increase shuffling efficiency, the sequences of AAV4 and 5 were adapted to AAV2 (“O”). (C) Schematic depicting the selection strategy. A viral library was injected into a mouse and 24 h later, the stellate cells were isolated from the liver. A PCR rescue was performed using total stellate cell DNA as template and used for production of secondary libraries. In total, four rounds of selection were performed. The cartoon of the mouse was taken from the servier medical art collection (<https://smart.servier.com/>). (D) Representative image of the PCR rescue. As a first step, a small-scale test PCR was performed using different template volumes as input material. Equal volumes of the PCR reactions were loaded and the condition yielding the highest intensity was used for large-scale PCRs.

Having confirmed successful DNA family shuffling, the selection in stellate cells was initiated in collaboration with the laboratory of Holger Willenbring (Figure 2-10 C). The library was purified by cesium chloride gradient density centrifugation (section 4.2.5.5) to remove empty particles and  $5 \times 10^{11}$  vector genomes (vg) were injected into one mouse expressing GFP specifically in stellate cells (section 4.2.7.1). After 24 h, the liver was perfused and stellate cells were sorted by FACS. In a quiescent state, stellate cells store Vitamin A in fat droplets which can be excited by UV light (Mederacke et al. 2015).

Therefore, cells which were double-positive for GFP and Vitamin A were used for further DNA isolation and production of secondary libraries. Prior to setting up a large-scale PCR, a test PCR was performed. Here, different template volumes were tested in the PCR reaction rather than specific ng amounts to determine the condition achieving the highest product yield. The rationale behind this strategy is that total DNA is isolated from the cells and the portion of AAV genomes is unknown. A low portion of AAV genomes could require higher template volumes and vice versa, high portions of AAV genome could lead to template inhibition during PCR and hence, lower template volumes are required. As seen in Figure 2-10 D a reduction in template leads to higher AAV *cap* recovery and thereby, the best PCR condition was identified.

In total, four rounds of selection were performed and 64 clones originating from the different selection rounds were sequenced. The sequences were sorted according to their C-terminal parental assignment (see below). When considering the individual selection rounds, enrichment at the C-terminus of AAV1 and AAV6 was observed during the first three selection rounds (Figure 6-4). Unexpectedly, the fourth round yielded chimeras with a C-terminal enrichment for AAV7 and this observation was recapitulated upon repetition of the fourth round with two more mice.

---

### 2.3.2. VALIDATION OF SINGLE CHIMERAS

---

In order to choose the most promising variants for individual experimental validation all clones were sorted into different groups based on their C-terminal parts (Figure 2-11). The groups identified were mostly derived from AAVrh10 (10-stretch), AAV7 (7-stretch) or AAV1 (1-stretch). Additionally, we specified groups that contained combinations. One showed a pronounced crossover from AAV7 to AAVrh10 and then back again to AAV7 (7-rh10-7). The other two were mixes of AAV1, AAV6 and AAV7 (1/6/7 mix) or AAV1 and AAV6 (1/6 mix). Only five out of 64 sequences did not fit to any of the groups identified. Next, at least two representatives from each group were chosen for further *in vivo* validation. Therefore, the candidate capsids were transferred from the replication-competent background into a helper context which allowed for packaging of an eYFP reporter. The vectors were produced in small-scale and purified by Iodixanol gradient density centrifugation (section 4.2.5.4).

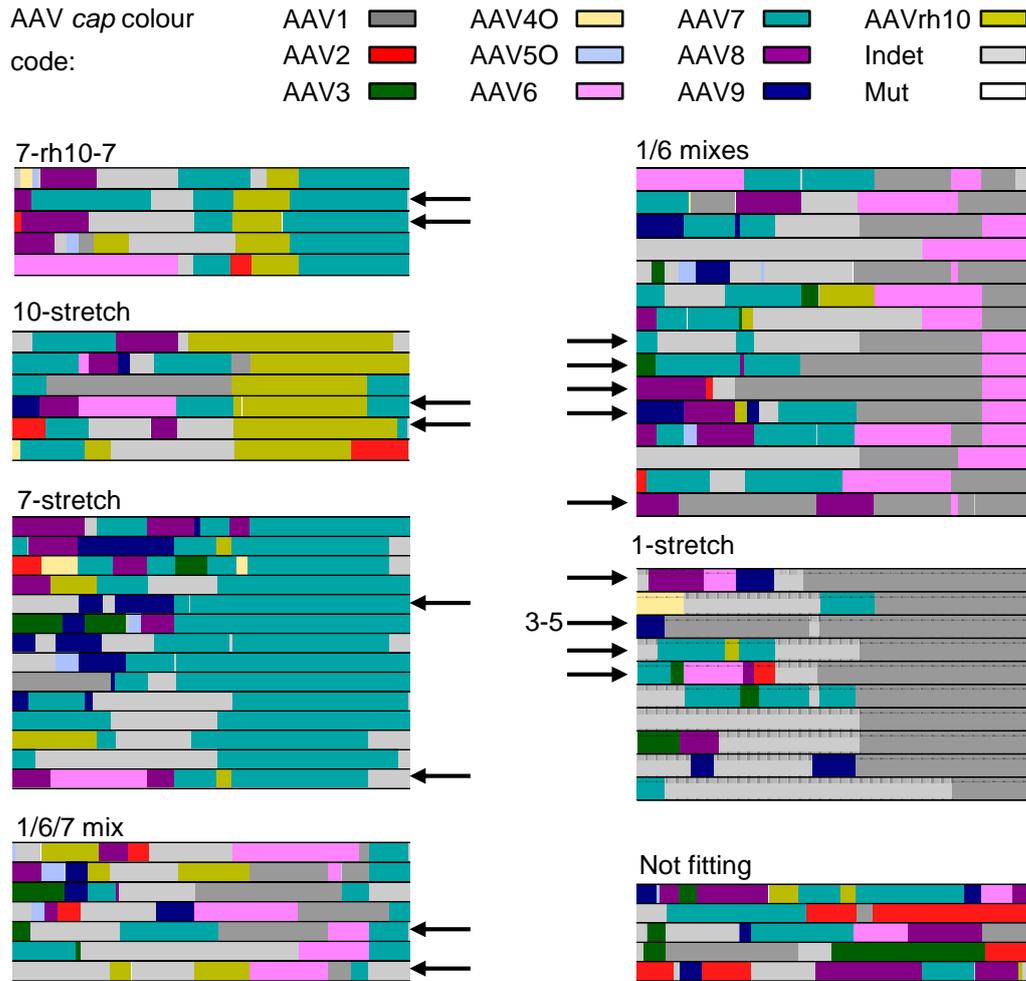


FIGURE 2-11: AMINO ACID SEQUENCE OF SELECTED VARIANTS

Each colored bar represents the complete amino acid sequence of a single chimeric capsid. The color code for the parental references is depicted at the top. The 1-100 library was subjected to an iterative *in vivo* selection in mice and AAV *cap* sequences were recovered from isolated stellate cells. After each round, chimeric amino acid sequences were isolated (eight from the first and second round, 18 from the third, 30 from the fourth) and finally sorted according to their C-terminal domain. Single chimeras (indicated by the arrow) from each identified group were randomly chosen for further *in vivo* validation. Capsid 3-5 is highlighted as it became the lead candidate in later analyses.

### 2.3.2.1. HIGH PRODUCTION CAPACITY UPON SELECTION

Interestingly, when the clones selected for validation were produced individually, we noted titers comparable or even higher to the parental wild-types (Figure 2-12). The latter (shown in white) were at the lower end with titers in the range of  $5 \times 10^{11}$  vg/ ml. Three chimeric variants produced also at a lower titer (4-8, 4b2-08, 4-5) and six achieved a titer at wild-type level (4-10 to 3-10). Eight chimeric capsids yielded titers in the range of  $10^{12}$  to  $10^{13}$ , *i.e.*, over ten-fold higher than all wild-types.

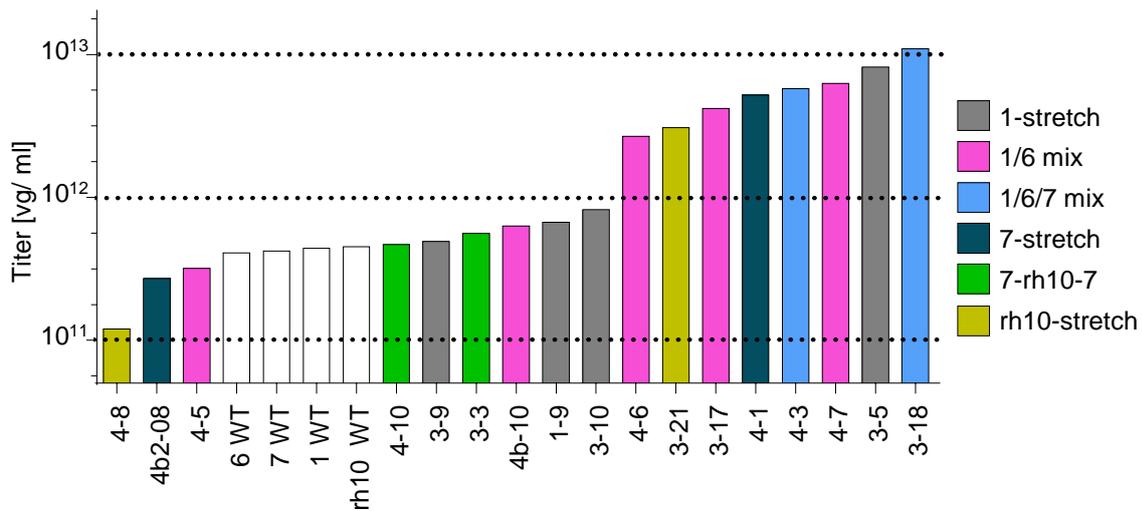


FIGURE 2-12: TITERS OF CHIMERIC CAPSIDS

The indicated capsid variants were produced by transfection of 15 plates and iodixanol density centrifugation purification. The chosen color code refers to the group shown in Figure 2-11.

From these observations, the question arose whether this increase in production capacity might be linked to novel properties of the AAPs, as previously observed with AAP DJ (section 2.2.2). Therefore, AAPs derived from low- and high-producing chimeras (4-8, 4b2-08, 4-5 and 3-5, 3-17, 3-18, 3-21, 4-3, 4-6, 4-7), as well as AAPs from chimeras producing comparable to the wild-types (3-3, 3-9), were cloned into the expression construct and their rescue potential was tested with the AAP knockout mutants of the most abundant parental serotypes, *i.e.* AAV1, AAV6, AAV7 and AAVrh10. The full AAP sequences are depicted in the supplements in Figure 6-2. All selected AAP variants were able to rescue at least one of the AAP knockout mutants (Figure 2-13). In almost all cases, AAV1 and AAV6 could be rescued back to wild-type levels whereas AAV7 and AAVrh10 were less efficiently rescued. Exceptions were the AAP variants derived from 3-3, 3-17, 4-3 and 4-7 that potently rescued the AAV7 and AAVrh10 AAP mutants. However, no striking differences between the different AAPs were observed such as an increase in transduction efficiency seen for the AAV3 AAP knockout mutant (section 2.2.1). In conclusion, the increase in production efficiency was most likely not mediated by AAP.

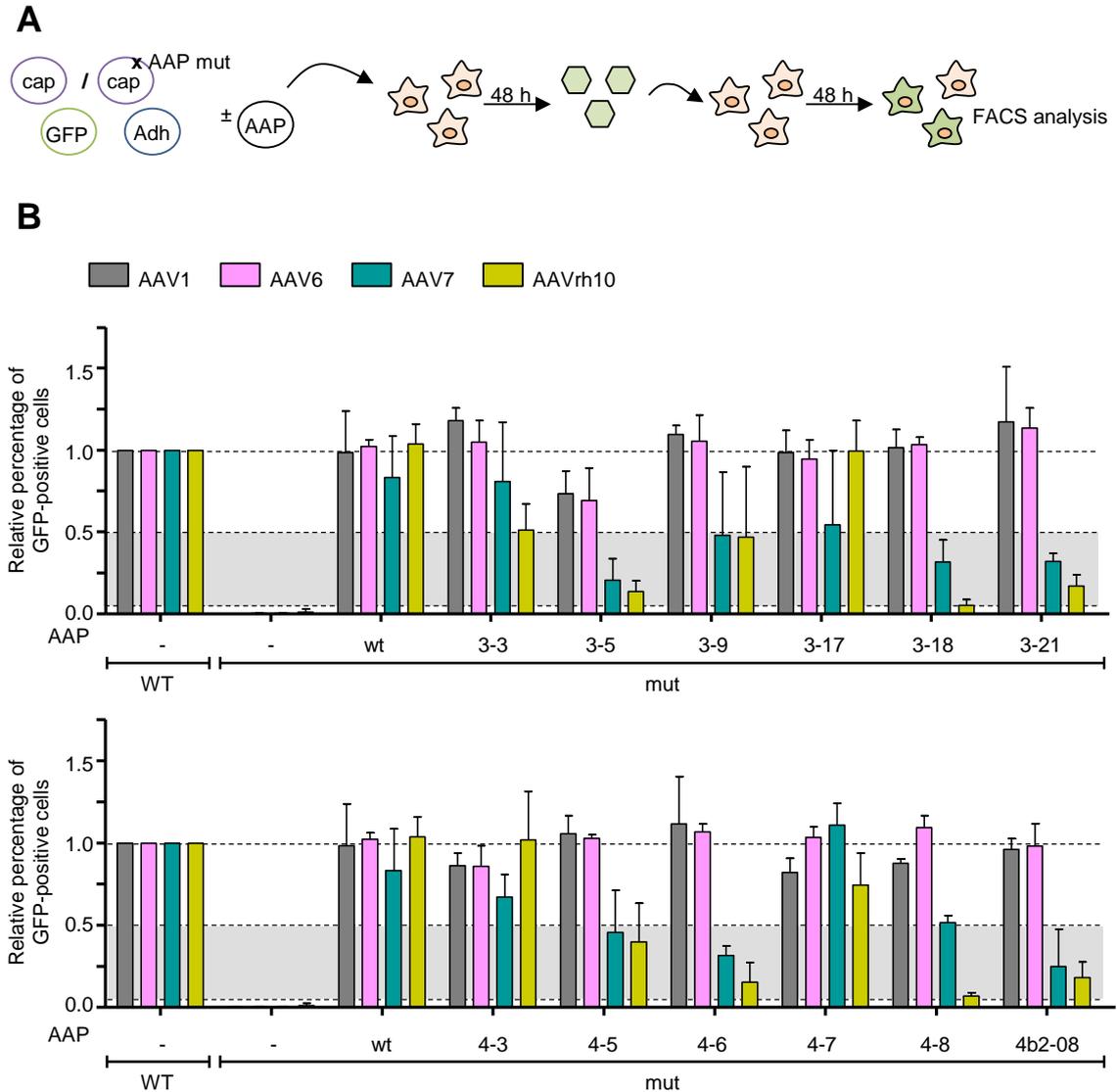


FIGURE 2-13: RESCUE ABILITY OF AAPs WHICH UNDERWENT SELECTION

(A). Different AAPs (Figure 2-12) subcloned from chimeras which underwent selection in stellate cells were tested for their potential to rescue the indicated AAP knockout plasmids. (B) The amount of GFP-positive cells was compared relative to the respective wild-type control (set to one) and the means with standard deviation from three independent transfections are depicted. Comparable to Figure 2-7 the intermediate performers with a rescue efficiency of 5 to 50% is indicated by the grey box. This data set adapted from (Herrmann, Grosse, et al. 2018).

### 2.3.2.2. INITIAL VALIDATION ON SINGLE CANDIDATE LEVEL

The produced variants were injected into mice at a dose of  $1 \times 10^{11}$  vg per mouse. Ten days later, the mice were sacrificed and liver slices were processed for immunostainings. The vector delivered an eYFP reporter and stellate cells were stained for desmin, which is a commonly used marker for this cell type. The wild-types, AAV1 and AAV6 mostly transduced hepatocytes (Figure 2-14 A). In contrast, the tested chimeras showed

different transduction profiles and most notably novel tropisms were identified. This indicates that the applied selection pressure lead to enrichment of promising variants.

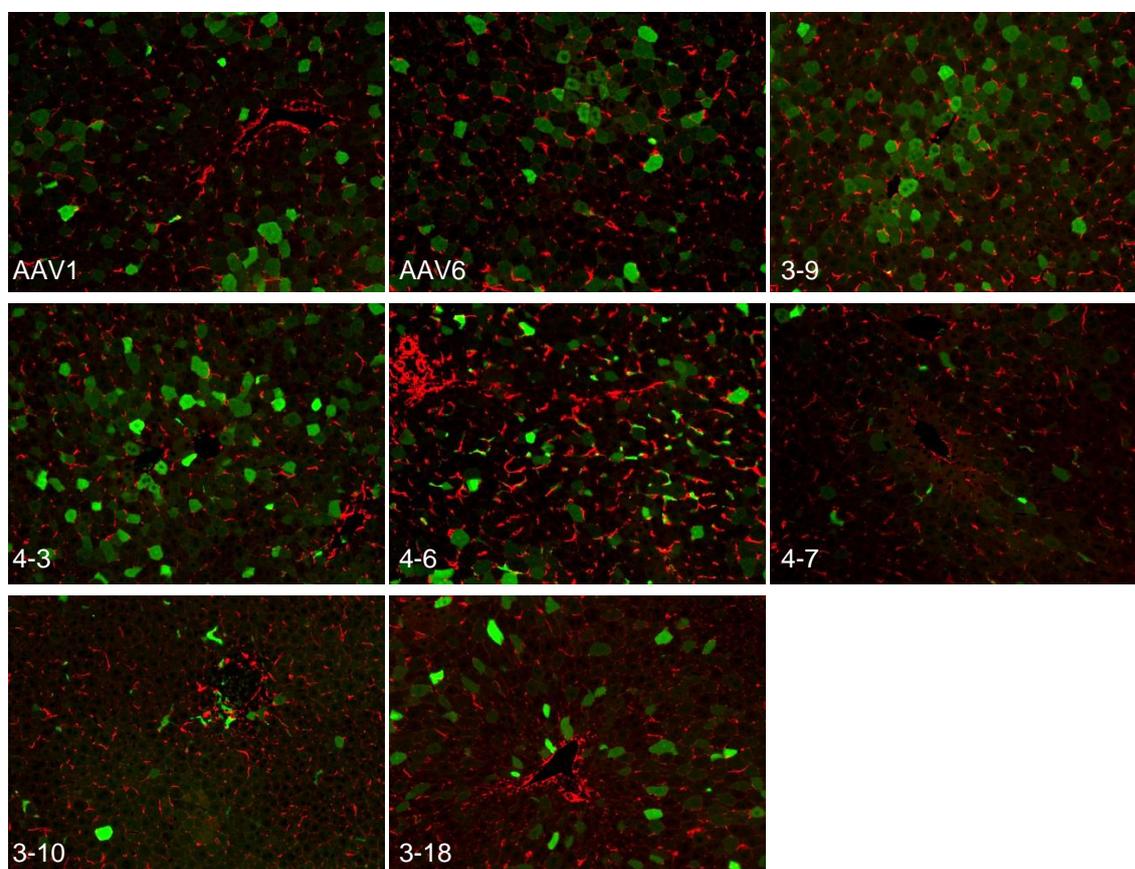
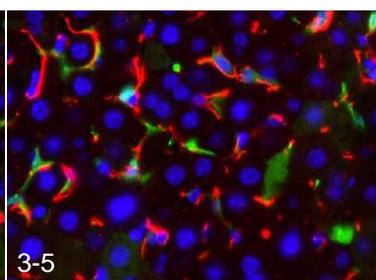
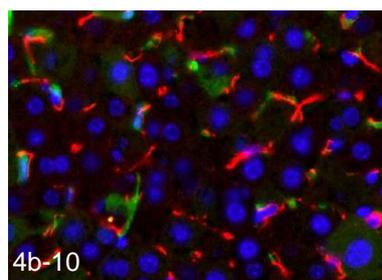
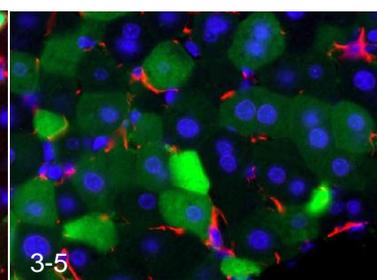
**A****B****C**

FIGURE 2-14: IMMUNOSTAININGS OF LIVER SLICES

Different AAV vectors enriched during selection in stellate cells were validated *in vivo*. (A) The indicated wild-types or chimeras were injected into mice and the livers were processed for immunostainings. Shown are representative images of the transduction efficiency for some of the vectors tested. The first number indicates the round of selection and the second number refers to the clone. The reporter (green) and stellate cell marker desmin (red) were stained. (B) Two more variants are shown in a higher magnification. Additionally, the nuclei were stained with DAPI. (C) For the variant 3-5 a different tropism was observed in another mouse. The brightness was increased by 20% for all pictures. The images were acquired by Regina Espanol-Suner (Willenbring lab).

In detail, a couple of variants efficiently transduced hepatocytes, such as 3-9 and 4-3. Others were low transducers in general, e.g., as seen with 3-18 and 4-7. Most interestingly, some capsids developed novel tropisms, namely, 4-6, 3-10, 3-5 and 4b-10 (Figure 2-14 A, B). In these cases, a low rate of hepatocytes was transduced and cells were targeted which closely localized with the desmin staining.

It has to be noted, however, that the liver is a complex organ and that the cells are in close proximity, which together complicated the identification and quantification of vector-targeted cells. Accordingly, the staining was not always distinct enough to determine whether the signals co-localized or whether a nearby cell had been transduced. Additionally, we observed variations in tropism between mice as exemplified with variant 3-5, which targeted stellate cells in one mouse but mostly hepatocytes in another (Figure 2-14 C).

For this reason, an alternative strategy was followed in collaboration with Steven Dooley (University hospital Mannheim, Germany). The AAV variant 3-5 was injected into mice and, instead of staining liver slices, the liver was perfused and single cell types were cultivated upon isolation. The cultured cells were then stained for the reporter and for a cell-specific marker. The identity of hepatocytes was confirmed by HNF4A staining, stellate cells were identified by desmin staining. Kupffer cells were recognized by F4/80 and, lastly, LSECs were stained with LYVE1. Notably, the cell isolations appeared homogenous as most nuclei (column 3, Draq5) localized with the cell-specific marker (column 2) (Figure 2-15 A). Interestingly, while barely any hepatocytes and LSECs were transduced by the 3-5 vector, the vast majority of stellate cells and Kupffer cells were transduced (Figure 2-15 A, B). This finding independently confirmed the promising result from the immunostainings (Figure 2-14).

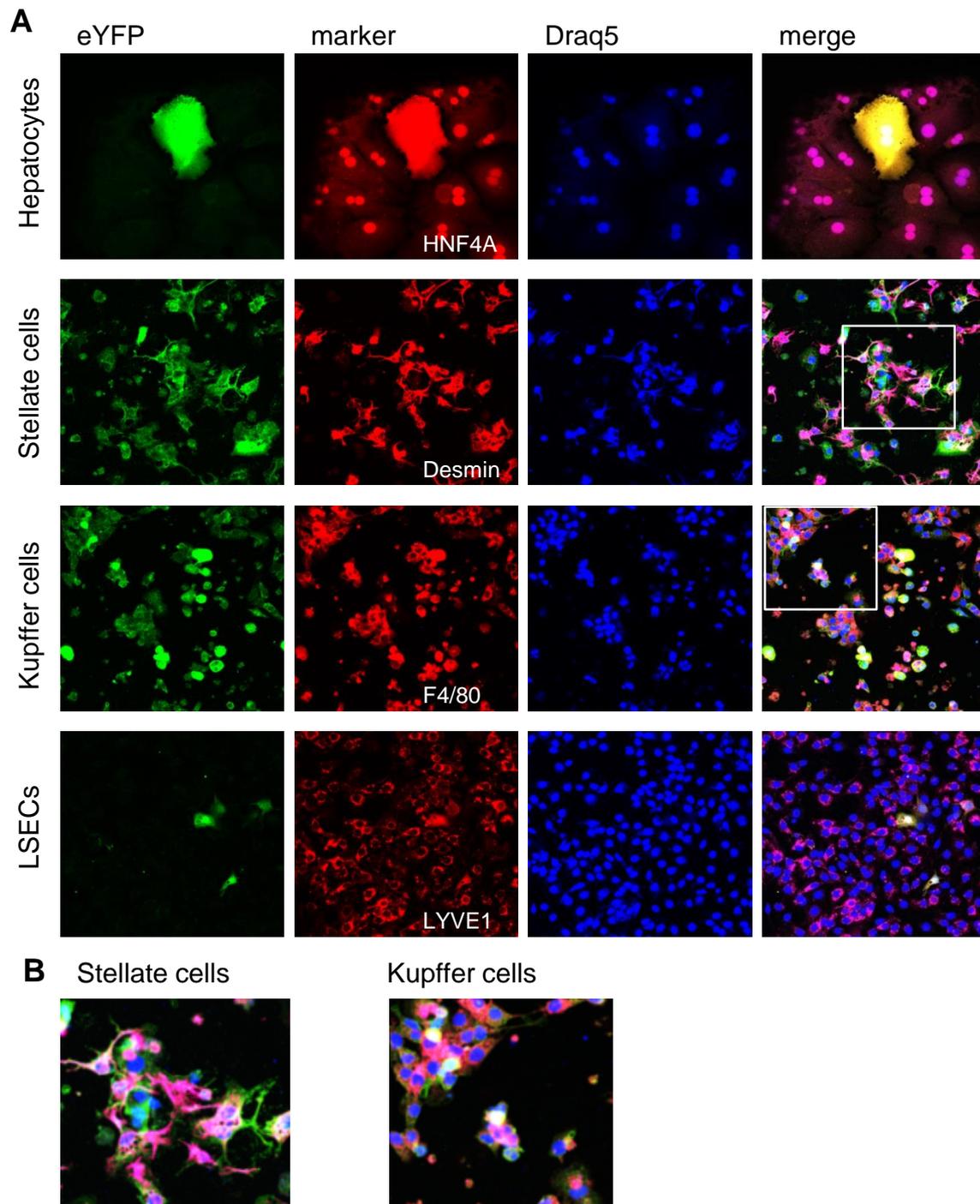


FIGURE 2-15: CELL SEPARATION AND STAINING

Determination of transduced cell types after AAV3-5 application in mice. (A, B) The variant 3-5 was injected into mice and the four cell types were isolated. The cells were taken into culture and stained for the reporter (green), a cell-type specific marker (red) and the nucleus (DRAQ5). An overlay of all channels is depicted in column "merge". (A) In the overlays for stellate cells and Kupffer cells, part of the image is highlighted by a box. In (B) this section was enlarged for a better visualization of the different stainings. The pictures were acquired by Christof Dormann (Dooley lab).

### 2.3.2.3. POOLED APPROACH BY NEXT-GENERATION SEQUENCING

So far, the novel tropism displayed by AAV 3-5 was confirmed by an alternative approach and from the immunostainings three more variants are potentially interesting (4-6, 3-10 and 4b-10) (Figure 2-14). However, the validation on single-capsid level requires high numbers of mice and time. In order to cover more variants in parallel a bulk validation approach was developed based on next-generation sequencing (NGS) (section 4.2.7.2).

Previously, another member of the Grimm laboratory – Jonas Weinmann - has established a screening protocol where multiple barcoded, AAV capsid variants are injected into a mouse and the distribution of the capsids is determined on the DNA and RNA levels by NGS analysis. The barcode is a small, unique nucleotide stretch incorporated into the 3' untranslated region (UTR) of the eYFP reporter (Figure 2-16).

Here, all variants of interest were separately produced and equal virus amounts were pooled. Of note, a primary library existed in the laboratory containing 82 variants representing wild-type capsids and peptide insertion variants thereof. A secondary library was generated where 28 chimeras derived from the *in vivo* HSC selection of this screen (denoted AH) were chosen including variants that were already tested *in vivo* (Figure 2-14) and variants which were newly sequenced (Figure 6-5). Additionally, chimeras from an unrelated selection performed by a different lab member (denoted JEA) and additional published capsid variants were included.

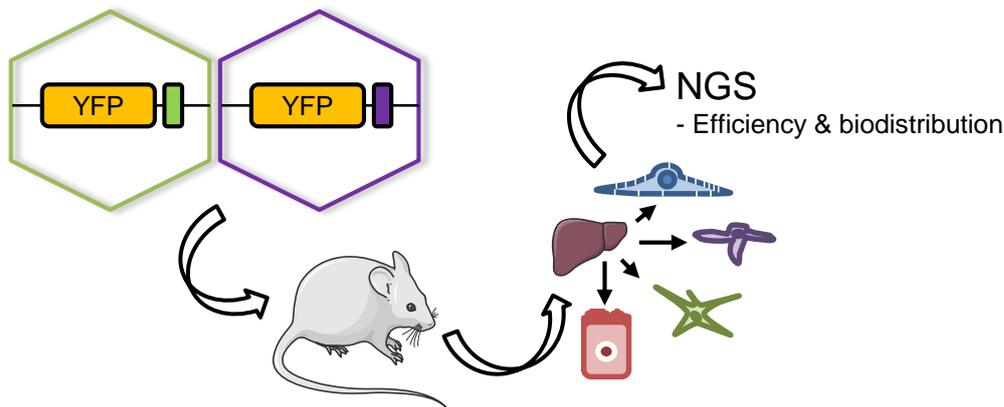


FIGURE 2-16: BULK VALIDATION BY NEXT-GENERATION SEQUENCING

Each capsid variant tested contained a unique barcoded vector genome. All variants were mixed in equal amounts and injected into mice. Ultimately, the livers were perfused and single cell types were isolated, namely, hepatocytes, stellate cells, LSECs and stellate cells. From each sample, the genomic DNA and RNA was isolated followed by qualitative and quantitative NGS analysis of barcode distribution. The cartoon of the mouse was taken from the servier medical art collection (<https://smart.servier.com/>).

To resolve the tropism of these capsids within the liver, four cell types were isolated from each individual mouse, namely, hepatocytes, stellate cells, LSECs and Kupffer cells. The two libraries were mixed and injected into mice at a total dose of  $1.57 \times 10^{12}$  vector genomes. Theoretically, each of the resulting 157 variants injected is represented with  $1 \times 10^{10}$  particles. Genomic DNA (gDNA) as well as RNA were isolated and the latter was processed for cDNA synthesis. Importantly, the vector-encoded barcode will be co-transcribed due to its positioning within the 3'UTR, thus allowing for monitoring of actual transgene expression on the RNA level and comparison with vector genome delivery on the DNA level. All samples including the input library were then processed for NGS and the recovered values were normalized to the abundance of a given variant in the input library.

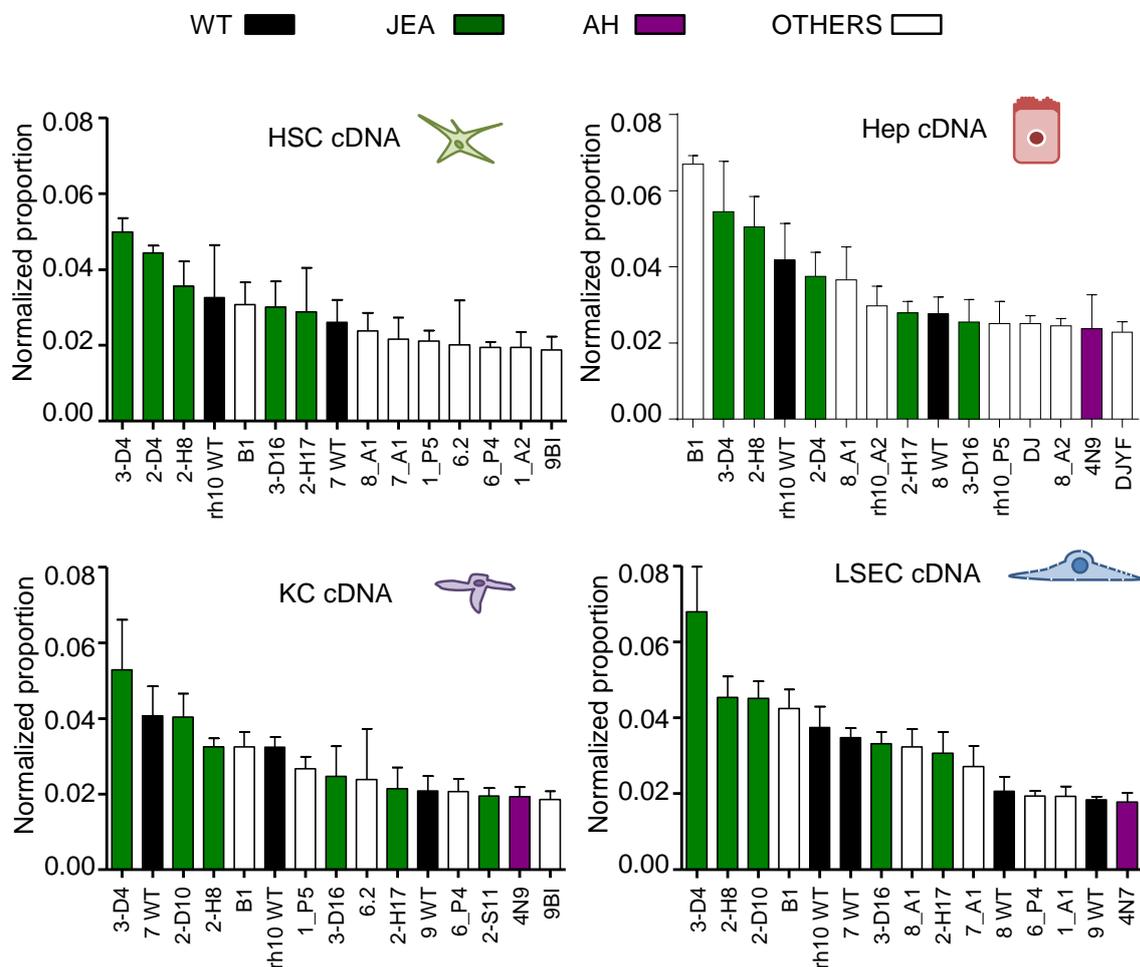


FIGURE 2-17: CDNA PROFILES FOR LIVER CELL SUBTYPES

Capsids were ranked by transgene expression (cDNA) and the top 15 hits are shown. Within a given cell type, the sum of all barcode read counts was set to 1.0 and the proportion of each variant determined. The bars were colored according to their origin: wild-types (WT) are colored in black, variants originating from an unrelated screen (JEA) in green, chimeras originating from this study (AH) in purple and all other variants in white. HSC, hepatic stellate cell; Hep, hepatocyte; KC, Kupffer cell

Within a given cell type, the proportion of each variant contributing to the transgene expression on the RNA/cDNA level was determined (Figure 2-17). Surprisingly, none of the chimeras selected in the *in vivo* HSC screen appear in the top 15 ranking in stellate cells. Instead, the chimeras originating from an unrelated screen in muscle (JEA) showed pronounced off-targeting to the different liver cell types. The two serotypes AAV7 and AAVrh10 were most effective among all wild-types in stellate cells and were found as well in the other fractions. There were two variants identified in this study which appeared in the top 15 of other cell types, namely, 4N9 (Hep, KC) and 4N7 (LSEC).

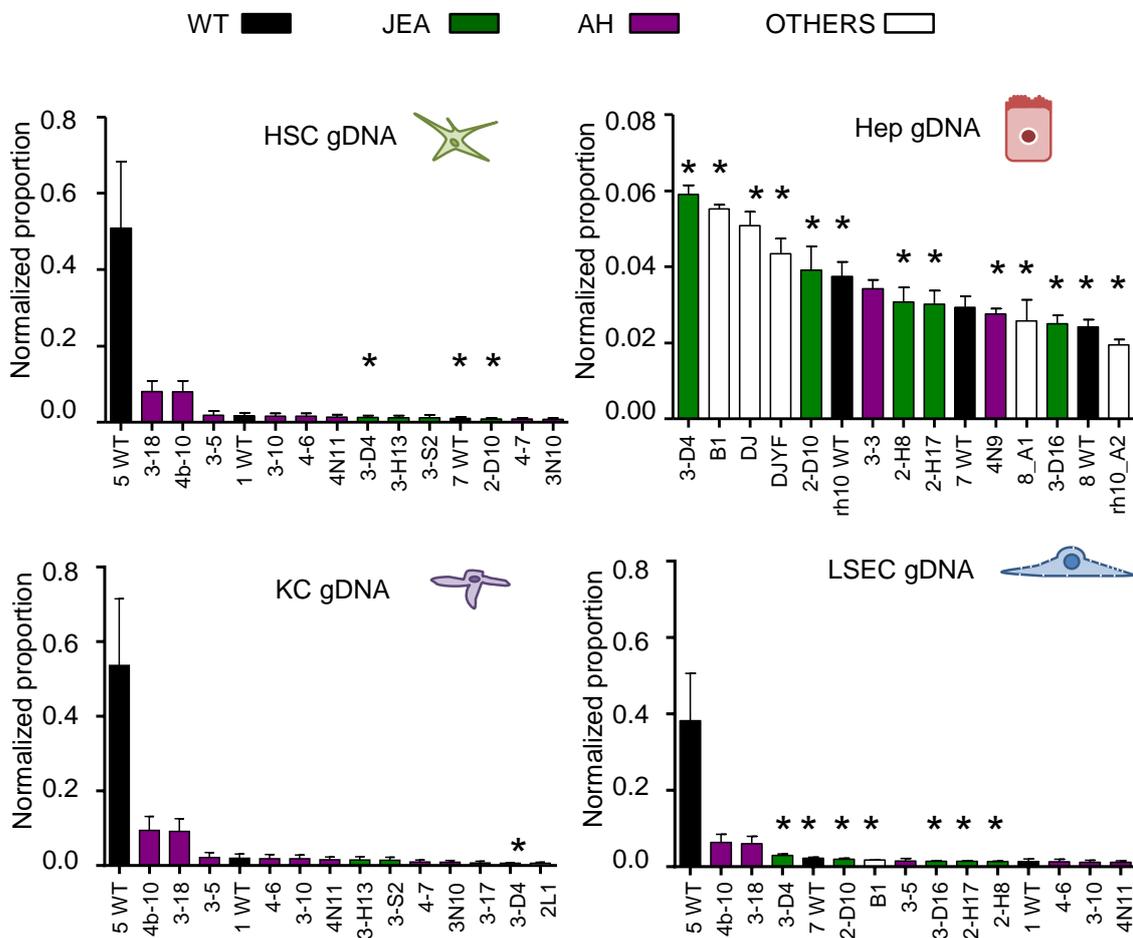


FIGURE 2-18: GDNA PROFILES FOR LIVER CELL SUBTYPES

Vector-encoded barcodes were ranked on the gDNA level. Shown are the top 15 hits. Variants that showed up in the top 15 list on the expression level (cDNA) as well are marked by an asterisk (\*). Note that the scale of the Y axis for the hepatocytes differs from the others. More details are found in the legends of Figure 2-17. Additionally, a version was included where AAV5 wild-type was excluded from the graphs for a better visualization (Figure 6-6).

In addition to cDNA expression, the capsids were ranked on the gDNA level as well. Notably, whereas the graphs looked comparable among the various cell types for cDNA, this was different for the gDNA ranking. Within the non-parenchymal cells (NPC), *i.e.*, stellate cells, Kupffer cells and LSECs, a large proportion of AAV5 can be found which is absent from hepatocytes (Figure 2-18). In stellate cells, a high number of the *in vivo* HSC selected variants now appear in the top 15 rank. In total, there are three wild-types (AAV1, AAV5, and AAV7), four originate from the unrelated screen (JEA), and eight were selected variants from this study.

Intriguingly, the ranking is highly similar among the different NPCs and six of the eight top hits appear in both, LSECs and Kupffer cells (4b-10, 3-18, 3-5, 4-6, 3-10 and 4N11). As the gDNA and cDNA graphs resembled each other for hepatocytes, the variants were indicated which occur on both, gDNA and cDNA level (compare the asterisks in Figure 2-18). Curiously, hepatocytes displayed a good correlation between gDNA and cDNA (13/15) and LSECs were intermediate (7/15). In contrast, stellate cells (3/15) and Kupffer cells (1/15) had a very low coupling of gDNA with cDNA.

To assess this observation further, a higher sample size was used. Hence, the gDNA rank was assigned to the cDNA rank of the top 50 hits on the transgene expression level. If one would assume a perfect transcription of gDNA into mRNA/cDNA, a line with the slope of 1.0 would be expected. The data for a higher sample size was plotted and analyzed by linear regression. Of note, the linear regression analysis considers the ranking only and the top 50 were chosen thereby excluding the last ranks where a very low amount to no read counts were determined. In addition, the analysis does not account for the differences in abundance. For example, AAV5 was the most abundant wild-type in the NPCs with a fair amount of distance to the second rank, yet the graphs do not contain this vital information. Nonetheless, this particular type of data depiction provides a mechanistic explanation for the unique performance of the tested chimeras.

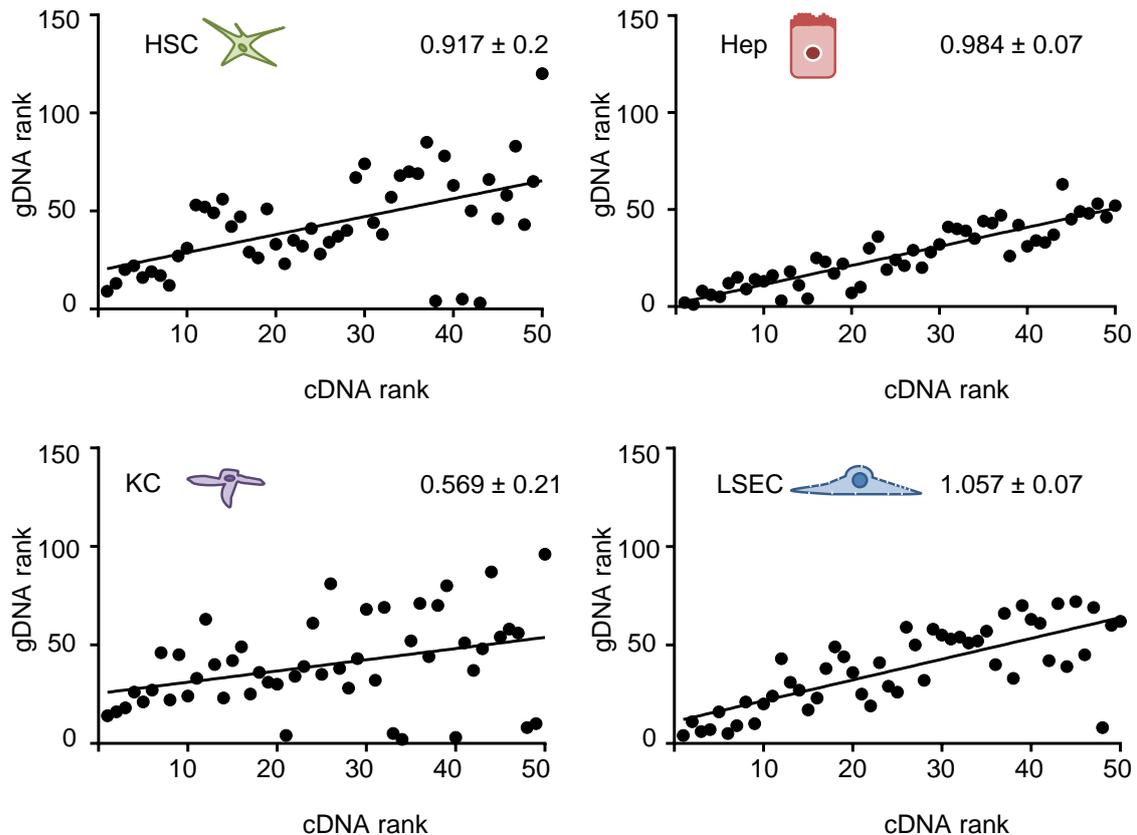


FIGURE 2-19: COMPARISON OF THE CDNA WITH THE GDNA RANK

The correlation of the cDNA rank with the gDNA rank was investigated. The top 50 in the cDNA ranking were studied more closely (plotted on the X-axis). The accompanying gDNA ranks of the single variants were plotted on the Y-axis. Each dot represents the cDNA and associated gDNA rank of a single variant. A linear regression was fit into the data points and the slope with its standard deviation was determined (indicated by the numbers at the top right of each graph).

The slopes for hepatocytes and LSECs were close to 1.0, namely 0.984 and 1.057, respectively (Figure 2-19). In contrast, stellate cells have a linear regression with a slope of 0.917 which becomes even flatter in Kupffer cells (0.569). This confirms the previous observation that hepatocytes and also LSECs have a coupling of gDNA presence to RNA expression. Whereas Kupffer cells and to a lesser extent, stellate cells, are uncoupled and high gDNA amounts do not automatically translate into high transgene expression.

At last, the amino acid sequences of the chimeras selected in stellate cells and used in the NGS screen were studied (Figure 6-5). Curiously, on the gDNA level, these variants showed an enrichment of AAV1 and AAV6 at the C-terminus. This was strikingly different on the cDNA level where the C-terminus was predominantly derived from AAV7.

So far, the distribution of the different variants within a specific cell type was analyzed. By application of an alternative normalization strategy, the distribution of a given variant among the different cell types can be analyzed, too. Here, the normalization strategy considers the vector genomes present in one cell and, therefore, the analysis compares the efficiency of a variant per cell and does not account for the complete cell population.

The biodistribution of the top 15 lead candidates on cDNA level for HSC are depicted in Figure 2-20 A. Of note, the variants originating from an unrelated screen (JEA) occurred in most cell types thereby exhibiting a broad transduction profile (e.g. 3D4 and 2D4). In addition, enrichment in a certain cell type can be observed, such as in Kupffer cells/HSC (6\_P4 and 1\_A2) or in hepatocytes (B1 and 8\_A1). Next, the biodistribution of *in vivo* HSC selected chimeras was studied further to see if enrichment for HSC can be observed. As observed before, some of the variants showed a broad transduction profile (e.g. 4N6 and 4N7) (Figure 2-20 B). Interestingly, these chimeras were derived from AAV7/AAVrh10 and likewise, these parental wild-types exhibited a broad distribution among the different cell types, too. Most importantly, there were variants, such as 3-5, 4N1 and 4b-10, which displayed off-targeting from hepatocytes and were more abundant in Kupffer cells/HSC. Of note, whereas 3-5 and 4b-10 had a higher preference for Kupffer cells than for HSC, the chimera 4N1 showed an increased tropism for HSC than for Kupffer cells. The corresponding wild-types, AAV1 and AAV6, showed a comparable biodistribution. Curiously, vectors that targeted HSC expressed in a highly similar efficiency in Kupffer cells as well.

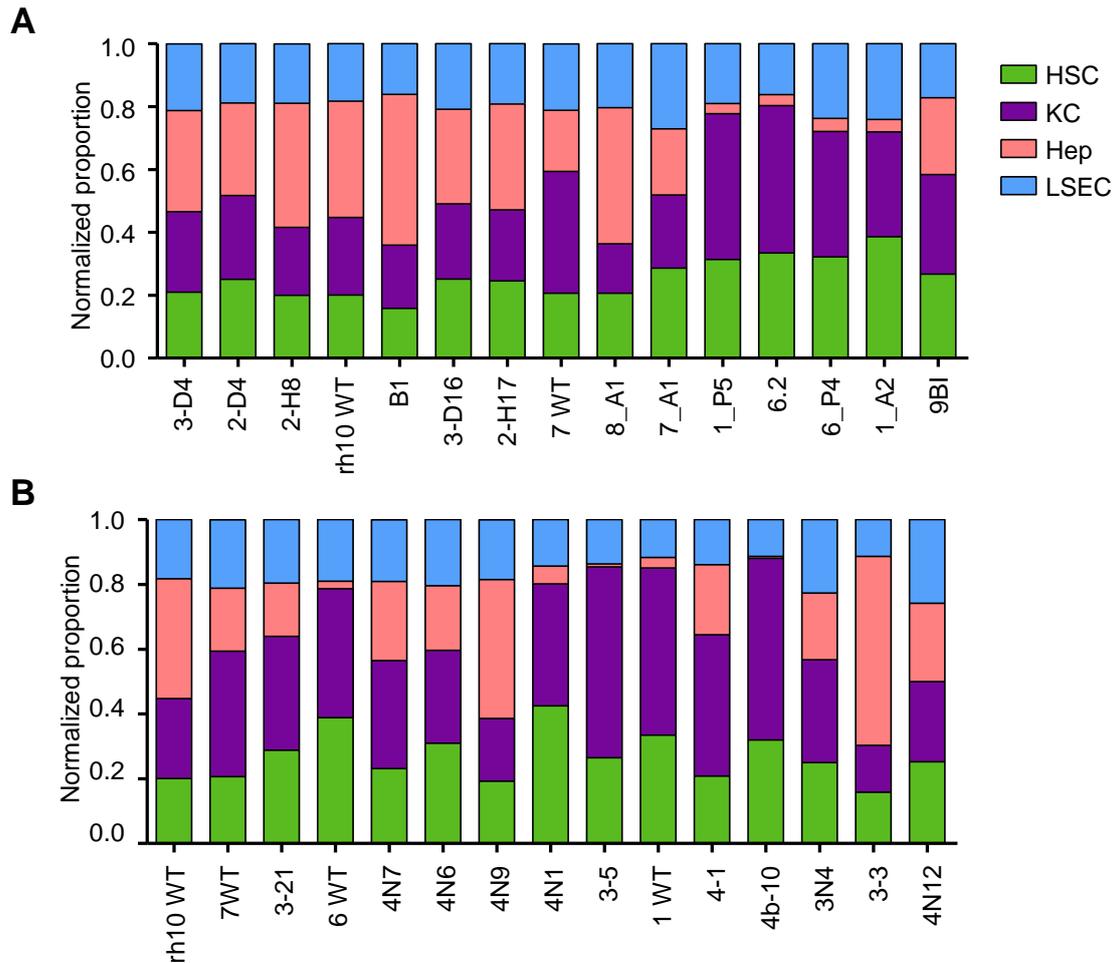


FIGURE 2-20: BIODISTRIBUTION WITHIN THE LIVER

(A) The biodistribution of the top 15 lead candidates on cDNA level in stellate cells is depicted as a stacked bar graph. The four cell types analyzed were stellate cells (HSC), Kupffer cells (KC), hepatocytes (Hep) and LSECs. (B) The selected variants along with the respective wild-type controls were sorted according to their cDNA rank. More information about the exact rank is provided in Figure 6-5.

The variants which will be considered for future *in vivo* validation studies are summarized in Figure 2-21 A. For hepatocytes, the chimeric capsid AAV DJ turned out to be highly specific and was the best candidate identified. In previous *in vivo* validation experiments AAV 3-5 mediated efficient transduction of both, Kupffer cells and HSC (Figure 2-14, Figure 2-15), and this was recapitulated in the biodistribution analysis. Additionally, the chimera 4N1 is of interest due to its preference for HSC over Kupffer cells. The corresponding wild-types were included as appropriate controls. For LSECs the overall efficiency was rather low and a vector with higher specificity for this cell type could not be identified. Another benefit from selection is the increase in production efficiency (Figure 2-12).

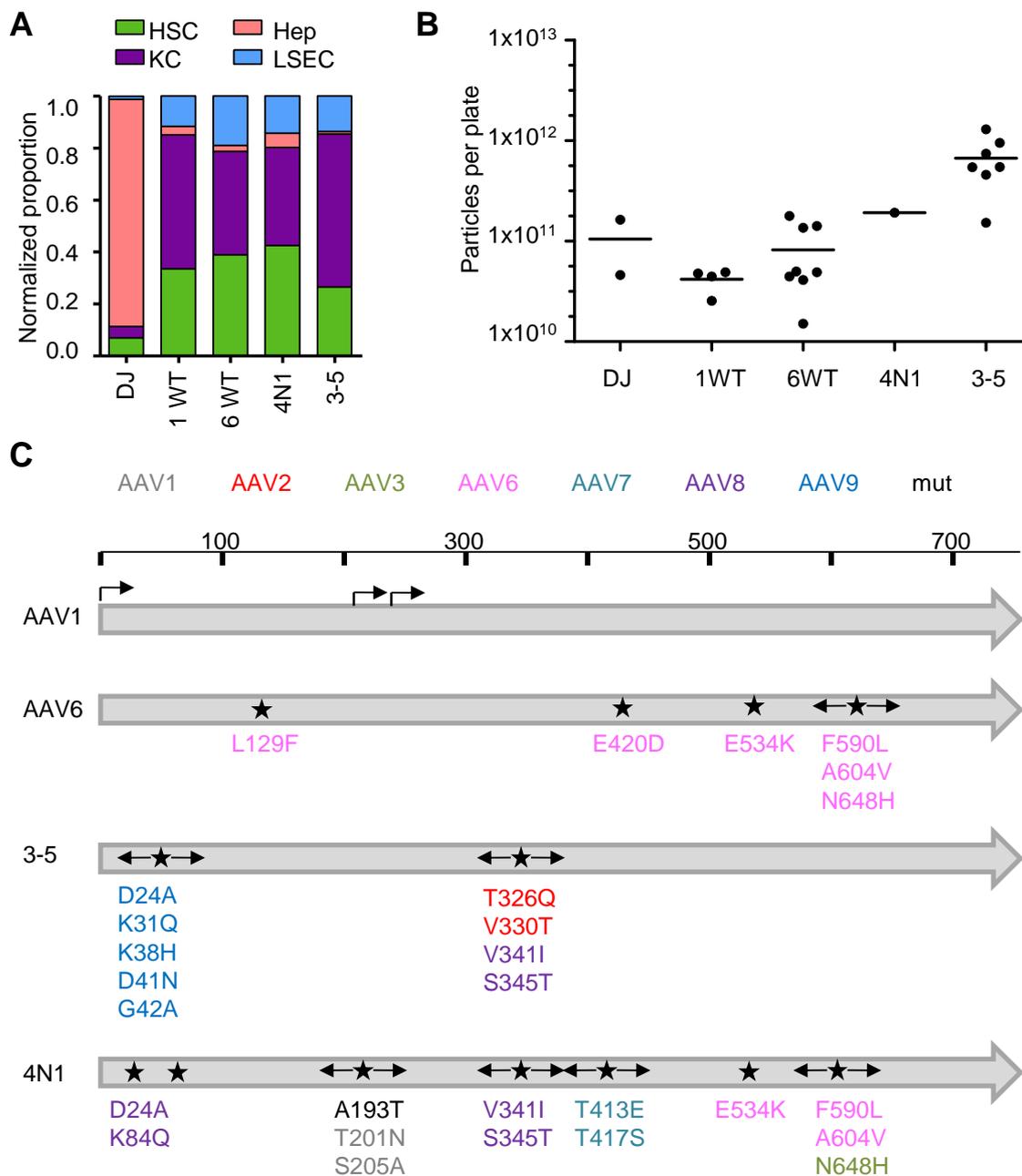


FIGURE 2-21: DETAILS ON THE MOST PROMISING VARIANTS

The NGS screen resulted in the identification of interesting variants for further validation. (A) The biodistribution of the variants is shown in a side-by-side comparison. (B) Likewise, their production efficiency was compared based on the particle amount produced per plate during transfection. The data originate from a database in the Grimm lab collecting the details of all virus productions. (C) The two lead candidates 3-5 and 4N1 share the highest homology on the protein level with AAV1. For a better comparison AAV6 is included as well. The approximate amino acid positions differing from AAV1 are indicated within the grey bars by the stars and arrows. The specific changes are noted below using the color code indicated above AAV1. The arrows in the bar representing AAV1 indicate the approximate translation start sites for the different VPs.

The AAV variant 3-5 has a difference of nine amino acids compared to AAV1 and produces 16 times better, with an average particle yield per plate of  $4.17 \times 10^{10}$  (AAV1) and  $6.71 \times 10^{11}$  (3-5), respectively (Figure 2-21 B, C). Of note, five of the amino acids are located near to the N-terminus and are buried within the capsid. The second candidate 4N1 is more diverse and has a difference of 13 different amino acids compared to AAV1.

Finally, the aim was the identification of promising AAV variants for the use as cell-type specific vectors. The *in vivo* HSC selection yielded promising variants (3-5 and 4N1) which showed hepatocyte detargeting and a preference for HSC and Kupffer cells. The NGS bulk validation provided many interesting observations and allowed to dissect the tropisms and efficiencies of the different AAV variants tested.

#### 2.3.2.4. ENHANCING SPECIFICITY BY VECTOR DESIGN

The NGS screen has yielded interesting variants for further validation, but most vectors still exhibited off-targeting. While the selection process had reduced the targeting to hepatocytes, vectors entering stellate cells also transduced Kupffer cells. For this reason, we reckoned that the vector cassette itself needs to be improved as well in order to obtain the required specificity. As described below, this was achieved by pursuing two complementary strategies.

The first element that we aimed to improve is the promoter that drives transgene expression. Cell-type specific promoters have the advantage that expression is limited to the cell type where the promoter is active. For stellate cells, the periostin promoter was acquired during the collaboration with Holger Willenbring. Likewise, collaboration with Antonia Follenzi (Università del Piemonte Orientale, Italy) provided the cd11b and VEC promoters that are active in Kupffer cells or LSECs, respectively. All promoters were cloned into an AAV vector encoding an eGFP reporter (section 4.2.3.4). Single-stranded vectors were chosen due to the size of the promoters that are incompatible with self-complementary vectors. However, the use of conventional single-stranded AAV genomes comes at the cost of a slower onset of transgene expression.

Therefore, an alternative strategy was pursued in parallel where a self-complementary vector with a ubiquitous CMV promoter was adapted. Here, microRNA (miRNA) binding sites (bs) were incorporated into the 3'UTR of the eGFP reporter. To achieve detargeting from hepatocytes, four miR-122 bs were incorporated. Additionally, a combinatorial vector cassette was generated where next to the miR-122 bs, two bs for miR142-3p were inserted. This miRNA is expressed in cells of hematopoietic origin and should

mediate off-targeting from between Kupffer cells (Merlin et al. 2017). An overview over the different constructs and the expected expression profile is depicted in Figure 2-22 A. Next, the vector cassettes were tested for functionality *in vitro*. The chimeric capsid 3-5 was chosen for production of crude lysates and different cell lines were transduced. Two days later, the amount of GFP-positive cells was determined by FACS analysis. The cell lines were chosen to model the different cell types found in the liver. Huh7 cells are derived from a hepatic carcinoma and served as a surrogate for hepatocytes. As expected, reporter expression was abolished from constructs bearing the miR-122 bs (Figure 2-22 B). Similarly, the miR142-3pbs diminished expression in RAW264.7 cells which are derived from macrophages. Thereby, detargeting from Kupffer cells and hepatocytes should be achievable. The other cells included were LX-2 which originated from human stellate cells and human umbilical vein endothelial cells (HUVEC) which model endothelial cells, *i.e.*, LSECs. The miRNA bs tested had no effect on transgene expression in these two cell lines.

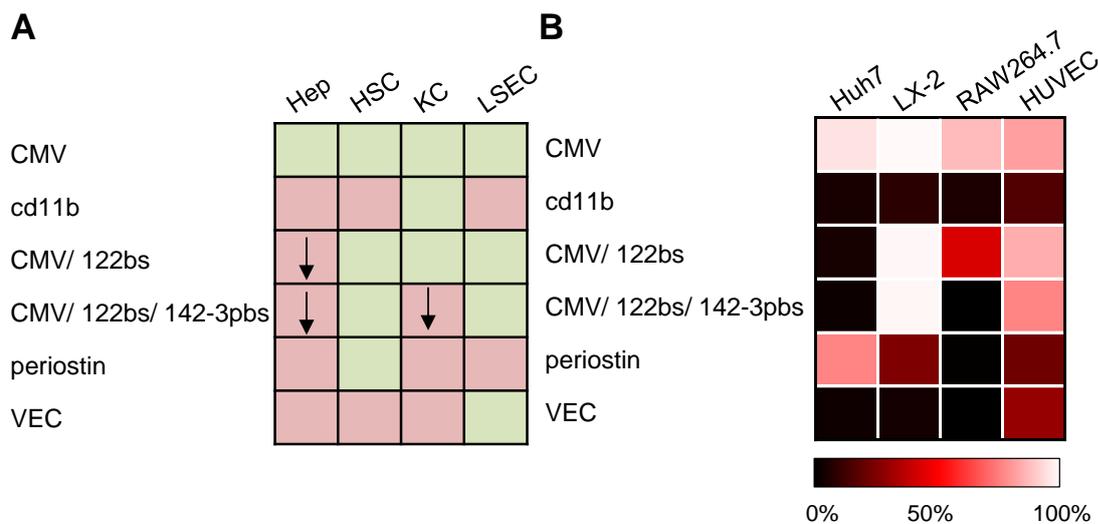


FIGURE 2-22: ADAPTATION OF THE VECTOR CASSETTE

Increase of vector cassette specificity by using different promoters and incorporating miRNA binding sites. (A) Scheme depicting the expected performance of the listed vector elements (left) in different liver cell subtypes (top). Successful transgene expression is colored green and inactive elements are shown in red. The arrows indicate the expected miRNA-dependent down-regulation of transgene expression. (B) All elements were packaged into the chimera 3-5 in small scale. The resulting crude lysates were used to transduce a variety of cell lines (indicated at the top) and the amount of reporter (eGFP) positive cells was determined 48 h later by FACS. The mean of  $n=3$  is depicted as a heat map ranging from no infection (0%) to 100%. The color scale is indicated below. Hep, hepatocytes; HSC, hepatic stellate cells; KC, Kupffer cell; bs, miRNA binding site

The cell-type specific promoters were not as efficient in mediating specificity as the miRNA bs. The cd11b promoter was acquired for Kupffer cells and is inactive in most cell types, albeit HUVECs to a small extent. Periostin was chosen for stellate cells and the promoter is active in all cell lines except for RAW264.7. In contrast, the VEC promoter specifically conveyed expression in HUVECs and thereby limited transgene expression to the endothelial cell line.

Overall, the generated vector cassettes were capable of restricting the transgene expression to specific cell types by application of miRNA bs and the VEC promoter. The results from the *in vitro* experiment are promising for future *in vivo* validations.

---

## 2.4.OPTIMIZATION OF THE SELECTION PROCESS

---

While the selection was performed, ideas for improvement arose to advance future selections. When many samples are processed in parallel, there is a risk of cross-contamination during PCR rescues. Similarly to the abovementioned NGS approach, a novel barcoding strategy was thus implemented to increase the stringency during selection and to prevent possible cross-contaminations. Furthermore, it can serve as an indicator of enrichment during selection. Additionally, an alternative sequencing approach by SMRT sequencing (PacBio) was assessed.

---

### 2.4.1. BARCODE MEDIATED PCR RESCUE

---

In order to increase the stringency of the PCR rescue during selection, a novel barcoding strategy should be implemented. Therefore, the replication-competent backbone used for the selection of AAVs required adjustments for the insertion of the barcodes (Figure 2-23 A). Therefore, an additional restriction site was introduced after the 3' end of *cap*. The NdeI restriction site was chosen as it was neither present in the acceptor backbone nor in the 10 serotypes used in this study and was transferred by oligo annealing between the Ascl site and the Spel site (Figure 2-23 A, section 4.2.3.3). Thereby two positions were created where the barcodes can be inserted. Eventually, a reverse primer covering the barcode sequence can be used for a specific PCR rescue. The idea was to tag individual libraries by use of one barcode at the Ascl/ NdeI interface (see section 2.4.2). Additionally, the second barcode at the NdeI/ Spel interface could be then used to separate the PCR rescues between single selection rounds or organs by introduction of a new barcode sequence for every sample.

Initially, a proof-of-principle for the selectivity of the barcodes was performed in an *in vitro* setting (section 4.2.8.1). Therefore, three different libraries were tagged with a barcode, namely, libraries composed of serotypes 2, 8 and 9 (289); 1, 7-9 and rh10 (1789rh10); and 1-rh10 with shuffling optimized versions of AAV4 and 5 (1-100). The individual libraries were produced in small scale as crude lysates. Next, HEK293T cells were infected for subsequent PCR rescue. The libraries were either used separately (conditions 1-3 in Figure 2-23 B) or as a mixture where each library is present at the same volume (condition 4 in Figure 2-23 B). The next day, the cells were harvested and lysed by Proteinase K digestion. A PCR was performed on all four samples (single and mixed infections) with either library-specific primers (lib) or a general primer (G) common in all libraries.

As expected, the general primer could rescue *cap* in all samples (Figure 2-23 C). The library-specific primers cover the corresponding unique barcode and, hence, sample-specific amplification was expected. Indeed, this specificity was observed as in the single-infection samples, only the respective barcodes serving as reverse primer binding site could perform the rescue (highlighted by arrows). Next, the samples were sequenced to confirm the identity of the rescued *cap* variants. Initially, PCRs using the common primer “G” were performed on conditions 1-3 (Figure 2-23 B) followed by subcloning and sequencing. The clones recovered originated from the expected library exemplified by Salanto analysis which correlated with the parents from the input library (Figure 6-7) and the presence of the expected library-specific barcode.

Next, the sample containing a mixture of libraries (condition 4) was analyzed. By using the common primer “G”, a mixture of clones can be isolated evidenced by the numbers and the type assignment (Figure 2-23 D). Clones from all three input libraries were observed. However, a majority originated from the 289 library (6 out of 10). If the library-specific barcodes were used for the PCR step, only clones originating from the respective library were recovered, once again confirming the specificity.

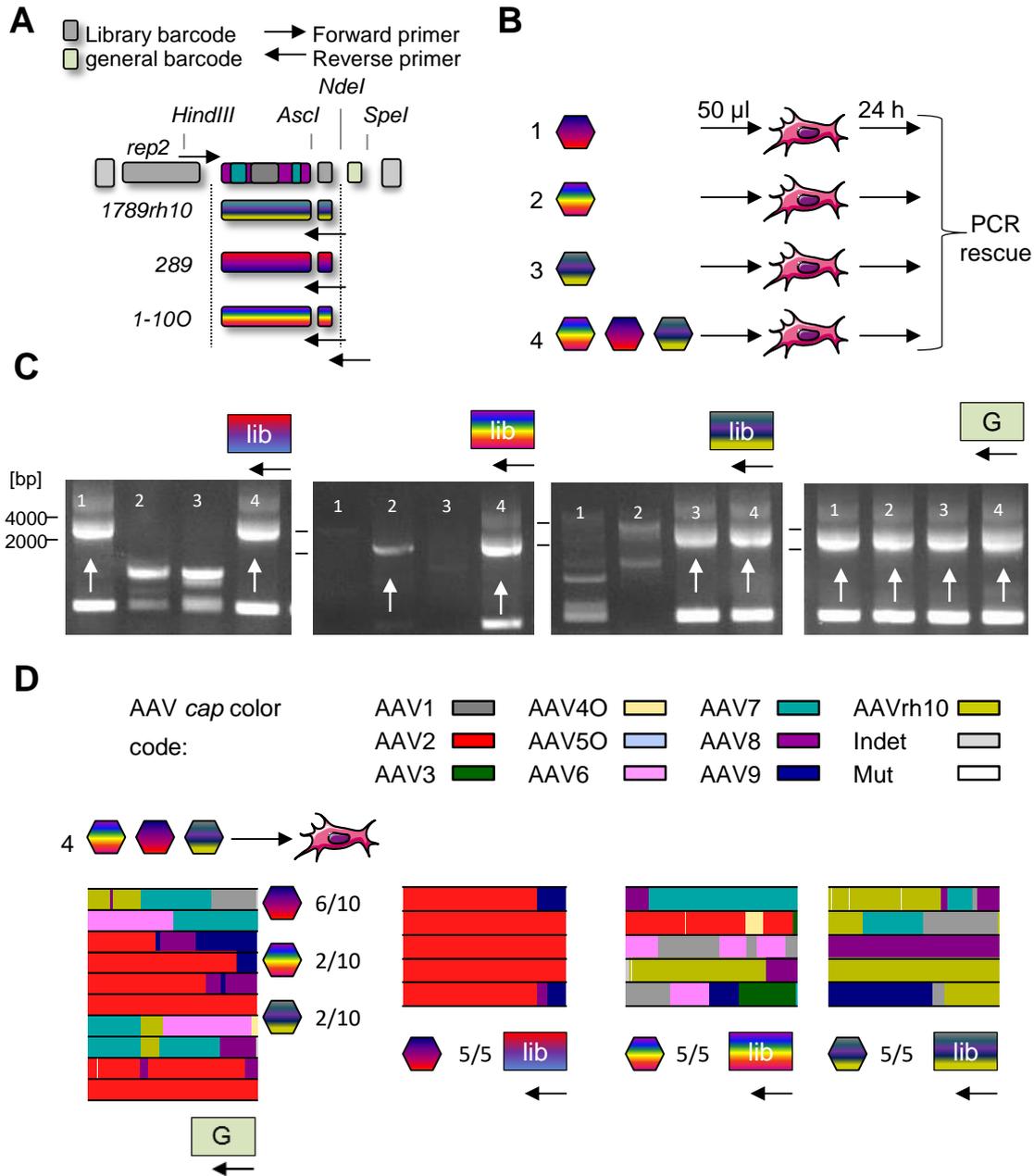


FIGURE 2-23: BARCODED PCR RESCUE

(A) The altered backbone design is depicted. The *NdeI* site was introduced between the existing *Ascl* and *SpeI* site allowing for the insertion of two different barcodes at the same time. One can be used to tag a library (lib) and the other can specify e.g. a given organ or selection round. For a proof-of-principle, three different libraries were tagged with a unique lib barcode and, additionally, a general (G) barcode common for all was inserted. (B) Crude lysates were produced and 50  $\mu$ l were used to infect HEK293T cells either separately (1-3) or as a mixture (4). After 24 h, the cells were harvested and a PCR rescue was performed. (C) For the PCR, all samples were tested with different combinations of reverse primer and equal volumes were analyzed on a 1% agarose gel. Sample numbering correlates with the conditions in (B). The barcode exploited as reverse primer binding site is depicted above the gel image and expected bands are highlighted with the arrow. (D) The products from the PCR rescues for condition 4 were subcloned for further analysis. The approximately last 800 bp of *cap* were sequenced with a reverse primer and single clones were analyzed with Salanto. The color code is depicted above the image and below. Numbers indicate the identified barcodes from all clones tested. Adapted from (Herrmann et al. manuscript in preparation).

---

#### 2.4.2. LIBRARY TRACKING IN VIVO

---

During the course of the study, a 1-100 library was generated by DNA family shuffling and selected in stellate cells *in vivo* (section 2.3.1). This library is referred to as “original” (ori). Although enrichment of certain C-terminal domains was observed, the initial library had a certain bias towards AAV1, AAV6 and AAV7. For future selections, a more evenly distributed library representing a higher variety at the C-terminus should be beneficial. Therefore, another 1-100 library was shuffled under the same conditions originally used. This library (“new”) again showed a large proportion of AAV1 and AAV6, exemplified by the grey and pink color at the 3′ end in Figure 2-24 A. In order to circumvent this bias, the shuffling was repeated once more but this time, AAV1 and AAV6 were counted as one serotype in the DNase I digestion. This resulted in a third library (“red”) with a reduced amount of these serotypes and an increased proportion of AAVrh10.

To test the hypothesis that a mixture of AAV libraries could be beneficial for selection due to the presence of a higher C-terminal diversity, an *in vivo* pilot study was performed to obtain proof-of-principle. Therefore, four shuffled libraries (289, 1-100 ori/red/new) were tagged with a unique barcode and produced separately by cesium chloride gradient density centrifugation (section 4.2.8.2). After titration, equal virus amounts were mixed and injected intravenously into two female mice. Two days later, they were sacrificed and selected organs were extracted followed by DNA isolation (Figure 2-24 A). For the PCR rescue, a common primer was used for the amplification of *cap*. The recovered variants were subcloned and single clones were sequenced from the 3′ end. Then, the present barcodes were identified, counted and the relative percentage was determined (Figure 2-24 B). For all organs from the first selection round, 96 sequences were recovered. A second round of selection was performed with the libraries recovered from heart, liver and kidney and here, 48 clones were sequenced.

After the first round of selection, the distribution among organs looked similar for the 1-100 libraries. The 289 library was less abundant in most organs except for kidney where it occupied approximately a quarter of the reads. However, after another round of selection, changes could be observed. The 289 library was now less abundant in the kidney, comparable to the other organs after the first round of selection. Instead, an increase of the libraries 1-100 “ori” and “new” could be observed.

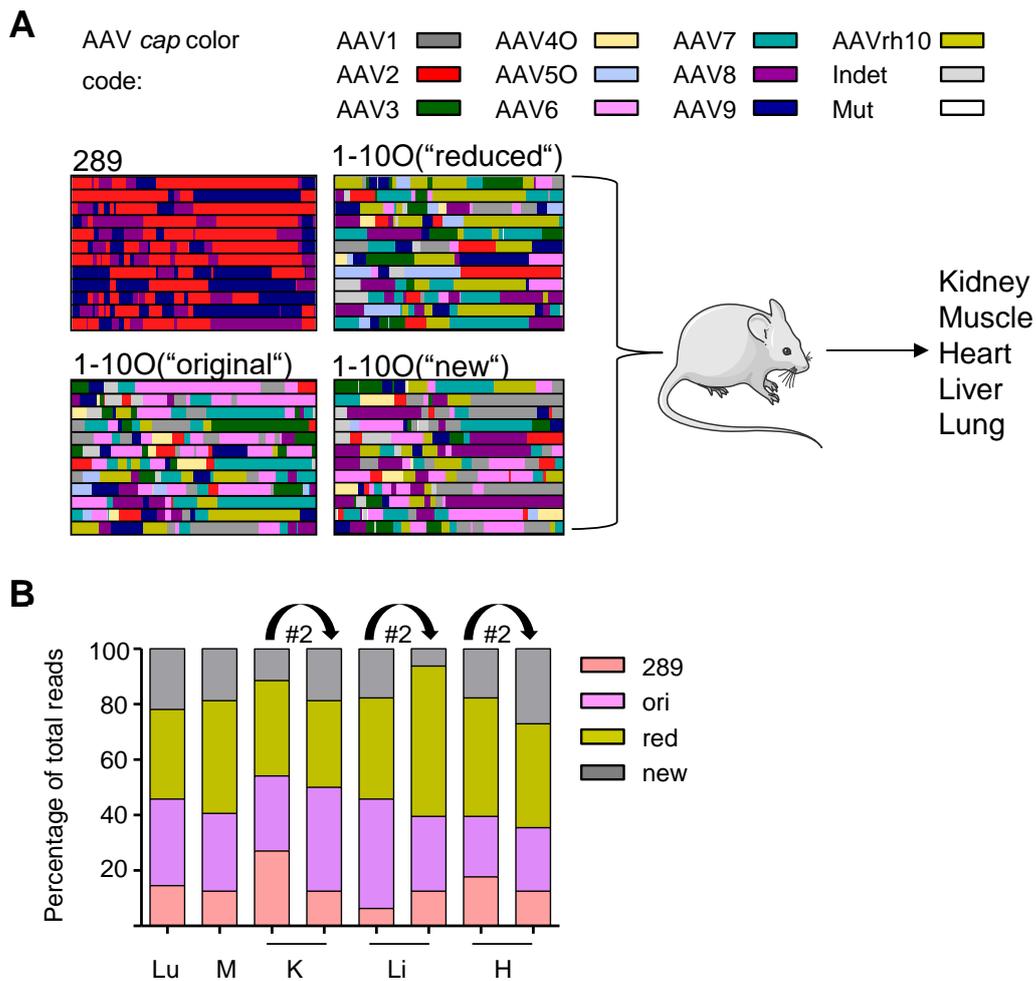


FIGURE 2-24: LIBRARY TRACKING IN VIVO

Validation of the library barcoding strategy in mice *in vivo*. (A) Scheme of the experimental setting. Four different libraries were tagged with a specific barcode and a mixture was injected into mice. After 48 h, the indicated organs were isolated and processed for DNA isolation. The color code represents the parental serotypes of the libraries. Full-length *cap* genes are shown on the nucleotide level. (B) The isolated DNA was used for PCR rescues with a common primer and the last 800 bp of the *cap* gene were sequenced. From the first selection round, 96 clones were analyzed and 48 from the second. The percentage of individual barcodes is depicted as a stacked bar graph. Lu, lung; M, muscle; K, kidney; Li, liver; H, heart; #2, round 2. The cartoon of the mouse was taken from the servier medical art collection (<https://smart.servier.com/>). Adapted from (Herrmann et al. manuscript in preparation).

In contrast, these two libraries became less abundant in the liver upon another selection round. Instead, an enrichment of the 1-100 “red” library and the 289 library was observed. The liver is the only organ where the 289 library was further enriched upon selection. In the heart, an enrichment of the 1-100 “new” library can be observed at the cost of the “red” library. Overall, differences in the distribution of individual libraries were observed indicative of enrichments.

In summary, we showed that the barcoding approach was capable to mediate a more stringent PCR rescue, even when a mixture of libraries was present. Furthermore, future selection studies could benefit from using a mixture of libraries instead of an individual one, and the presented strategy could also be applied to track the library origin.

---

#### 2.4.3. MONITORING LIBRARY ENRICHMENT BY SMRT SEQUENCING

---

Last but not least, yet another aspect that needs to be improved during generation and selection of AAV capsid libraries is the sequencing method. To date, the standard procedure is Sanger sequencing of single clones. This method is cheap, simple and fast, but, it is also very low-throughput especially since three sequencing reactions are required to cover the complete *cap* gene (section 2.1.1). Thus, only a fraction of the whole clone population within a capsid library is covered. PacBio/SMRT is an alternative sequencing method that bears the seminal advantage that it reads on average 10 to 15 kb with a coverage of  $3.5\text{--}7.5 \times 10^4$  reads (Rhoads and Au 2015).

During the selection in stellate cells performed in this work (section 2.3.1), a presumable shift from AAV1 and AAV6 towards AAV7 was observed by Sanger sequencing. To independently validate this result, the PCR rescue product obtained after the third round was sent for PacBio/SMRT sequencing. A more detailed analysis of the data is found in the Bachelor thesis of Niklas Beumer (also performed in the Grimm laboratory), who kindly isolated the sequences of 200 random clones. These sequences were aligned and assigned, using another pivotal feature of the Salanto suite. Briefly, at each nucleotide position the proportion of each reference sequence, undetermined (indet) and mutated (mut) was determined. This finally reflects the likeliness of a certain serotype to occur at a given position and the data was depicted as a line graph as seen in Figure 2-25. The lines for Sanger sequencing appear like a staircase (Figure 2-25 A) whereas, in stark contrast, the PacBio/SMRT-derived analysis appears much smoother (Figure 2-25 B). The reason for the latter is the difference in the number of analyzed sequences, which is approximately 10-fold higher in the case of the PacBio/SMRT strategy.

Moreover, in the analysis based on Sanger sequencing of 18 clones, an enrichment of AAV1 and AAV6 was observed at the C-terminal part. Interestingly, there seemed to be a preference of AAV1 over AAV6 at certain positions (indicated by the arrows). Similarly, this trend can be observed in the sequences derived from PacBio/SMRT sequencing. However, here, an overall high abundance of AAV7 became visible which was missed in the Sanger sequencing-based analysis.

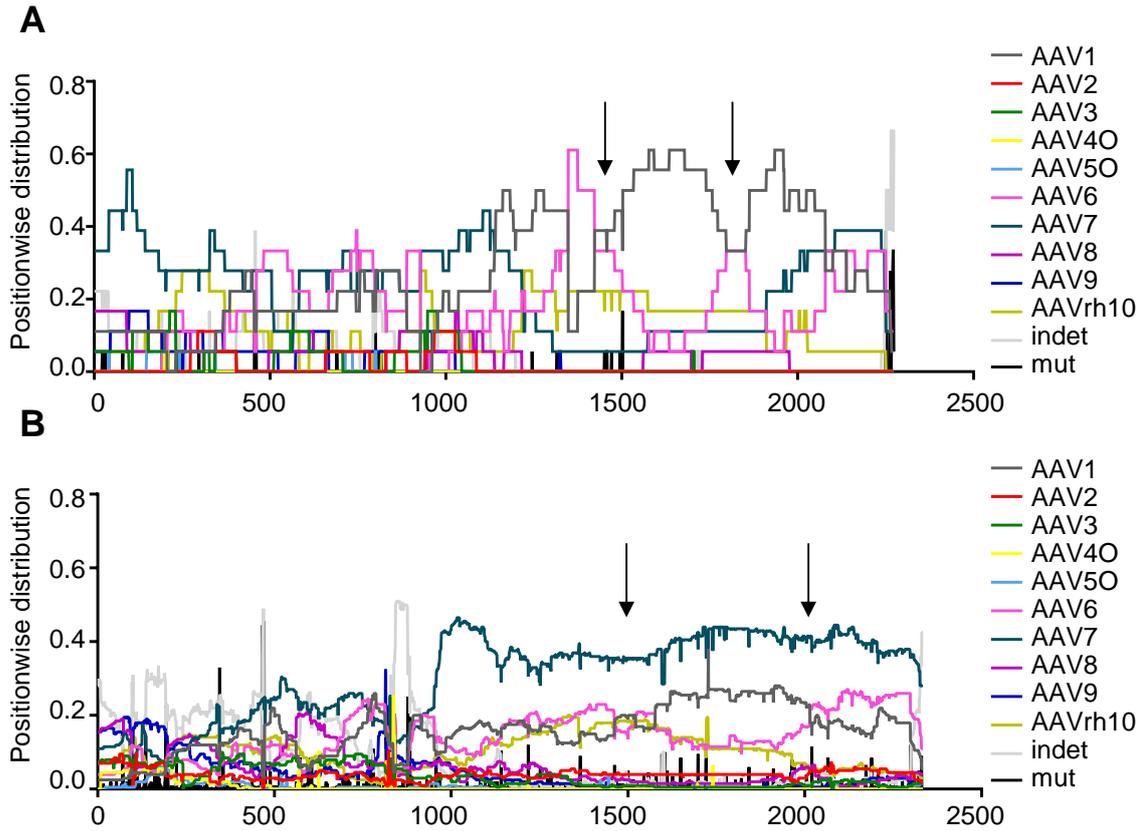


FIGURE 2-25: SEQUENCE ANALYSIS BASED ON DIFFERENT METHODS

The positionwise distribution of the parental sequences revealed a higher accuracy of PacBio/SMRT sequencing. The complete nucleotide sequence was analyzed by Salanto. The probability (Y-axis) of a given parental serotype at a certain position (X-axis) is depicted and colored according to the color scheme. The arrows indicate a preference of AAV1 over AAV6 at a certain position. (A) The analysis is based on the analysis of 18 clones by Sanger sequencing. (B) The analysis is the result of 200 sequences originating from a PacBio/SMRT sequencing run.



---

## 3. DISCUSSION

---

Over the last few decades, viral vectors based on AAVs have become increasingly popular for a variety of biological or medical applications. Their attractiveness and potential are best exemplified by the fact that the first two gene therapies approved in the Western World are based on AAV vectors (section 1.1.3). The development of clinical AAV gene therapies is additionally fostered by basic research and our resulting, steadily improving knowledge about disease initiation and progression. A major example that was in the center of the present work is the liver, a vital organ exerting multiple functions in physiology and disease that is composed of different cell types. Of these, stellate cells play an important role for the progression of liver fibrosis, a potentially fatal disease lacking options for curative treatment (section 1.2.2). Accordingly, viral vectors that are specific for these or other liver cell types would be of great value for the biomedical community. A most promising strategy to achieve this vital aim is the molecular engineering of novel viral capsids and genomes that mediate cell type-specific expression of therapeutic transgenes, ideally from a non-invasive peripheral infusion. To this end, the present work pursued two complementary goals: (i) to dissect and optimize seminal steps in the workflow for AAV capsid evolution by DNA family shuffling, and (ii) to apply this technology in order to enrich novel, liver cell type-specific AAV capsids through iterative selection in adult mice *in vivo*.

---

### 3.1. TECHNICAL IMPROVEMENTS

---

The first major aim of this work was to broaden our current understanding of the molecular processes underlying DNA family shuffling of AAV capsid genes, and to then use the newly gained knowledge to improve and simplify the protocol for future users of this powerful technology. As described above and discussed in more detail below, the steps that were investigated comprise (i) fragmentation of input capsid genes, (ii) purification of products from the re-assembly PCR, (iii) DNA barcoding as a novel means for library multiplexing and tracking, and (iv) use of PacBio/SMRT instead of Sanger sequencing for analysis of shuffled full-length capsid genes.

Initially, we studied one of the very early steps in DNA family shuffling, namely, the fragmentation of full-length *cap* genes into smaller pieces for subsequent re-assembly in a primer-less PCR. To date, the vast majority of previous studies comprising the pioneering work from Grimm and colleagues (Koerber, Jang, and Schaffer 2008; Li et al. 2008; Grimm et al. 2008) has used enzymatic fragmentation with DNase I as method of

choice. However, this procedure is difficult to standardize because the digestion efficiency is influenced by multiple parameters, including the age of the enzyme and the experience of the operator. We therefore evaluated the possibility to overcome this bottleneck by replacing the DNase I reaction with physical fragmentation by Covaris-based ultrasonication.

Our results show that when using the Covaris machine for fragmentation, only the pool with the largest fragments (800 bp) was capable of producing full-length *cap* (Figure 2-1). This confirms and extends previous data from our group, and it supports the hypothesis that DNA family shuffling is dependent on longer fragments above a certain threshold (Große 2016). Interestingly, when DNA family shuffling was modeled by another group *in silico*, it was found that small fragments of 100 bp would convey a four times higher crossover rate as compared to larger fragments of 400 bp (Maheshri et al. 2006). Concurrently, however, the overall yield would be decreased by a factor of over six-fold. Additionally, when modeling the reassembly frequency, the best result was obtained for fragments of 400 bp in size as compared to 100 bp and 200 bp fragments. Taken together, this model suggests that a minimal fragment size of 400 bp yields the best trade-off of between crossover frequency and reassembly efficiency. Of note, this former theoretical prediction is fully congruent with the new experimental data obtained in the present work, as the pool of 300 bp fragments obtained by Covaris ultrasonication failed to recover *cap* in most cases.

Furthermore, our sequencing analyses show that the C1 library created through ultrasonication was composed of fewer fragments with an overall larger fragment size, in contrast to libraries derived from DNase I digestion that contained shorter fragments with a higher abundance (Figure 2-2). Also this correlation is in line with previous data from our group (Große 2016).

Finally, we compared different conditions for DNase I digestion, in particular varying reaction times, to assess their influence on the overall efficiency of the DNA family shuffling protocol. Importantly, our data suggest that all of them worked comparably well, based on our analysis of average fragment size and abundance of fragments.

In conclusion of this part, our thorough comparison of two fragmentation methods, enzymatic or physical, provided compelling evidence that DNase I digestion is the more robust approach. This was supported by the observations that (i) all DNase I digests tested recovered *cap*, and that (ii) the reproducibility of full-length *cap* assembly is higher for DNase I (6/6 samples recovered *cap*) as compared to Covaris (only 4/9 reactions recovered *cap*). Furthermore, the robustness is illustrated by the finding that (iii) the

sequence analysis gave comparable results among individual DNase I digests, despite different appearances of the digests on the gel in terms of fragment size distributions. As a whole, our results allow us to conclude that DNase I digestion is in fact much more robust and reproducible than originally anticipated, and we thus continue to recommend this method as the preferred option for initial DNA fragmentation during AAV DNA family shuffling.

Additionally, we were able to improve the yields of the recovered full-length *cap* PCR product, which was urgently needed as roughly 1.3 µg of digested PCR product are required for standard library generation. This improvement was made possible by showing that bands generated during nested PCR steps with a size between 4 to 5 kb could be resolved into the 2.2 kb *cap* band upon digestion with appropriate restriction enzymes (Figure 2-3). Our interpretation of this finding is that the restriction sites flanking *cap* serve as a linker sequence, leading to concatamer formation and permitting resolution into single molecules through enzymatic digestion. Accordingly, the PCR purification protocol was changed from the standard gel purification to column purification, which, as shown above (Figure 2-3), resulted in a substantial increase in final yields of PCR product for subsequent cloning. This optimization is highly beneficial for future library productions as it additionally saves time, work and materials.

Next to the steps for library production, we also succeeded at optimizing the downstream selection process of shuffled AAV capsid libraries. A first improvement was the incorporation of specific barcodes into the library backbones, fueled by our motivation to increase the stringency of the PCR rescue for secondary library generation. This was urgently implied by observations in our group and many others working with AAV libraries (D. Grimm, personal communication) of occasional contamination with unrelated *cap* sequences during PCR reactions. To decrease this risk, we came up with the idea that incorporation of a barcode specific for e.g. a given target organ or a particular selection round can be used as a reverse binding primer site during PCR rescue. Because this barcode can be designed to be unique for each scenario, it will serve as a unanimous identifier for a specific library and will thus largely reduce or eliminate the risk of cross-contamination. This expectation was clearly validated by our results from the *in vitro* experiments, proving that it is indeed possible to deliberately isolate individual libraries or clones from a mixture (Figure 2-23).

Based on these encouraging first data, we can readily envision a second benefit of our new library barcoding strategy next to the reduction of contaminations, namely, the ability to use a mixture of different libraries for selection instead of only one (Figure 2-24).

Thereby, the complexity and variability of the input library, or rather the pool thereof, can be expanded even further, while the barcodes can be exploited to track the origin of each individual library within the pool at any time. This ability to multiplex and deconvolve libraries, respectively, is essential because if enrichment of a certain library in a pool is observed early during the selection, this will indicate that candidates originating from this library are most relevant. Hence, subsequent selection rounds can focus on this subset of libraries and thus continuously increase the stringency and chances of overall success. This strategy of co-infecting with multiple different libraries at once was applied in the past (albeit lacking the additional benefit of the barcodes) and yielded the promising variant EP1.9 (section 1.1.4.2, (Asuri et al. 2012)). In this previous example, libraries created by different technologies were chosen, *i.e.* shuffling, peptide insertion and error-prone PCR, which ultimately enabled the identification of the library origin. As an important advance, our novel barcoding strategy will now also allow to mix and disentangle several shuffled libraries, whose identification would otherwise be difficult, especially when libraries with a similar parental composition are used.

Finally, we have made first attempts to improve yet another critical step in the AAV DNA family shuffling protocol, which is the sequencing-based analysis of the starting libraries or of selected clones. This was motivated by the fact that the current standard - Sanger sequencing - requires three different sequencing reactions to cover the complete *cap* gene, the results of which are then manually assembled into the full-length sequence. This labor-, time- and cost-intensive work load limits the amount of clones which can be analyzed in parallel and therefore necessitates the implementation of alternative, more powerful and high-throughput-compatible sequencing methods. One such option could be Illumina sequencing, but the maximum read length of this technology is limited to approximately 250 bp (Rhoads and Au 2015), which makes it unsuitable for sequencing of full-length shuffled AAV genomes in a single run due to their size of 2.2 kb (*i.e.*, in principle, the same problem as with Sanger sequencing). Fortunately, in recent years, the SMRT technology of Pacific Biosciences (thus also called PacBio sequencing) has become affordable and thus very attractive as a novel means for sequencing of shuffled AAVs. The latter is owing to the facts that PacBio/SMRT sequencing can easily cover the entire *cap* gene in one run, and that the read depth is still in a range (typically >15.000) that is orders of magnitude above that of conventional Sanger sequencing. Here, the power and potential of this new approach were exemplified and the sample after the third selection round was sequenced. Sanger sequencing predicted enrichment for AAV1 and AAV6 but unexpectedly, the picture changed after the fourth round where enrichment of AAV7 was found in three independent mice (Figure 6-4). The complete PacBio

sequencing run encompassed approximately 13600 reads providing approximately  $10^3$  more reads than Sanger sequencing. Analysis of 200 clones retrieved from PacBio sequencing revealed that AAV7 was already the most abundant serotype after the third round and the same interesting enrichment of AAV6 over AAV1 was identified seen in Sanger sequencing (Figure 2-25). For future studies this sequencing technology is highly promising as much more clones can be covered and, ideally, the most abundant variants can be thereby identified.

### 3.2. THE SHUFFLING OF AAP IS NOT DETRIMENTAL

When AAP was first reported in 2010, the authors immediately raised the concern that during shuffling of *cap*, the overlapping AAP frame will be shuffled along which might in turn be detrimental for production of viral libraries and selection (Sonntag, Schmidt, and Kleinschmidt 2010). In earlier studies the fact was exploited that VP3 alone can assemble particles as long as AAP is present. By using VP3 only expression constructs the AAP frame is not complete and as a consequence particle assembly is diminished. In contrast, the experimental system established by our group is in a more natural context as the whole VP frame is used where AAP was endogenously knocked out (Große 2016). This allowed us to assess the functionality of AAP - wild-type or shuffled - in a *trans*-complementation assay. Briefly, endogenous AAP was knocked out in AAV helper plasmids of various serotypes so that during production of recombinant particles, either stuffer DNA or AAP could be supplemented *in trans*. Subsequently, expression of a vector-encoded GFP reporter in cells transduced with the resulting crude rAAV lysates was monitored as a surrogate marker of infectious particle production. However, this assay is indirect, and neither AAP expression nor the generation of assembled AAV particles has been confirmed more directly on the protein level to date. During this study, these critical gaps were filled by, firstly, confirming the functionality of the AAP expression plasmids by Western blotting. As hoped for, this analysis showed the complete absence of residual AAP expression in case of the knockout AAP versions, thereby verifying our combinatorial strategy to disrupt the AAP start codon and to concurrently introduce a stop codon (Figure 2-4 A). Intriguingly, the different AAPs ran at varying sizes in the gel although a similar size would be expected based on their sequences. Of note, while this study was conducted, another group presented a similar system and our observation is in accordance with their results (Earley et al. 2017). Taken together, this implies that AAP may be post-translationally modified in a serotype-specific manner, which is an interesting hypothesis for future work.

Secondly, in this study, we analyzed the outcome of the *trans*-complementation assay more directly on the AAV protein/capsid level as well. Therefore, native dot blots using capsid-specific antibodies were performed under non-denaturing conditions, in order to detect assembled AAV particles in the presence or absence of AAP. As an important control, the samples were additionally heated to promote capsid disassembly. This confirmed that the impaired transduction originates from a lack of assembled particles (Figure 2-4 B). Curiously, in the case of the AAV1, 4 and 5 AAP knockout mutants, the ADK antibodies detected a structure which failed to reveal the B1 epitope (hidden inside the assembled particle) upon heating. A possible explanation is that the VP subunits form stable aggregates in the absence of AAP that are recognized by the ADK antibodies and that are largely resistant to treatment with heat.

Next, the dependency of a given AAV variant on its own cognate AAP was assessed by testing different AAP variants. Therefore, an experimental array was designed where each AAV variant was rescued with all different wild-type AAPs, followed by measurement of transduction activity (Figure 2-5 B). A major finding was that most of the serotypes tested highly depend on AAP. Notable exceptions were the AAV4 and AAV5 mutants that maintained residual transduction activity even in the absence of AAP, which might reflect the aforementioned structures detected in the dot blot. Furthermore, a high degree of interchangeability was observed with most AAPs, which is in agreement with, and substantially expands, previously published data (Sonntag et al. 2011). This follow-up study reported the assembly of AAV1, AAV2, AAV8 and AAV9 VP3-only particles through *trans*-complementation with AAP2.

Additionally, AAP4 was only able to rescue its own knockout and the one derived from AAV5. Likewise, AAP5 failed to rescue a variety of AAP knockout mutants derived from multiple AAV serotypes, expanding results of a recent study that used VP3-only particles (Earley et al. 2017). These two AAPs, which are inefficient with most serotypes, have the most divergent sequence (Figure 2-5 C), prompting the conclusion that the compatibility of capsid variants with (exogenous) AAP is determined by the degree of homology. In this respect, it is curious that the AAV5 mutant tolerated a variety of other AAPs to mediate an intermediate rescue.

Another interesting observation was made for AAV3, as some AAPs even increased the rescue of the AAP3 knockout mutant over wild-type level. This may indicate a stronger AAP dependency of AAV3 and hence a beneficial effect of AAP supplementation during production. Furthermore, the mutants derived from AAV7 and AAVrh10 were also unique by displaying an intermediate rescue efficiency as compared to the other AAP

knockouts. In conclusion, we found that most of the wild-type AAPs are readily interchangeable but there are exceptions that are worthy of further investigation.

Motivated by these results, we next investigated shuffled AAPs and started with AAP from AAV DJ, denoted as AAP DJ, as this capsid was one of the first shuffled chimeras reported and was available in the laboratory (Grimm et al. 2008). Because it is composed of serotypes AAV2, 8 and 9, we tested the ability of AAP DJ to rescue these three parental capsids. Indeed, AAP DJ potently rescued the corresponding parental AAP knockout mutants to wild-type levels and, interestingly, a more efficient rescue of the AAV2 mutant was mediated in comparison to AAP8 and AAP9 (Figure 2-6 C). Akin to AAV2, it potently rescued AAV3 beyond wild-type levels as well (Figure 2-6 D) but in the Western blot two bands appeared for AAPDJ in the Western blot, similar to AAP8 (Figure 2-6 B). Intriguingly, these observations imply that this chimeric AAP has acquired characteristics from its multiple different parents.

When our laboratory has recently studied the localization of these different AAPs, we found AAP2 to be present in the nucleolei, whereas AAP8 and AAP9 were situated in the nucleoplasm (Grosse et al. 2017). AAPDJ was observed in the nucleoplasm as well and therefore resembled two of its three parents (Herrmann, Große, et al. 2018). This further exemplifies the chimeric functionality that AAV/P DJ has acquired during shuffling and molecular evolution.

Besides, these different AAP localization patterns that we and others (Grosse et al. 2017) (Earley et al. 2017) observed fuel questions about the exact mechanism of AAP during capsid assembly. For instance, it is known that AAV2 assembly occurs in the nucleolei but some AAPs are actually excluded from this compartment based on our data and could still rescue the AAV2 AAP knockout mutant upon *trans*-complementation.

AAV DJ originally resulted from a combination of positive and negative selection pressure (Grimm et al. 2008), leaving open the possibility that shuffling of AAP might still be detrimental during the early steps of library generation by DNA shuffling. Therefore, we next analyzed additional shuffled AAPs that we arbitrarily selected from five different unselected libraries. When juxtaposed with the AAV2 mutant, 54 out of 60 (90%) shuffled AAPs were able to rescue the AAP2 knockout above the 5% threshold level, with most of them restoring the transduction levels back to wild-type (Figure 2-7 B). The six AAP clones initially classified as dysfunctional could rescue at least one of their wild-type parents. Of note, three more AAP variants (19, 39 and 51) were tested with their respective wild-type AAP knockout mutants as well as they barely passed the 5% threshold in the original assay with AAV2 and for these, robust rescue could be observed

with non-AAV2 serotypes (Table 6-1). In conclusion, all of the 60 chimeric AAPs tested were capable of promoting assembly of at least one of the AAP knockout versions.

The rescue efficiencies of chimeric AAPs varied depending on the parental wild-type mutant rescued, which led us to the question whether AAP supplementation would result in increased rAAV particle yields. When the wild-type serotypes, which endogenously express their cognate AAP already, were produced in the presence of a mixture containing all 10 AAPs, the transduction efficiency of the AAP-supplemented samples relative to the stuffer control was around one in all cases (Figure 2-8 B, C). This indicates that the extra AAP supplied during production had no effect on the yields of infectious particles. This experiment was subsequently expanded to chimeric capsids as well. Five cognate capsids of the previously tested chimeric AAP variants were chosen for production of recombinant particles and depending on the library origin, the corresponding wild-type AAP variants were supplied. Out of 25 chosen capsids, 19 were capable of transducing at least one of the cell lines tested and comparable to the wild-types the supplementation with extra AAP had no influence on particle yields in most cases (Figure 2-8 D). This is in accordance with recently published data from our group, showing that AAP overexpression in a baculovirus system likewise did not result in increased titers (Grosse et al. 2017). Taken together this suggests that excess AAP supplementation during library production has no effect, either positively or negatively, on particle production. Furthermore, this supports our conclusion that chimerism of AAP is not disadvantageous.

There was one exception, namely, capsid 18 whose own AAP failed to support the production of transduction-competent particles, albeit it was present in this assay. A remarkable increase in efficiency was observed upon exogenous AAP supplementation, however, mostly mediated by AAP5 and to a lesser extent by AAP2 (Figure 2-9). Six chimeric capsids failed to achieve transduction, which might be a consequence of either a defect in assembly or the development of novel tropisms incompatible with the cell lines tested. Strikingly, four of these six capsids originated from the 24589 library, including variant 18. Additionally, four AAP representatives of this library were not efficient at rescuing the AAV2 mutant (Figure 2-7 B). From the six variants which failed to support the AAV2 mutant, five contain an N-terminus derived from either AAV4 or AAV5 (supplements, Figure 6-2). As noted before, AAV4 and AAV5 are more divergent from the other serotypes, implying that sequence homology especially within the N-terminus might play a crucial role for the compatibility of a capsid with AAP. Congruent with this hypothesis, Tse et al. have very recently reported a series of AAP deletion and substitution mutants and showed that the AAP N-terminus is crucial for VP recognition

(Tse et al. 2018). In one particularly noteworthy experiment, they replaced part of the N-terminus of AAP1 with AAP5 and observed a drop in AAV1 vector yield. This supports our aforementioned data and conclusion, that the AAP N-terminus is most important for conveying compatibility with a given AAV capsid.

Furthermore, we studied the functionality of AAPs which had undergone co-selection with their cognate capsids. The underlying rationale was our interesting observation that after selection, some chimeric vectors produced better in comparison to the wild-types with which they shared a large proportion of the AAP C-terminus. One ensuing hypothesis was that an extra supply of AAP might increase particle yields, akin to what we had observed for the AAV3 mutant that was boosted beyond wild-type level with certain AAP variants. Therefore, we hypothesized that the observed increase in production efficiency was mediated by AAP. However, our data from the *trans*-complementation assay showed that in most cases, the evolved AAPs mediated a rescue back to wild-type levels, but not further (Figure 2-13 B). This was especially true for AAV1 and AAV6. The other two mutant versions of AAV7 and AAVrh10 were not so readily rescued which is in accordance with previous data (Figure 2-5 B). This implies that the increase in production efficiency is most likely not associated with AAP itself but rather a result from other factors, such as capsid stability as intermediate production and purification steps were involved, which are in turn prone to favoring variants that produce at high titers and/or are more stable during density gradient centrifugation. Most importantly, the chimeric AAPs resulting from the selection were functional and, hence, not limiting for the selection.

When AAP was initially reported, there was a concern in the field that the inadvertent shuffling of AAP might be detrimental for the generation of viral libraries by DNA family shuffling. Luckily, recently published first results from our group already indicate that this risk may have been over-estimated (Große 2016). In the study presented here, previous experiments (Figure 2-7 B) were expanded to a higher sample size and additional information was gathered on the effect of DNA shuffling on AAP functionality.

The overall conclusion is that inadvertent shuffling of AAP is not as detrimental as thought of, supported by our findings that (i) an AAV2 mutant could be rescued by 90% of chimeric AAPs tested and that the remaining variants rescued at least one other parental serotype. This, in combination with the high interchangeability of wild-type AAPs, indicates a substantial compatibility between AAP and capsid variants and thus a high degree of compatibility. Furthermore, (ii) the supplementation of AAP during production of already AAP-expressing rAAV had negligible influence on transduction

efficiency, suggesting that the particle yields were not increased by the extra AAP. In addition, we studied evolved AAP variants and showed that (iii) AAP DJ combines characteristics from several parents. Lastly, (iv) 12 additional evolved AAPs also displayed broad *trans*-complementation ability.

Of note, the high degree of compatibility seemed to be dependent on homology as incorporation of the more divergent serotypes AAV4 and AAV5 lead to reduced interchangeability. We therefore conclude that it should be beneficial for future studies to identify crucial domains in AAP that govern its function. In this context, the recent study identifying the AAV N-terminal part as particularly important for VP recognition is highly interesting (Tse et al. 2018). Consequently, this region could be kept conserved to prevent adverse shuffling at this position and thus disruption of AAP activity. Alternatively or in addition, supplying extra AAP during production might turn out to be beneficial for some capsids variants (such as clone 18). Concurrently, it is tempting to study the domains important for AAP functionality by applying DNA family shuffling directly to AAP. For instance, the AAPs from highly compatible AAPs, such as AAP2 or AAPrh10, could be shuffled with a more divergent AAP, such as AAP5. This fundamental approach, *i.e.*, the use of DNA family shuffling to dissect sequence-function relationships, has recently been exploited to identify the determinants of the AAVrh10 capsid that mediate crossing of the blood-brain barrier (Albright et al. 2018). In this study, the authors shuffled AAV1, which is non-permissive to the brain, with AAVrh10 and then validated several variants *in vivo*. Eventually, they identified an eight amino acid long “footprint”, which upon engraftment into AAV1 led to brain transduction. Likewise, DNA family shuffling was used to study the domains of the human Argonaute (Ago) proteins Ago1-4 which are involved in the RNA interference pathway (Schurmann et al. 2013). In this example, the biological mechanisms underlying RNA cleavage were further unraveled by shuffling of the slicing competent Ago2 with the slicing deficient Ago proteins, such as Ago3 and Ago4.

From the data obtained in this work, one can imagine different phenotypes which might occur during production of viral libraries (depicted in Figure 3-1). Simply put, a shuffled capsid is either capable to assemble particles (P1, P2) or not (P4). The latter could have several reasons, including premature stop codons in the AAP or VP sequence, or sterical hindrance of amino acids preventing capsid formation. Based on the high interchangeability of AAPs to *trans*-complement capsid variants, phenotype P3 is another option as well. In this scenario, the cognate AAP is incompatible with the own VP proteins and thus fails to support capsid assembly, as it was the case for capsid variant 18 in this work. Yet, the presence of an AAP originating from another shuffled

variant might rescue this AAP-dependent defect in capsid assembly. Concerning this phenotype, two considerations should be kept in mind. Firstly, the data presented here indicates that albeit this phenotype is possible, it is rather rare and predominantly occurred upon shuffling of less homologous parental sequences. A supplementation of AAPs during production might therefore be considered whenever divergent sequences are shuffled and incompatibility issues are more likely to arise. This should stabilize variants that would otherwise be lost although they might have had interesting properties for a certain selection strategy or downstream application, respectively.

The second and most important consideration concerns the phenomenon of genotype-phenotype linkage, *i.e.*, encapsidation of a genome into the capsid that it encodes among a myriad of possible candidates in the same cell. Until recently it was not clear if such a linkage exists at all. One strategy to prevent packaging of a genome into an unrelated capsid is to transfect minimal amounts of capsid-encoding library DNA into the cells, to yield a low plasmid copy number per cell (*e.g.* (Maheshri et al. 2006)). Strikingly, however, other groups transfected much higher amounts of DNA and still succeeded at identifying functional chimeras (*e.g.* (Grimm et al. 2008)). Notably, a recent study reported an astonishingly high correlation of packaged genomes upon co-transfection in a replication- and packaging-competent AAV context, that was relieved as soon as the ITRs were removed from the *cap*-encoding plasmid (Nonnenmacher, van Bakel, et al. 2015). Two possible explanations for this unexpected observation are a *cis*-packaging mechanism and a pioneer plasmid which outgrows other present plasmid variants. The authors hypothesized that the initial rescue of the viral genome from the plasmid is rather inefficient, yet as soon as this genome has been excised, it will replicate and outgrow the others. The resulting excess amounts of this particular variant in the cell automatically results in an abundance of the encoded capsid and thus increases the likelihood of correct genotype-phenotype linkage. In order to rescue a capsid with an incompatible cognate AAP (P3) *trans*-complementation with a compatible AAP must occur within the same cell, *i.e.*, both must be expressed at the same time. In case the pioneer plasmid hypothesis fosters the observed phenotype-genotype linkage, residual translational activity of AAP for the remaining, outgrown, plasmids is necessary.

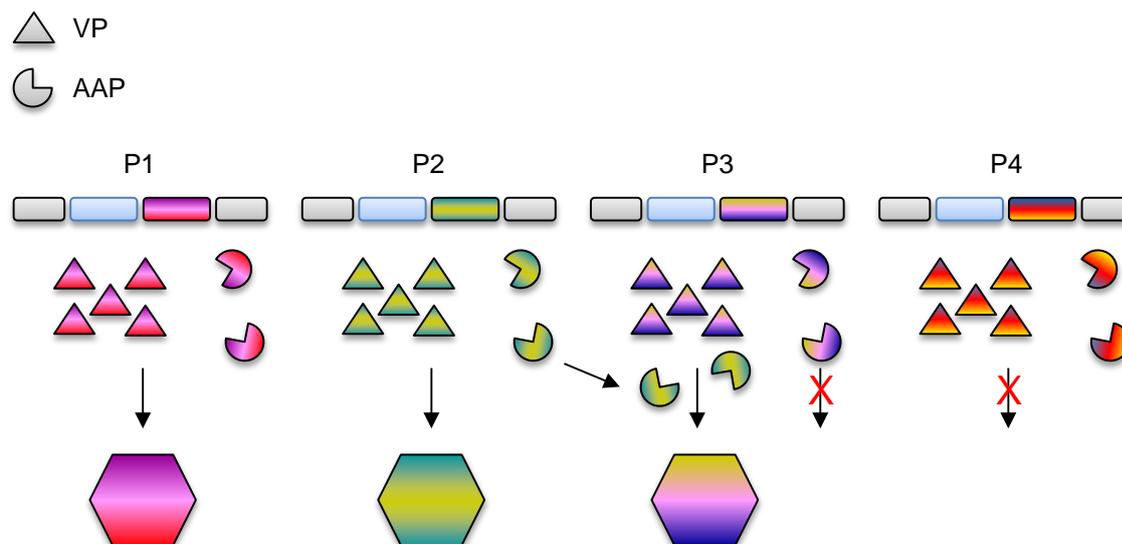


FIGURE 3-1: POSSIBLE PHENOTYPES DURING LIBRARY PRODUCTION

The model depicts different phenotypes (P) which might occur during particle production. In the best case, the VPs are compatible with the cognate AAP and capsids are assembled (P1, 2). However, it is also possible that the VPs have a sterical defect and cannot assemble, or that a premature stop codon has been introduced by the shuffling procedure (P4). Finally, it is conceivable that the VPs are incompatible with their respective AAP (P3) but can be rescued by AAP originating from another capsid (P2).

### 3.3. SELECTION IN STELLATE CELLS

Stellate cells belong to the non-parenchymal cell fraction in the liver and are highly relevant for liver disease progression as they, for example, participate in liver fibrosis. Because tools for targeted genetic manipulation of these cells remain limited, a major goal of this work was to identify synthetic AAV variants from a library of shuffled capsids that can fill in this gap.

In order to identify promising AAV variants efficient at stellate cell targeting, four rounds of *in vivo* selection were performed by peripheral injection of an AAV library into adult mice, followed by stellate cell isolation, total DNA purification and subsequent PCR-based *cap* rescue of AAV genomes that had entered the cells.

Sanger sequencing of single clones, performed to monitor and track library composition, revealed a number of interesting changes particularly in the C-terminus. Typically, the N-terminus is buried within the capsid and is important for intracellular trafficking steps (section 1.1.1.3), whereas the C-terminus forms the capsid surface and is involved in cell binding. Therefore, the C-terminal part was sorted according to the respective parental serotypes, resulting in the identification of different groups that were named according to

the most abundant parental reference, *i.e.*, 1-stretch, 7-stretch, 10-stretch, 7-rh10-7, 1/6 mix and 1/6/7 mix. From each group, single variants were further validated *in vivo* by immunostainings of liver slices (Figure 2-14).

Intriguingly, a variety of different phenotypes were observed. While some variants transduced a low number of cells overall (3-18 and 4-7), others exhibited a pronounced tropism towards hepatocytes (3-9 and 4-3). In addition, several capsid chimeras had developed a novel tropism, as best seen for 4-6, 3-10, 3-5 and 4b-10. With these capsids, fewer to no hepatocytes were expressing the reporter and most positive cells were phenotypically different from hepatocytes. Instead, reporter eGFP expression co-localized with desmin staining, *i.e.*, a marker for stellate cells.

As the liver is a complex organ composed of different cell types in very close proximity, the method of analysis was changed from studying liver slices to separated cell populations. As a pilot, the 3-5 vector was injected into mice and different cell types were isolated from each mouse (Figure 2-15), taken into primary cell culture and stained for different markers. Now that single cell types could be assessed more readily and more unanimously, it became evident that both, stellate cells and Kupffer cells were targeted efficiently by capsid 3-5.

Curiously, 3-5 exhibited different tropisms in the liver slices (Figure 2-14), namely, activity in either NPCs or hepatocytes. This observation was not confirmed when single cell types were analyzed, either by single-cell type staining (Figure 2-13) or by NGS analysis. It needs to be tested whether this is related to mouse strain or the method. The latter could be tested by performing a partial hepatectomy for immuno stainings of liver slices and perfusion of the remaining liver for single-cell type isolation.

One inherent drawback for the single-capsid validation approach is the limited amount of variants which can be screened due to the high number of animals and the manpower required. Therefore, an alternative vector barcoding and NGS strategy was pursued with the aim to validate 28 interesting chimeras in parallel (section 2.3.2.3). Firstly, all variants within one cell type were ranked by their efficiencies of transgene expression, as measured on the RNA/cDNA level. This revealed a similar overall picture for stellate cells and the other cell types, namely, hepatocytes, LSECs and Kupffer cells (Figure 2-17). Unexpectedly, though, none of the capsid variants selected in stellate cells appeared in the top 15 ranking on the cDNA level. Instead, several variants from an unrelated screen performed by another group member were enriched, next to peptide insertion variants and the wild-types AAV7 and AAVrh10.

Interestingly, the picture looks drastically different for the ranking on the gDNA level (Figure 2-18). In stellate cells, the majority of variants ranking high in this list are now indeed derived from the *in vivo* screen performed in this work. This indicates that the applied selection procedure was actually successful at enriching variants that could enter the target cell population, *i.e.*, stellate cells. A most likely explanation for the striking discrepancy of the results of the cDNA and gDNA ranking lies in the selection strategy itself. As described above, whole cells were taken for total DNA isolation as well as for PCR amplification and subcloning of enriched AAV capsid genes. Consequently, there is a high probability that variants were isolated that had delivered their genetic cargo into the cell (or had remained attached to the outside), but had failed to traffic to and/or enter the nucleus and mediate transgene expression. (Grimm and Zolotukhin 2015). Alternatively, it is conceivable that some capsid structures are very stable and thus prevent disassembly, which is a prerequisite for vector genome release and transgene expression. In fact, the observed increase in titer for selected clones might correlate with such an increase in capsid stability. Also of note in this context, when Thomas et al. investigated the reason for AAV8's superior transduction profile over AAV2 in the mouse liver (Thomas et al. 2004), they found capsid stability and uncoating kinetics to be key to the rapid onset of transgene expression. While AAV8 uncoats and thus releases vector genomes very rapidly in the mouse liver, the AAV2 capsid is stable for a longer period of time and therefore takes longer to form double-stranded, transcription-competent vector DNA in the nucleus.

Generally, the gDNA profile of the other NPCs shared similarities with the one from stellate cells. Particularly surprising was the high abundance of AAV5 wild-type gDNA in all NPCs that was not observed in the parenchymal cells, *i.e.*, hepatocytes, and that did not correlate with the cDNA data for AAV5. This finding is of particular interest as currently clinical studies are underway which use liver-directed targeting with AAV5 vectors (*e.g.* (D'Avola et al. 2016; Rangarajan et al. 2017)). From the data obtained AAV5 is not as efficient to translate its high gDNA content into transgene expression which leads to the question whether the best candidate was chosen. Ideally, the vector should exhibit high transgene expression with little gDNA content, *i.e.*, less particles entering the cells, thereby reducing the overall vector dose. High numbers of particles pose the danger of immune responses as reaction to capsid fragments correlates with the dose (Crudele et al. 2015).

Furthermore, the gDNA lead candidates identified in stellate cells were not restricted to this cell type but were also found in the other NPC fractions. Especially, Kupffer cells had the highest degree of agreement with stellate cells. Additionally, a high correlation of

cDNA and gDNA ranks was observed for hepatocytes and LSECs as many variants showed up at similar position on the two lists and this trend was confirmed upon linear regression analysis of a higher sample size. This implies that capsids which enter the cells (gDNA) efficiently transcribe RNA (cDNA) from their encapsidated genomes. This directly influences the selection strategy as a selection on DNA level will most likely yield promising variants which additionally express delivered transgenes well.

In contrast, Kupffer cells and to a lesser extent stellate cells showed a low level of accordance between gDNA and cDNA which was confirmed further with the linear regression analysis. Here, high gDNA amounts do not translate into high expression levels. These findings are of great interest and relevance as they imply that the AAV vector-host interaction mechanisms appear to be different in stellate cells as well as in Kupffer cells compared to hepatocytes and LSECs. Consequently, a selection purely on the DNA level - as performed in the present pilot study - will not automatically lead to the identification of efficient, *i.e.*, transcriptionally active vectors.

We therefore propose an alternative selection strategy that is based on the RNA level and thus provides a much higher chance to enrich variants that successfully uncoat and express their genome. However, this is technically challenging due to the biology of AAVs. First of all, there is only minimal expression of the capsid gene in the absence of a helper virus (section 1.1.1.4). Furthermore, there are (at least) two transcripts for the VP proteins, and in order to recover the complete *cap* gene, the largest VP1 transcript is needed, which is less abundant (section 1.1.1.1). One approach in this direction was implemented by Deverman *et al.* who identified the AAV9-based peptide variant PHP.B by using their so-called CREATE system (Deverman *et al.* 2016). Here, the AAV *cap* gene is flanked by a floxed primer binding site that becomes reverted upon cell-type specific expression of Cre recombinase and can then be harnessed for a PCR rescue of enriched *cap* sequences.

The second kind of analysis that is enabled by the data set is to compare the abundance of a given capsid variant in different cell types. Here, an alternative normalization strategy is pursued that considers the vector load in a given cell type. Hence, the proportions acquired are valid per cell and disregard the complex organization of the organ itself. Among the top 15 lead candidates on the cDNA level for stellate cells, the chimeras derived from an unrelated screen in muscle (JEA) were found to have an even distribution between the different cell types (Figure 2-20). However, this property is not optimal for a cell-type specific vector. Ideally, the capsid itself is already as specific as

possible or at least displays increased on-targeting. For this reason, the JEA chimeras, that exhibit high transgene expression, are not the best option.

Most interesting in the context of this work was whether the *in vivo* HSC selected chimeras had become more specific for stellate cells during iterative selection. In fact, most chimeras exhibited a mixed phenotype with activity among different cell types, akin to the characteristics of the parental AAV7 and AAVrh10. This result is in accordance with the selection strategy, as variants were enriched that were efficient at entering stellate cells. However, such a positive on-target selection does not automatically also convey specificity. Ideally, the application of a negative selection pressure would be highly beneficial to remove variants entering other cell types than the target cell, *i.e.* stellate cells in this study. In an *in vivo* setting this will pose to be difficult due to the complex organ architecture. In an *in vitro* assay one could use the library to firstly infect the off-target cells and then transferring the medium containing the remaining AAV variants to the target cell population.

Another curious observation was that the abundance of a given variant in stellate cells was highly comparable to that in Kupffer cells. This apparent connectivity between these two cell types was unexpected, but could possibly be explained by the Kupffer cell's nature as a macrophage as they ingest pathogens. Possibly, some AAV variants escape the degradation pathways to express their transgenes which could explain the shallow slope in the cDNA and gDNA rank comparison, which indicates an uncoupling of vector DNA presence and transcription. Alternatively, considering the fact that stellate and Kupffer cells had comparable cDNA and gDNA profiles, it might be that certain aspects of virus-host interactions are common as well, thus enriching the same AAV capsid variants during selection. Last but not least, the purification method is relevant, too. If contaminating cell types were still present after stellate cell purification, AAV vectors infecting these will contaminate the PCR rescue reactions as well. Indeed, the formation of stellate cell/Kupffer cell doublets that could have been co-isolated has been reported (Bartneck et al. 2015). However, here, the purification of stellate cells occurred based on FACS sorting which is considered pure (>99% according to (Mederacke et al. 2015)).

The overall aim of the screen was the identification of AAVs specific for a certain liver cell type, and indeed, good candidates were identified that should be studied further. A most notable example is the chimera DJ that was originally selected *in vitro* in hepatocytes and that, remarkably, maintained its hepatocyte tropism *in vivo* (Figure 2-21). The overall efficiency of targeting LSECs was low, and no preferred candidates have been identified yet, in line with reported challenges to transduce LSECs with AAVs

(Buchholz, Friedel, and Buning 2015). Here, a separate selection should be considered and based on the existing gDNA/cDNA correlation a selection on DNA level could be conducted. The selection in stellate cells yielded variants which efficiently entered the target cells and showed off-targeting from hepatocytes. This promising result is encouraging for subsequent selection steps and as the tested variants were not efficient a mixture of different libraries (shuffled, peptide insertions and combinations thereof) could lead to further diversification thereby increasing chances for the identification of promising variants.

Initially, the variant 3-5 was the lead candidate due to its performance in stellate cells and Kupffer cells in the single-cell type experiment. The NGS screen showed that the efficiency is even slightly higher in Kupffer cells than in stellate cells. Accordingly, this capsid will be considered further for both cell types. Likewise, 4N1 will be validated as well as it exhibited a higher efficiency for stellate cells over Kupffer cells. One advantage of the two chimeras over the assigned parental wild-types AAV1 and AAV6 is the increase in production efficiency. Especially for experiments in mice or higher species, the ability to obtain high-titer stocks is critical as it will allow to reduce the injection volume. Moreover, the increase in production efficiency reduces work loads and costs as productions can be scaled down.

The choice of capsid is the first important step for efficient target cell transduction. The NGS screen established that a very low proportion of hepatocytes were transduced in the biodistribution analysis by the lead candidates 3-5 and 4N1, but this will eventually translate into a higher hepatocyte number as they constitute the majority of the cells in the liver. In addition, the vector dose will be increased by over 10-fold for the actual experiments in comparison to the NGS screen. Therefore, additional levels of restriction were implemented by enhancing the specificity of the vector cassette itself and all generated constructs were tested *in vitro* with cell lines modeling the different cell populations, as described in the Results section. To achieve detargeting from hepatocytes, vectors were generated bearing miRNA-122 bs. This miRNA is abundantly and selectively expressed in hepatocytes and was shown to mediate AAV vector detargeting before (e.g. (Geisler et al. 2011; Xie et al. 2011)). As the capsids are not able to discriminate between Kupffer cells and stellate cells, binding sites for miRN-142-3p were additionally incorporated. This miRNA is abundant in cells from hematopoietic origin and was shown to mediate Kupffer cell-specific knockdown *in vivo* in a lentiviral vector (Merlin et al. 2017). In the *in vitro* setting, the VEC promoter restricted transgene expression to a specific cell line, namely HUVECs. Likewise, the VEC promoter and the

cd11b promoter were previously shown to being functional in mice in a cell-type specific manner (Merlin et al. 2017).

Ultimately, we believe that a triple strategy needs to be pursued in order to achieve maximum specificity and efficiency (Figure 3-2). Already, good capsids were identified for hepatocytes (DJ), Kupffer cells (3-5) and stellate cells (3-5 and 4N1). In case of residual off-targeting, genes of interest can be transcribed from a cell-type specific promoter such as VEC or cd11b to further restrict the expression to the on-target cell. Analogously, expression can be turned off by the incorporation of cell-type specific miRNA binding sites as observed in the *in vitro* study by miRNA122 or 142-3p binding sites. In the future, these vectors will be validated in combination with the chosen capsids.

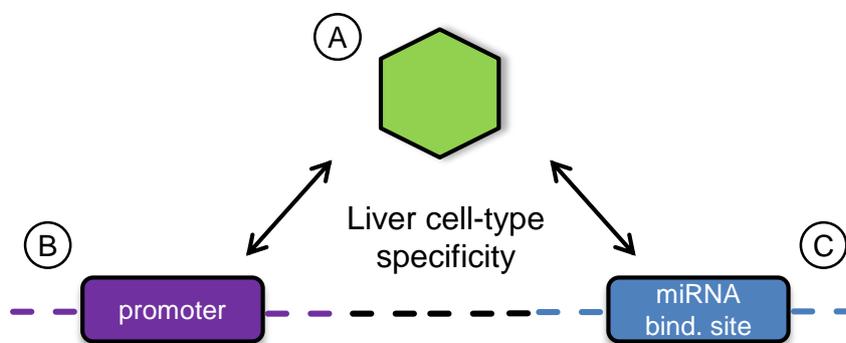


FIGURE 3-2: TRIPLE STRATEGY TO MEDIATE SPECIFICITY

A strategy is depicted for the generation of liver cell-type specific AAV vectors. There are several levels where specificity will be mediated: the chosen capsid (A), the choice of promoter (B) and the incorporation of miRNA binding sites to mediate mRNA degradation in off-target cells.

### 3.4. OVERALL CONCLUSIONS

In this study, we dissected and improved - where possible - different crucial steps in the protocol for AAV DNA family shuffling, with the overall aim to streamline and optimize this exciting technology and to thereby make it even more accessible to members of the biomedical community. Our key results include our demonstration that enzymatic digestion based on DNaseI is the more robust approach as compared to physical fragmentation, despite the predicted advantages of the latter. Additionally, we improved the methodology for purification of the *cap* DNA pool after the nested PCR step, by replacing gel extraction with direct column purification, which saves time, work and consumables.

On top, through extensive and comprehensive *trans*-complementation studies, we relieved concerns about AAP impairment as an adverse by-product of shuffling of the *cap* gene, that were originally triggered when the overlap in ORFs of AAP and *cap* were discovered. While our data clearly show that chimeric AAPs mostly retain functionality, we also made a variety of surprising observations, such as the different susceptibility of AAV serotypes to complementation with ectopic AAP that is most pronounced for the least homologous AAV4 and AAV5. Concurrently, we uncovered possibilities for improvement in the future, including the addition of excess AAP during library production with the aim to boost library viability and/or to avoid the early loss of interesting clones whose own capsid and AAP may be incompatible. Similarly, we made highly relevant observations that will help to advance future selection strategies. To our best knowledge, our study is the first to perform an AAV capsid library selection in the liver at higher resolution, *i.e.*, on the level of specialized cellular subtypes rather than on the level of the whole organ. Our approach to concomitantly study the distribution of different AAVs in hepatocytes, stellate cells, Kupffer cells and LSECs, is unique as it is the only so far that respects the distinct functions and receptor compositions of single specialized cells within the liver. Notably, our proof-of-concept selection in stellate cells succeeded as we were able to enrich chimeras that deliver substantial amounts of vector DNA to these cells.

At the same time, we learned two vital additional lessons that can be harnessed in future reiterations of this selection scheme or in modifications thereof in other tissues. Firstly, we identified a discrepancy between the ranking of lead candidates on the gDNA versus the RNA level, *i.e.*, the amounts of vector DNA that had entered the cells did not always correlate with the amount of transcripts. This observation of capsid- and cell type-specific gDNA/RNA uncoupling implies that future studies should aim to establish a rescue on the mRNA instead of the DNA level, to foster the enrichment of capsids that are transcriptionally most active. Secondly, albeit the *in vivo* selection in stellate cells enriched efficient capsid variants, it became clear that it may be difficult to solely convey specificity through the capsid itself. Therefore, to further restrict gene expression to a specific cell type, we developed new combinatorial vector cassettes that carried cell-specific promoters and miRNA binding sites, and that can easily be adapted to mediate optimal specificity and efficiency in other target cells as well.

In summary, we have identified a variety of possible improvements for future AAV capsid library generation and selection strategies, based on which we conclude with the following recommendations: (i) an AAP cocktail should be supplemented during production in cases where less homologous AAVs are present in the parental library; (ii)

to increase chances to enrich transcriptionally active capsids, a selection based on the mRNA level is preferred; (iii) a repertoire of cell-type specific promoters and miRNAs should be gathered and combined with transductional targeting mediated by the capsid; (iv) unique library-specific barcodes can be incorporated into the backbones to reduce the risk of cross-contamination during rescue steps and to concurrently enable multiplexed library screening; and, finally, (v) initial library quality and progress of the selection should be monitored by next-generation sequencing technology such as PacBio/SMRT, to cover a larger sample size and to thus reduce bias that is invariably introduced by conventional Sanger sequencing. Based on our cumulative experience from this work and from the past decade, we are highly optimistic that these and other improvements developed by our colleagues in the field will synergize and accelerate both, the wider application of molecular AAV evolution technology and the isolation of powerful new gene therapy vectors with unprecedented efficiency and specificity.

## 4. MATERIALS AND METHODS

### 4.1. REAGENTS, BUFFERS AND DEVICES USED

TABLE 4-1: LIST OF DEVICES

Application	Device	Company
Bacterial incubators	Heraus function line incubator	Thermo Fisher Scientific (Waltham, USA)
	Shaking Incubator Multitron	INFORS HT (Basel, Switzerland)
	Shaking Incubator Minitron	INFORS HT (Basel, Switzerland)
Cell culture/ sterile working	Countess	Thermo Fisher Scientific (Waltham, USA)
	HERA safe sterile work bench	Thermo Fisher Scientific (Waltham, USA)
Centrifugation	HERA cell 150 incubator	Thermo Fisher Scientific (Waltham, USA)
	Allegra X-12R centrifuge	Beckman Coulter (Brea, USA)
	Argos flexifuge	Biozym Scientific GmbH (Hessisch Oldendorf, Germany)
	Avanti J-26 XP centrifuge	Beckman Coulter (Brea, USA)
	Avanti J-25 centrifuge	Beckman Coulter (Brea, USA)
	Beckman tube sealer	Beckman Coulter (Brea, USA)
	Benchtop centrifuge 5415R	Eppendorf (Hamburg, Germany)
	Benchtop centrifuge 5417R	Eppendorf (Hamburg, Germany)
	500 ml centrifuge bottles	Beckman Coulter (Brea, USA)
	Fixed angle type 70 Ti rotor	Beckman Coulter (Brea, USA)
	Fixed angle type 70.1 Ti rotor	Beckman Coulter (Brea, USA)
	Galaxy minister	VWR (Fenenay-sous-Bais, France)
	Optima™ L-90K Ultracentrifuge	Beckman Coulter (Brea, USA)
	Optima™ Ultracentrifuge tubes (26x77 mm)	Beckman Coulter (Brea, USA)
	Ultracentrifuge tubes 16x76 mm	Seton Scientific (Petaluma, USA)
Electroporation	Gene Pulser Xcell	Bio-Rad (Hercules, USA)
Flow cytometry	Cytomics FC500MPL analyzer	Beckman Coulter (Brea, USA)
Gel electrophoresis	Gel Doc XR	Bio-Rad (Hercules, USA)
	MINI-SUB CELL GT	Bio-Rad (Hercules, USA)
	Mitsubishi P93D	Mitsubishi Electric (Cypress, USA)
	SUB CELL GT	Bio-Rad (Hercules, USA)
	UV-Transilluminator	Biostep GmbH (Jahnsdorf, Germany)
Microscopy	CKX 419F	Olympus cooperation (Tokyo, Japan)
	U-RPL-T	Olympus cooperation (Tokyo, Japan)
other applications	accujet pro	BrandTech Scientific (Essex, UK)
	Bunsen burner	Carl Friedrich Usbeck KG (Radevormwald, Germany)
	Covaris focused-ultrasonicator S2	Covaris (Woburn, USA)

	Bio-Dot® SF Microfiltration Apparatus	Bio-Rad (Hercules, USA)
	Lauda Aqualine AL5	DJB Labcare (Buckinghamshire, UK)
	Magnetic stirrer	Thermo Fisher Scientific (Waltham, USA)
	Microwave	Sharp Electronics (Hamburg, Germany)
	Mixing block MB-102	Biozym Scientific GmbH (Hessisch Oldendorf, Germany)
	Nanodrop 2000	Thermo Fisher Scientific (Waltham, USA)
	NanoVue Spectrophotometer	Thermo Fisher Scientific (Waltham, USA)
	pH meter PB-11	Sartorius (Göttingen, Germany)
	Pipettes	Gilson (Middleton, Germany)
		Eppendorf (Hamburg, Germany)
	Qubit fluorometer	Thermo Fisher Scientific (Waltham, USA)
	Refractometer	Exacta Optech (San Prospero, Italy)
	Shaker DRS-12	neoLab (Heidelberg, Germany)
	Sonorex ultrasonic bath	Bandelin (Berlin, Germany)
	St5 cat shaker	M. Zipperer GmbH (Staufen, Germany)
	Tissue lyser LT	QIAGEN (Hilden, Germany)
	Tube roller TRM-V	neoLab (Heidelberg, Germany)
	Vacuum pump	Promega (madison, USA)
	Vortex Genie2	Scientific Industries (Bohemia, USA)
	Water bath TW12	Julabo Labortechnik (Seelbach, Germany)
	Weighing scale	KERN & SOHN GmbH (Balingen, Germany)
PCR	C1000 Touch thermal cycler	Bio-Rad (Hercules, USA)
	Corbett RG6000	QIAGEN (Hilden, Germany)
	FlexCycler	analyticjena (Jena, Germany)
	QX200™ Droplet Generator	Bio-Rad (Hercules, USA)
	QX200™ Droplet Reader	Bio-Rad (Hercules, USA)
	vapo protect	Eppendorf (Hamburg, Germany)
Western Blot	Film developing cassettes	Dr. Goos-Suprema GmbH (Heidelberg, Germany)
	Mini-PROTEAN Tetra cell chamber	Bio-Rad (Hercules, USA)
	PowerPac basic/ HV/ HC	Bio-Rad (Hercules, USA)
	Trans-Blot® SD Semi-Dry Electrophoretic	Bio-Rad (Hercules, USA)
	ChemoCam (ECL Imager)	INTAS Science Imaging Instruments (Göttingen, Germany)
	X-OMAT 2000 processor (film developer)	KODAK (Rochester, USA)

---

TABLE 4-2: LIST OF CHEMICALS, REAGENTS AND KITS

Reagent	Name	Company
Chemicals	Acetic acid	VWR chemicals (Fenenay-sous-Bais, France)
	Agarose	Biozym Scientific GmbH (Hessisch Oldendorf, Germany)
	Albumin Fraktion V (BSA)	Roth (Karlsruhe, Germany)
	Ammonium persulfate (APS)	GRÜSSING GmbH (Filsum, Germany)
	Bromophenol blue	CHROMA (Bellows Falls, USA)
	Calcium chloride (CaCl <sub>2</sub> )	Roth (Karlsruhe, Germany)
	Cesium chloride (CsCl)	Roth (Karlsruhe, Germany)
	DMSO	Merck (Darmstadt, Germany)
	Dodecylsulfate-Na-salt-pellets (SDS)	SERVA Electrophoresis GmbH (Heidelberg, Germany)
	Ethanol absolute	SIGMA-ALDRICH (St. Louis, USA)
	Ethidium bromide	Roth (Karlsruhe, Germany)
	Ethylendiamintetraacetate (EDTA)	GRÜSSING GmbH (Filsum, Germany)
	Glucose	MERCK (Darmstadt, Germany)
	Glycerol	VWR chemicals (Fenenay-sous-Bais, France)
	HEPES	Roth (Karlsruhe, Germany)
	Hydrochloric acid (HCl)	SIGMA-ALDRICH (St. Louis, USA)
	Iodixanol (Optiprep™)	Progen (Heidelberg, Germany)
	Isopropanol	SIGMA-ALDRICH (St. Louis, USA)
	Magnesium chloride (MgCl <sub>2</sub> )	Applichem (Darmstadt, Germany)
	Magnesium sulphate (MgSO <sub>4</sub> )	MERCK (Darmstadt, Germany)
	Methanol	SIGMA-ALDRICH (St. Louis, USA)
	Milk powder	Roth (Karlsruhe, Germany)
	MOPS	SERVA Electrophoresis GmbH (Heidelberg, Germany)
	Nuclease-free water	Ambion, Thermo Fisher Scientific (Waltham, USA)
	PBS Dulbecco without Ca <sup>2+</sup>	MERCK (Darmstadt, Germany)
	PEG8000	Promega (Madison, USA)
	Phenol red	MERCK (Darmstadt, Germany)
	Polyethylenimine (PEI), linear	Polysciences Inc. (Eppelheim, Germany)
	PonceauS	SIGMA-ALDRICH (St. Louis, USA)
	Potassium acetate (KAc)	GRÜSSING GmbH (Filsum, Germany)
	Potassium chloride (KCl)	GRÜSSING GmbH (Filsum, Germany)
	Rotiphorese Gel 40 (19:1)	Roth (Karlsruhe, Germany)
	Sodium chloride (NaCl)	GRÜSSING GmbH (Filsum, Germany)
	Sodium hydroxide (NaOH)	SIGMA-ALDRICH (St. Louis, USA)
	TE buffer	Thermo Fisher Scientific (Waltham, USA)
	TGS (Tris/Glycine/SDS buffer) 10x	Bio-Rad (Hercules, USA)
	Tris-HCl/ Tris	Roth (Karlsruhe, Germany)
	Tween20	Roth (Karlsruhe, Germany)
	UltraPure™ TEMED	Thermo Fisher Scientific (Waltham, USA)
	β-Mercaptoethanol	Roth (Karlsruhe, Germany)
Enzymes	Antarctic Phosphatase	NEB (Ipswich, USA)

	Benzonase	MERCK (Darmstadt, Germany)
	DNase I	Thermo Fisher Scientific (Waltham, USA)
	HotStar Hifidelity Polymerase	QIAGEN (Hilden, Germany)
	OneTaq® 2X Master Mix with Standard Buffer	NEB (Ipswich, USA)
	Phusion Hot Start II DNA Polymerase	Thermo Fisher Scientific (Waltham, USA)
	Proteinase K	Roche (Penzberg, Germany)
	RNaseA	QIAGEN (Hilden, Germany)
	Restriction Enzymes	NEB (Ipswich, USA) / Fermentas (St. Leon-Rot, Germany)
	T4 DNA Ligase	NEB (Ipswich, USA)
Kits	ddPCR™ Supermix for Probes (No dUTP)	Bio-Rad (Hercules, USA)
	DirectPCR Lysis Reagent	Viagen Biotech Inc (Los Angeles, USA)
	DNA clean & concentrator-5	zyzo research (Irvine, USA)
	DNeasy Blood & Tissue Kit	QIAGEN (Hilden, Germany)
	NucleoBond® Xtra Midi / Maxi	Macherey-Nagel (Hœrdt, France)
	PureYield™ Plasmid Midiprep kit	Promega (Madison, USA)
	SensiMix™II Probe Kit	Bioline (London, UK)
	QIAprep Spin Miniprep Kit	QIAGEN (Hilden, Germany)
	QIAquick Gel Extraction Kit	QIAGEN (Hilden, Germany)
	QIAquick PCR Purification Kit	QIAGEN (Hilden, Germany)
	Qubit dsDNA BR Assay Kit	Thermo Fisher Scientific (Waltham, USA)
	Western Lightning® PLUS-ECL	PerkinElmer (Waltham, USA)
Media (eukaryotic)	DMEM GlutaMAX +4.5 g/ L D-Glucose	Gibco by Thermo Fisher Scientific (Waltham, USA)
	-pyrovat	
	Endothelial Cell Growth Medium	Promocell (Heidelberg, Germany)
	Endothelial Cell Growth Medium SupplementMix	Promocell (Heidelberg, Germany)
	1x DPBS	Gibco by Thermo Fisher Scientific (Waltham, USA)
	Fetal Bovine Serum Gold (FBS)	Gibco by Thermo Fisher Scientific (Waltham, USA)
	Minimum Essential Medium Non-essential amino acids (100x)	Gibco by Thermo Fisher Scientific (Waltham, USA)
	Penicillin-Streptomycin	Gibco by Thermo Fisher Scientific (Waltham, USA)
	0.25% Trypsin/ EDTA	Gibco by Thermo Fisher Scientific (Waltham, USA)
Media (prokaryotic)	Ampicillin	Roth (Karlsruhe, Germany)
	Bacto™ Agar	BD (Franklin Lakes, USA)
	Bacto™ Trypton	BD (Franklin Lakes, USA)
	Bacto™ Yeast Extract	BD (Franklin Lakes, USA)
Standards	100 bp DNA ladder	Thermo Fisher Scientific (Waltham, USA)
	1 kb DNA ladder plus	Thermo Fisher Scientific (Waltham, USA)
	PageRuler™ Plus Prestained Protein Ladder	Fermentas (St. Leon-Rot, Germany)
	Colorplus prestained protein ladder	NEB (Ipswich, USA)

Oligonucleotides	dNTPs (dATP, dCTP, dGTP, dTTP)	NEB (Ipswich, USA)
------------------	--------------------------------	--------------------

TABLE 4-3: BUFFER COMPOSITIONS

Material	Composition
Ampicillin stock solution	5% (w/v) Ampicillin (50 mg/ ml)
Benzonase buffer	50 mM Tris/HCl pH 8.5 150 mM NaCl 2 mM MgCl <sub>2</sub>
10x DNA loading dye	50 mM Tris pH 7.6 60% glycerol 0.25% (w/v) bromophenol blue
6x Purple loading dye	NEB (Ipswich, USA)
ddPCR™ Droplet Reader Oil	Bio-Rad (Hercules, USA)
Droplet Generation Oil for Probes	Bio-Rad (Hercules, USA)
Freezing medium	10% (v/v) DMSO 90% (v/v) FBS
15% Iodixanol phase	25% (v/v) Iodixanol, 75% (v/v) PBS-MK-NaCl
25% Iodixanol phase	41.66% (v/v) Iodixanol, 58.33% (v/v) PBS-MK, phenol red until it appears red
40% Iodixanol phase	66.67% (v/v) Iodixanol, 33.33% (v/v) PBS-MK
60% Iodixanol phase	100% Iodixanol (v/v), phenol red until it appears yellow
LB medium	1% (w/v) Bacto Tryptone 0.5% (w/v) Bacto Yeast Extract 1% (w/v) NaCl
LB Amp medium	LB medium supplemented with 1 ml 5% Ampicillin per L
LB plates	LB medium plus 1.5% (w/v) Bacto Agar
LB Amp plates	LB plates supplemented with 1.5 ml 5% Ampicillin per L
Miniprep P1 (pH 8.0)	50 mM Tris/HCl pH 8.0 100 µg/ml RNase A 10 mM EDTA
Miniprep P2	200 mM NaOH 1% SDS
Miniprep P3 (pH 5.1)	2.8 M KAc
NaHEPES resuspension buffer	50 mM HEPES 150 mM NaCl 25 mM EDTA
PBS-MK	1 mM MgCl <sub>2</sub> 2.5 mM KCl in PBS
PBS-MK-NaCl	1 mM MgCl <sub>2</sub> 2.5 mM KCl 1M NaCl in PBS

#### 4 Materials and methods

---

40% PEG/ NaCl solution	40% (w/v) PEG8000 1.915 M NaCl
Phenolred solution	0.5% (w/v) phenol red
Protein sample buffer 2x	2 mM EDTA 100 mM Tris/ HCL pH 7.5 4% SDS 20% glycerol 10% $\beta$ -mercaptoethanol 0.02% bromphenolblue
Running gel buffer (pH 8.8)	1.5 M Tris 0.4% SDS [(v/v) from a 10% stock solution]
SOB medium (pH 7.0)	2% Bacto Tryptone 0.5% (w/v) Bacto Yeast Extract 10 mM NaCl 2.5 mM KCl 10 mM $MgSO_4$ (addition after autoclaving) 10 mM $MgCl_2$ (addition after autoclaving)
SOC medium	SOB medium plus 20mM glucose
Stacking gel buffer (pH 6.8)	500 mM Tris 0.4% SDS [(v/v) from a 10% stock solution]
50x TAE	1M Acetic acid 50mM EDTA 2M Tris
10x TBS	250 mM Tris/HCl, pH 7.4 1.25 M NaCl
TBST	1x TBS 0.05% Tween20
TFBI buffer (pH 5.8)	16 mM $CaCl_2$ 13.2 % (v/v) glycerol 30.6 mM KAc 100 mM KCl 80 mM $MgCl_2$
TFBII buffer (pH 8.0)	76 mM $CaCl_2$ 13.2 % (v/v) glycerol 10 mM KCl 4.8 mM MOPS
10x TGS (Running buffer)	Bio-Rad (Hercules, USA)
Topping solution (RI= 1.3710)	3.27 M CsCl (0.55g/ ml) In NaHEPES resuspension buffer
Transferbuffer	1x TGS 20% Methanol in $H_2O$

---

TABLE 4-4: LIST OF MATERIALS USED

Material	Company
Amicon Ultra-15 centrifugal filter units (100,000 NMWL)	MERCK (Darmstadt, Germany)
BD plastipak (syringe)	BD (Franklin Lakes, US)
BD microlance3	BD (Franklin Lakes, US)
Cell culture flasks (75/ 175 cm <sup>2</sup> )	Greiner bio-one (Frickenhausen, Germany)
Cell culture plates (6/12/96 well)	Greiner bio-one (Frickenhausen, Germany)
Cell culture plates 15cm	Nunc, Thermo Fisher Scientific (Waltham, USA)
Cell lifter	Corning (New York, USA)
Centrifuge tube 500 ml	Corning (New York, USA)
Coster 50ml reagent reservoir	Corning (New York, USA)
Countess <sup>TM</sup> cell counting chamber slides	Thermo Fisher Scientific (Waltham, USA)
Cuvettes (polystyrene)	Sarstedt (Nümbrecht, Germany)
Covaris reaction tubes (microTUBE AFA fibre Snap-Cap 6x16 mm)	Covaris (Woburn, USA)
Dialysis tubing /Type 20/32 inch, wall thickness 0.02 mm, MWCO 14,000)	Roth (Karlsruhe, Germany)
DG8 <sup>TM</sup> Cartridges for QX200 <sup>TM</sup> /QX100 <sup>TM</sup> Droplet Generator	Bio-Rad (Hercules, USA)
DG8 <sup>TM</sup> Gaskets for QX200 <sup>TM</sup> /QX100 <sup>TM</sup> Droplet Generator	Bio-Rad (Hercules, USA)
Electroporation Cuvettes (25x 1mm)	peqlab (Erlangen, Germany)
Erlenmeyer flasks	Thermo Fisher Scientific (Waltham, USA)
Filter tips	Sarstedt (Nümbrecht, Germany), Mettler-Toledo (Columbus, USA)
Glass bottles	DURAN group (Wertheim, Germany)
Glass culture tubes	DURAN group (Wertheim, Germany)
Microlance canules 21G 0.8x40mm, 19G 1.1x40 mm	BD (Franklin Lakes, USA)
Nitrocellulose membrane	Whatman (Maidstone, UK), Ahlstrom (Helsinki, Finland)
PCR tubes 0.2 ml 8-Strip	STARLAB (Hamburg, Germany)
Pasteur capillary pipettes (230 mm)	neoLab (Heidelberg, Germany)
Petri dishes	Greiner bio-one (Frickenhausen, Germany)
Pipette tips	Greiner bio-one (Frickenhausen, Germany), Kisker (Steinfurt, Germany)
qPCR tubes 0.1ml strips and lids	QIAGEN (Hilden, Germany)
Reaction tubes (0.5, 1, 2ml)	SARSTEDT (Nümbrecht, Germany), Eppendorf (Hamburg, Germany)
Reaction tubes 5ml	Eppendorf (Hamburg, Germany)
Reaction tubes 15 ml, 50 ml	Greiner bio-one (Frickenhausen, Germany)
Reaction tubes 50 ml	BD (Franklin Lakes, US)
Serological pipettes (5, 10, 25 ml)	Greiner bio-one (Frickenhausen, Germany)
Slide-A-Lyzer dialysis cassette (MWCO 20,000)	Thermo Fisher Scientific (Waltham, USA)
Sterile filter (0.22 µM pore size)	Greiner bio-one (Frickenhausen, Germany)
Steritop filter (0.22 µM)	MERCK (Darmstadt, Germany)
twin.tec® PCR Plates semiskirted	Eppendorf (Hamburg, Germany)
VacConnectors	QIAGEN (Hilden, Germany)

Whatman paper 3mm	Whatman (Maidstone, UK)
X-ray films Amersham Hyperfilm™ ECL	GE Healthcare (München, Germany)

TABLE 4-5: LIST OF SOFTWARES APPLIED

Application	Program	Vendor
Alignment analysis tool	Salanto	Developed in close collaboration with Christian Bender (Schurmann et al. 2013; Große 2016) ( <a href="https://bitbucket.org/benderc/salanto/wiki/Home">https://bitbucket.org/benderc/salanto/wiki/Home</a> )
Alignments	AlignX	Thermo Fisher Scientific (Waltham, USA)
	ClustalX2	(Larkin et al. 2007)
Covaris	SonoLite 2.07	Covaris (Woburn, USA)
ddPCR	QuantaSoft™ Software, Regulatory Edition	Bio-Rad (Hercules, USA)
Figure design	MS office	Microsoft Corporation (Redmond, USA)
Flow cytometry	MXP software	Beckmann Coulter (Brea, USA)
Gel pictures	Quantity One	Bio-RAD (Hercules, USA)
Graphical analysis	GraphPad Prism	GraphPad Software, Inc. (La Jolla, USA)
qRT-PCR	Rotor Gene 6000 Series	QIAGEN (Hilden, Germany)
Picture analysis	Fiji	(Schindelin et al. 2012)
Running scripts	Python 3.5	The Python Software Foundation (Wilmington, USA)
Sequence visualization	Serial cloner 2.6	<a href="http://serialbasics.free.fr/Serial_Cloner.html">http://serialbasics.free.fr/Serial_Cloner.html</a>

TABLE 4-6: LIST OF OLIGONUCLEOTIDES

Name with internal numbering	Description	Sequence 5' to 3'
#36 cap mid	binds in the middle of <i>cap</i> of all serotypes except 4, 5, 12 and po1; sequencing primer	GAAATTGGCATTGCGATTCC
#37 cap mid	binds in the middle of <i>cap</i> of serotypes 4, 5, 12 and po1; sequencing primer	GATTGGCATTGCGATTCCAC
#176 DJ rev	rev primer, sequencing in whc background	GTCGCAAACACTCACGTG ACCTC
#459 SAfor	<i>cap</i> amplification after first PCR	GACTACAAGGACGACGATG ACAAG
#460 SArev	<i>cap</i> amplification after first PCR	CACTGAATTCTCATCAGGCG AAG
#822 LSeqFor	sequencing/ cloning primer for shuffled libraries	GATCTGGTCAATGTGGATTT G
#823 LSeqRev	sequencing/ cloning primer for shuffled libraries	GACCGCAGCCTTTTGAATG TC
#824 LSeqFor Nested	Primer for nested PCR during selection	ACTGCATCTTTGAACAATAA AT
#826 LSeqRev Nested	Primer for nested PCR during selection	GGTTTATTACTAGTGGCGCG
#827 M13for	for initial <i>cap</i> amplification	GTA AACGACGGCCAGTGA G

#828 M13Rev	for initial <i>cap</i> amplification	GGAAACAGCTATGACCATG
#833 Rep2For	Binds Rep2 within native HindIII site	AGACGCGGAAGCTTCGATC AA
#835 CMV for	Binds at the end of the CMV promoter und was used to sequence transgenes	CGCAAATGGGCGGTAGGCG TG
#983 AAP2 NotI For	Cloning primer for AAP2 with NotI site	ataagaatg <u>cgggccgc</u> ATGGAGAC GCAGACTCAGTAC
#984 AAP2 EcoRI Rev	Cloning primer for AAP2 with EcoRI site (and AAP3)	ccggaattcTCAGGGTGAGGTA TCCATAC
#985 AAP4 NotI For	Cloning primer for AAP4 with NotI site	ataagaatg <u>cgggccgc</u> ATGGAGCA GGCGACGGACCC
#986 AAP4 EcoRI Rev	Cloning primer for AAP4 with EcoRI site	ccggaattcTCACGTACGGCAG TTCGTAC
#987 AAP5 NotI For	Cloning primer for AAP5 with NotI site	ataagaatg <u>cgggccgc</u> ATGGACCC AGCGGATCCCAGC
#988 AAP5 EcoRI Rev	Cloning primer for AAP5 with EcoRI site	ccggaattcTCAGCGTCGCGTA ACCGTAC
#989 AAP8 NotI For	Cloning primer for AAP8 with NotI site	ataagaatg <u>cgggccgc</u> ATGGCGAC TCAGAGTCAGTTC
#990 AAP8 EcoRI Rev	Cloning primer for AAP8 with EcoRI site (AAP1, 6)	ccggaattcTCATGAACACGTC CGCCGGG
#991 AAP9 NotI For	Cloning primer for AAP9 with NotI site	ataagaatg <u>cgggccgc</u> ATGGCGAC ACAGAGTCAGTCC
#992 AAP9 EcoRI Rev	Cloning primer for AAP9 with EcoRI site	ccggaattcTCATGAAAACGTC CGCTGG
#1292 AAP1/3/6/7 NotI For	Cloning primer for AAP1/3/6/7 with NotI site	ataagaatg <u>cgggccgc</u> ATGGCGAC TCAGAGTCAGTCC
#1293 AAPrh10 NotI For	Cloning primer for AAPrh10 with NotI site	ataagaatg <u>cgggccgc</u> ATGGCGAC TCAGAGTCAGTGC
#1294 AAP7/ rh10 EcoRI Rev	Cloning primer for AAP7/ rh10 with EcoRI site	ccggaattcTCATGAAGACGTC CGCCGGG
#1295 pSSV9-YFP-for	Cloning primer for eYFP via NheI	atgcagtagtgcgcccaccATGGT GAGCAAGGGCGAGGA
#1296 pSSV9-YFP-rev	Cloning primer for eYFP via KpnI	tgcattggtaccactagtggcgccctT ACTTGTACAGCTCGTCCA
#1297 pSSV9-p(A)-for	Cloning primer for BGH poly(A) via KpnI	atgcaggtaccTAGAGCTCGCTG ATCAGCCT
#1298 pSSV9-p(A)-rev	Cloning primer for BGH poly(A) via NotI	tgcattg <u>cgggccgc</u> CTCCCAGCA TGCCTGCTAT
#1314 spacer-for	Insert spacer with additional SpeI/AflII site with Ascl and KpnI overhangs	cgccatcattactagtagatagtactta <u>cttaagccaattggtac</u>
#1315 spacer-rev	Insert spacer with additional SpeI/AflII site with Ascl and KpnI overhangs	Caattgg <u>cttaagtaag</u> tactatctactag taatgatgg
#1512 For_ccdbAscl	Amplifies ccdb cassette with Ascl site, for	agactagtggcgccACTGGCTG TGTATAAGGGAG
#1513 rev_ccdbPacl	Amplifies ccdb cassette with Pacl site, binds REV	agccttaattaaTCGCGTGGATC CGGCTTACT
#1515 miR122-for	Clone miR122 binding sites via ClaI	atgcaatcgatAGCGGCCGCAC AAAC
#1516 miR122-rev	Clone miR122 binding sites via Sall	tgcattg <u>tcgac</u> GCTCTAGATGGA GTG
#1538 periostinREV	To clone the periostin promoter via NheI	atcgatg <u>ctagc</u> TTCAGCCCTGA GCTCCG
#1539 pSSV9-cd11b-for	Cloning primer for the cd11b promoter via Pacl	atgcattaattaaTCGACGGTATC GATAAGCTT
#1540 pSSV9-cd11b-rev	Cloning primer for the cd11b	tgcattg <u>ctagc</u> GGATCCCGAGAA

	promoter via NheI	CCTGGAAG
#1543 VEC-for	Cloning primer for the VEC promoter via PacI	atgcatta <u>ata</u> TCGTCGACTCTAGCTAGTAG
#1544 VEC-rev	Cloning primer for the VEC promoter via SpeI	tgcat <u>actagt</u> ACCGGTGGATCTCGAGTCTG
#1784 periostin-for	Cloning primer for periostin via PacI	atcgatt <u>ata</u> CCTCTGACTCACTCTC
#1789 miR122-for	Clone miR122 binding sites via AvrII	atgcac <u>ctagg</u> CGCACAAACACCA
#1790 miR122-rev	Clone miR122 binding sites via Sall	gctctag <u>tcgacgctctt</u> ataataaAGATGGAGTGTGACA
#1791 miR142-3p-for	Clone miR142-3p binding sites via PacI	atgcatta <u>ata</u> GCTTGCGGCCGCTCC
#1792 miR142-3p-rev	Clone miR142-3p binding sites via Sall	tgcat <u>gtcgac</u> CAGCGGCCTGTAGTGTTC
2/indet*	for primer for AAPx02 as it started as AAP2 and shuffled to another parent	ataagaat <u>g</u> cgccgcATGGAGACGCAGACTCAGTCCC
AAP2ko_for	Cloning primer for AAP knockout with NotI site	ataagaat <u>g</u> cgccgcCGGAGACGCAGACTCAGTAC
AAP8ko_for	Cloning primer for AAP knockout with NotI site	ataagaat <u>g</u> cgccgcCCGGCGACTCAGAGTCAGTTC
AAP9ko_for	Cloning primer for AAP knockout with NotI site	ataagaat <u>g</u> cgccgcCCGGCGACACAGAGTCAGTCC
rescueBCnew	Binds rev, for PCR rescues with barcodes via NdeI	atcgat <u>catatg</u> CATATGGGAAGGACAGGAGGTCACCT
rescueBCori	Binds rev, for PCR rescues with barcodes via NdeI	atcgat <u>catatg</u> AAGCCCGTGGAAGGCTTCTAG
rescueBC289	Binds rev, for PCR rescues with barcodes via NdeI	atcgat <u>catatg</u> TTAGCCGGCACACCTCTAGG
rescueBC1	Binds rev, for PCR rescues with barcodes, contains SpeI site	atcgat <u>actagt</u> TCACGACGTCTGATGTGGAGC

Underlined are sites for restriction enzymes, small letters indicate stuffers/ overhangs for enzymes, CAPITAL letters indicate sequence stretches complementary to the template

TABLE 4-7: PLASMIDS USED

Name with internal number	Description	Origin
#68 pIRES(blast) FLAG/HA-hAgo2-GA	CMV driven human Argonaute2 expression plasmid including a FLAG/HA tag; acceptor for AAP expression	Cloned by Nina Schürmann
#182 WHc1(SpeI)	Cap helper plasmid with <i>rep2</i> and <i>cap 1</i>	Cloned by Eike Kienle
#183 WHc2(SpeI)	Cap helper plasmid with <i>rep2</i> and <i>cap 2</i>	Cloned by Eike Kienle
#184 WHc4(SpeI)	Cap helper plasmid with <i>rep2</i> and <i>cap 4</i>	Cloned by Eike Kienle
#185 WHc5(SpeI)	Cap helper plasmid with <i>rep2</i> and <i>cap 5</i>	Cloned by Eike Kienle
#186 WHc6(SpeI)	Cap helper plasmid with <i>rep2</i> and <i>cap 6</i>	Cloned by Eike Kienle
#187 WHc7(SpeI)	Cap helper plasmid with <i>rep2</i> and <i>cap 7</i>	Cloned by Eike Kienle
#188 WHc8(SpeI)	Cap helper plasmid with <i>rep2</i> and <i>cap 8</i>	Cloned by Eike Kienle
#189 WHc9(SpeI)	Cap helper plasmid with <i>rep2</i> and <i>cap 9</i>	Cloned by Eike Kienle
#190 WHcrh10(SpeI)	Cap helper plasmid with <i>rep2</i> and <i>cap rh10</i>	Cloned by Eike Kienle
#193 WHc-CapDJ	Cap helper plasmid with <i>rep2</i> and <i>cap DJ</i>	Cloned by Eike Kienle

#545 pBS-sds-decoy-empty	Self-complementary vector bearing ITRs with CMV-eGFP	Cloned by Stefan Mockenhaupt
#552 pBSUF3rev-YFP-sds	Self-complementary vector bearing ITRs with CMV-eYFP with SV40 enhancer	Cloned by Eike Kienle
#641 pDONR207	<i>ccdB</i> gene	Thermo Fisher Scientific (Waltham, USA)
#778 pSSV9_Pac_Asc	Single-stranded replication competent vector with ITRs, <i>rep2</i> and a stuffer for <i>cap</i> insertion	Cloned by Stefanie Große
#1080 pIRES(blast) FLAG-AAP2	AAP expression construct under control of CMV with HA/ FLAG tag (based on #68)	Cloned by Stefanie Große
#1081 pIRES(blast) FLAG-AAP4	AAP expression construct under control of CMV with HA/ FLAG tag (based on #68)	Cloned by Stefanie Große
#1082 pIRES(blast) FLAG-AAP5	AAP expression construct under control of CMV with HA/ FLAG tag (based on #68)	Cloned by Stefanie Große
#1109 pIRES(blast) FLAG-AAP8	AAP expression construct under control of CMV with HA/ FLAG tag (based on #68)	Cloned by Stefanie Große
#1110 pIRES(blast) FLAG-AAP9	AAP expression construct under control of CMV with HA/ FLAG tag (based on #68)	Cloned by Stefanie Große
#1111 Adeno helper plasmid	Plasmid required for AAV production, expresses factor required from adenovirus	(Matsushita et al. 1998)
#1134 WHC2mutAAP	Cap helper plasmid with <i>rep2</i> and <i>cap 2</i> bearing a knockout in AAP	Cloned by Stefanie Große
#1183 pSSV9-CMV-Cas9	Single-stranded vector with Cas9 under control of a CMV promoter	Cloned by Stefanie Große
#1189 WHC4GA	Cap helper plasmid with <i>rep2</i> and <i>cap 4</i> (original version, sequenced optimized to <i>cap2</i> ) with PacI site	Cloned by Stefanie Große
#1190 WHC5GA	Cap helper plasmid with <i>rep2</i> and <i>cap 5</i> (original version, sequenced optimized to <i>cap2</i> )	Cloned by Stefanie Große
#1194 WHc3(Spel)	Cap helper plasmid with <i>rep2</i> and <i>cap 3</i>	Cloned by Eike Kienle
#1290 WHC4GA_w/o PacI	Cap helper plasmid with <i>rep2</i> and <i>cap 4</i> (original version, sequenced optimized to <i>cap2</i> ) without PacI site	Cloned by Stefanie Große
#1291 WHC4GA repaired	During sequence optimization the AAP frame was not accounted for resulting in many mutations. Here, stop codons were removed and the start codon re-installed.	Cloned by Stefanie Große
#1292 WHC5GA repaired	During sequence optimization the AAP frame was not accounted for resulting in many mutations. Here, stop codons were removed and the start codon re-installed. Additionally, <i>rep</i> was mutated as well and was fixed here.	Cloned by Stefanie Große
#1293 WHC5GA only repSTOP	During sequence optimization the AAP frame was not accounted for resulting in many mutations. Here, only the mutation in <i>rep</i> was fixed (AAP is not fixed).	Cloned by Stefanie Große
#1326 WHC8mutAAP	Cap helper plasmid with <i>rep2</i> and <i>cap 8</i> bearing a knockout in AAP	Cloned by Stefanie Große

#1327 WHC9mutAAP	Cap helper plasmid with <i>rep2</i> and <i>cap 9</i> bearing a knockout in AAP	Cloned by Stefanie Große
#1452 pBSc1	Bluescript plasmid with <i>cap 1</i> gene flanked by <i>PacI</i> / <i>AscI</i> and oligo binding sites for DNA family shuffling	Cloned by Eike Kienle
#1453 pBSc2	Bluescript plasmid with <i>cap 2</i> gene flanked by <i>PacI</i> / <i>AscI</i> and oligo binding sites for DNA family shuffling	Cloned by Eike Kienle
#1454 pBSc3	Bluescript plasmid with <i>cap 3</i> gene flanked by <i>PacI</i> / <i>AscI</i> and oligo binding sites for DNA family shuffling	Cloned by Eike Kienle
#1455 pBSc4	Bluescript plasmid with <i>cap 4</i> gene flanked by <i>PacI</i> / <i>AscI</i> and oligo binding sites for DNA family shuffling	Cloned by Eike Kienle
#1456 pBSc5	Bluescript plasmid with <i>cap 5</i> gene flanked by <i>PacI</i> / <i>AscI</i> and oligo binding sites for DNA family shuffling	Cloned by Eike Kienle
#1457 pBSc6	Bluescript plasmid with <i>cap 6</i> gene flanked by <i>PacI</i> / <i>AscI</i> and oligo binding sites for DNA family shuffling	Cloned by Eike Kienle
#1458 pBSc7	Bluescript plasmid with <i>cap 7</i> gene flanked by <i>PacI</i> / <i>AscI</i> and oligo binding sites for DNA family shuffling	Cloned by Eike Kienle
#1459 pBSc8	Bluescript plasmid with <i>cap 8</i> gene flanked by <i>PacI</i> / <i>AscI</i> and oligo binding sites for DNA family shuffling	Cloned by Eike Kienle
#1460 pBSc9	Bluescript plasmid with <i>cap 9</i> gene flanked by <i>PacI</i> / <i>AscI</i> and oligo binding sites for DNA family shuffling	Cloned by Eike Kienle
#1461 pBScr10	Bluescript plasmid with <i>cap rh10</i> gene flanked by <i>PacI</i> / <i>AscI</i> and oligo binding sites for DNA family shuffling	Cloned by Eike Kienle
#1494 pBSc4SO repaired	Bluescript plasmid with <i>cap 4</i> sequence-optimized gene with repaired AAP flanked by <i>PacI</i> / <i>AscI</i> and oligo binding sites for DNA family shuffling	Cloned by Stefanie Große
#1495 pBSc5SO repaired	Bluescript plasmid with <i>cap 5</i> sequence-optimized gene with repaired AAP flanked by <i>PacI</i> / <i>AscI</i> and oligo binding sites for DNA family shuffling	Cloned by Stefanie Große
#1500 WH_empty_Hind/Spe	Cap helper plasmid with <i>rep2</i> and a stuffer instead of <i>cap</i> allowing cloning (via <i>HindIII</i> / <i>SpeI</i> )	Cloned by Stefanie Große
#1544 whc_ccdb_PacI_AscI	Cap helper plasmid with <i>rep2</i> and a <i>ccdb</i> cassette instead of <i>cap</i> allowing cloning (via <i>PacI</i> / <i>AscI</i> )	Cloned during this study
#1608 pSSV9_Pac_Asc_ccdB	Comparable to #778 with a <i>ccdb</i> cassette instead of a stuffer	Cloned during this study
#1661 WHC1mutAAP	Cap helper plasmid with <i>rep2</i> and <i>cap 1</i> bearing a knockout in AAP	Cloned by Stefanie Große
#1662 WHC3mutAAP	Cap helper plasmid with <i>rep2</i> and <i>cap 3</i> bearing a knockout in AAP	Cloned by Stefanie Große
#1663 WHC6mutAAP	Cap helper plasmid with <i>rep2</i> and <i>cap 6</i> bearing a knockout in AAP	Cloned by Stefanie Große
#1664 WHC7mutAAP	Cap helper plasmid with <i>rep2</i> and <i>cap 7</i> bearing a knockout in AAP	Cloned by Stefanie Große
#1665 WHCrh10mutAAP	Cap helper plasmid with <i>rep2</i> and <i>cap rh10</i> bearing a knockout in AAP	Cloned by Stefanie Große
#1666 pIRES(blast) FLAG-	AAP expression construct under control	Cloned by Stefanie Große

---

AAP1	of CMV with HA/ FLAG tag (based on #68)	
#1667 pIRES(blast) FLAG-AAP3	AAP expression construct under control of CMV with HA/ FLAG tag (based on #68)	Cloned by Stefanie Große
#1668 pIRES(blast) FLAG-AAP6	AAP expression construct under control of CMV with HA/ FLAG tag (based on #68)	Cloned by Stefanie Große
#1669 pIRES(blast) FLAG-AAP7	AAP expression construct under control of CMV with HA/ FLAG tag (based on #68)	Cloned by Stefanie Große
#1670 pIRES(blast) FLAG-AAPrh10	AAP expression construct under control of CMV with HA/ FLAG tag (based on #68)	Cloned by Stefanie Große
#1824 pIRES(blast) FLAG-AAP DJ	AAP expression construct under control of CMV with HA/ FLAG tag (based on #68)	Cloned by Stefanie Große
#1913 pSSV9-CMV-YFP-spacer	Single-stranded vector with an eYFP under the control of a CMV promoter bearing a spacer with additional restriction sites between the eYFP and the poly(A)	Cloned during this study
#1914 pSSV9-VEC-YFP	Same as #1913 with the VEC promoter presumably mediating specificity in LSECs	Cloned during this study
#1915 pSSV9-cd11b-YFP	Same as #1913 with the cd11b promoter presumably mediating specificity in Kupffer cells	Cloned during this study
#2006 sc-eYFP-miR122	Self-complementary eYFP driven by CMV promoter bearing 4x miR122 binding sites	Cloned during this study
#2007 sc-eGFP-miR122	Based on #545 with 4x miR122 binding sites	Cloned during this study
#2008 sc-eGFP-miR122-miR142-3p	Based on #2008 bearing two additional miR142-3p binding sites	Cloned during this study
#2009 pSSV9-VEC-eGFP	Single-stranded eGFP reporter vector driven by the supposedly LSEC specific promoter VEC	Cloned during this study
#2010 pSSV9-cd11b-eGFP	Single-stranded eGFP reporter vector driven by the supposedly Kupffer cell-specific promoter cd11b	Cloned during this study
#2022 pSSV9-periostin-eGFP	Single-stranded eGFP reporter vector driven by the supposedly stellate cell-specific promoter periostin	Cloned during this study
Cd11b vector	Construct bearing the cd11b promoter, kindly provided by Antonia Follenzi	(Simard et al. 2006)
pBK06	Luciferase vector with additional miR122 binding sites	Cloned by Ben Kachel
pBK76	Luciferase vector with additional miR142-3p binding sites	Cloned by Ben Kachel
1.2-kb peri-lacZ	Construct bearing the truncated version of the periostin promoter, kindly provided by Holger Willenbring	(Lindsley et al. 2007)
VEC vector	Construct bearing the VEC promoter, kindly provided by Antonia Follenzi	(Gory et al. 1999)

---

## 4.2. METHODS

---

### 4.2.1. MOLECULAR BIOLOGICAL METHODS

---

#### 4.2.1.1. POLYMERASE CHAIN REACTION (PCR)

---

In order to amplify DNA for further subcloning the phusion HS II polymerase was used. In one reaction, 10  $\mu$ l GC or HF buffer, 1  $\mu$ l dNTPs (10 mM), 0.5  $\mu$ l polymerase, 1.5  $\mu$ l DMSO, 1.5  $\mu$ l forward primer (10  $\mu$ M), 1.5  $\mu$ l reverse primer (10  $\mu$ M) and 0.2  $\mu$ l template plasmid were mixed and filled up to 50  $\mu$ l with water. Usually, an initial heating step of 5 min at 95°C was performed followed by 40 cycles of denaturation (95°C, 15 s), annealing (50-60°C, 15 s) and elongation (72°C, 30 s per 1 kb). The annealing temperature was usually set to 56°C and further modified by gradient PCRs if necessary. Finally, a 10 min elongation step at 72°C was included with indefinite hold at 4°C.

A colony PCR was performed to test colonies after transformation for successful insert integration. Therefore, 25  $\mu$ l Taq 2x mastermix were mixed with 0.2  $\mu$ l forward primer (10  $\mu$ M), 0.2  $\mu$ l reverse primer (10  $\mu$ M) and 4.6  $\mu$ l water. A colony was picked from the plate with a pipet tip, streaked out on a new plate and used to stir within the PCR tube containing 10  $\mu$ l of the mix. After an initial heating step of 95°C for 5 min, 30 cycles of heating (95°C, 15 s), annealing (50-60°C, 15 s) and elongation (68°C, 1 min per 1 kb) followed. The run was finalized by an elongation step for 10 min at 68°C and kept indefinitely on hold at 4 °C.

In any case, the PCR product was separated on a 1% (w/v) agarose gel, at 80-120 V for 30 to 50 min and DNA was visualized by intercalating ethidium bromide under UV light. If necessary, a 6x loading buffer was added to the DNA prior loading. PCR products required for further cloning procedures were excised from the gel using a scalpel and samples were stored at -20°C.

#### 4.2.1.2. RESTRICTION DIGESTION

---

For subcloning purposes 20  $\mu$ l plasmid backbone were digested with 2  $\mu$ l of each appropriate restriction enzyme, 4  $\mu$ l of 10x buffer and filled up to 40  $\mu$ l with water. Digestion occurred depending on the enzymes' requirements usually with 10x CutSmart buffer at 37°C either over day for a couple of hours or overnight. In case of PCR products, a similar strategy was applied concerning the volume but digestion took place

over night. For all reactions the amount of total enzyme did not exceed one tenth of the total volume. If the digest of the vector yielded two differently sized fragments they had to be separated on a 1% (w/v) agarose gel and the appropriate band was extracted.

Test digestions were performed to either test the successful integration of an insert and/or the integrity of the ITRs. For this, 1  $\mu$ l 10x buffer was mixed with 0.25  $\mu$ l of each required enzyme and 0.5-2  $\mu$ l plasmid DNA was filled up to a total volume of 10  $\mu$ l with water. Usually, 0.5  $\mu$ l plasmid was used but if small digestion products or a low concentration was expected the amount was increased. After incubation for 1 h at the appropriate temperature the test digests were analyzed on a 1% (w/v) agarose gel. In order to test for insert integration the enzymes already used for cloning were used. In case of ITR integrity XmaI was used for ITR2 originating from AAV2 and PstI for ITR4 originating from AAV4. The ITRs were regarded as intact if the *in silico* predicted patterns matched the real one.

---

#### 4.2.1.3. LIGATION

For standard cloning purposes a ligation mix was set up containing 0.5  $\mu$ l T4 ligase, 1  $\mu$ l 10x ligase buffer, insert, vector and water filled up to a total volume of 10  $\mu$ l. Usually, an insert to vector ratio of 3:1 was used with a total DNA amount of 100 ng. In case of small fragments like miRNA binding sites or small oligonucleotides a 7:1 ratio was applied. The reaction occurred either over night at 16°C or over day for at least 3 h at room temperature.

---

#### 4.2.1.4. DNA ISOLATION

DNA purification from a 1% (w/v) agarose gel was performed using the QIAquick Gel Extraction Kit according to manufacturer's recommendation. The optional isopropanol step was always included as well as an additional washing step with buffer QG and a prolonged "empty" spin for 3 min instead of 1 min. For elution 30  $\mu$ l water was used for PCR products and 50  $\mu$ l for vectors. This was incubated 5 min at room temperature prior spinning. In case for the generation of initial and selected AAV libraries the QIAquick PCR purification kit was used. Again, instructions were followed with the prolonged empty spin and elution in 30  $\mu$ l for 5 min at room temperature prior spinning.

Usually, digested PCR products as well as digested vectors where only a small stuffer below 50 bp fell out were purified via the DNA clean & concentrator-5 kit according to

manufacturer's instructions. Vectors were eluted in 40  $\mu$ l water and PCR products in 20  $\mu$ l water.

After subcloning colonies were inoculated to isolate plasmid DNA the next day. In most cases a so called "dirty" mini protocol was applied with lab-made buffers P1, P2 and P3. Buffers P1 and P3 were stored at 4°C whereas P2 stayed at room temperature. For this, 2 ml of the culture were pelleted in a 2 ml tube at maximum speed for 2 min in a table top centrifuge. The supernatant was discarded and the pellet was resuspended in 300  $\mu$ l resuspension buffer P1. Next, 300  $\mu$ l lysis buffer P2 was added, tubes were gently inverted and incubated for 5 min at room temperature. For neutralization 300  $\mu$ l neutralization buffer P3 was added and gently inverted. After an incubation time of 5 min at room temperature the samples were spun down at max speed for 10 min. Carefully 800  $\mu$ l of the clear supernatant was transferred to a new tube containing 600  $\mu$ l isopropanol. The DNA was then pelleted at maximum speed for 15 min. The supernatant was discarded and an additional washing step with 70% ethanol was performed for 5 min at maximum speed. Again, the supernatant was discarded and remaining ethanol was removed carefully by pipetting. The pellet was dried and then eluted in 50  $\mu$ l water. Sometimes, the cultures were inoculated in the early morning, vortexed and then DNA was isolated in the afternoon. In such cases elution occurred in 40  $\mu$ l. Mini preps originating from the "dirty" prep protocol were used for sequencing purposes and for further subcloning. For small-scale transfections "clean" mini preps were prepared by using the QIAgen miniprep kit according to manufacturer's instructions with elution in 50  $\mu$ l water.

If larger plasmid DNA amounts were required a midi (80 ml) or maxi (300 ml) culture were inoculated and pelleted the next day at 4000 x g, 10 min at room temperature. For midi prep the DNA was isolated according to manufacturer's instruction using the PureYield™ Plasmid Midiprep kit. The recommended drying step of 30 s to 1 min was prolonged to 5 min in order to reduce ethanol contamination. Elution occurred for 5 min with 1 ml water. Similarly, for maxi prep the manual of the NucleoBond® Xtra Midi / Maxi kit was followed with the instructions for maxi preparations. Usually, 1.5 ml was used for elution.

---

## 4.2.2. PROKARYOTIC METHODS

---

TABLE 4-8: *ESCHERICHIA COLI* STRAINS USED

Strain	Competence	Vendor
<i>E. coli</i> ccdB Survival <sup>1</sup> ™ T1 <sup>R</sup>	Chemically competent	Thermo Fisher Scientific (Waltham, USA)
<i>E. coli</i> MAX Efficiency DH5 $\alpha$ <sup>™</sup>	Chemically competent	Thermo Fisher Scientific (Waltham, USA)
<i>E. coli</i> MegaX DH10B <sup>1</sup> ™ T1 <sup>R</sup>	Electrocompetent	Thermo Fisher Scientific (Waltham, USA)

---

### 4.2.2.1. CULTIVATION CONDITIONS

---

Bacteria were usually cultured in LB medium carrying the respective antibiotic resistance. If bacteria were cultivated on LB plates, 1.5% (w/v) agar was added to the medium and the antibiotic was added shortly before pouring of the plates. All media and plates were stored at 4°C until usage. For liquid cultures, glass culture tubes were used for mini cultures, 250 ml flasks for midi culture and 2 l flasks for maxi preps. Mini cultures were usually inoculated in 3 ml medium by picking a clone from a plate with a pipet tip and the larger formats used a 1/1000 volume of a preculture. Incubation occurred at 37°C shaking at 180 rpm. Bacteria plated out on plates were incubated at 37°C overnight in an incubator. In some cases a glycerol stock was prepared by mixing 800  $\mu$ l bacteria from a fresh culture with 200  $\mu$ l glycerol. After careful inversion the tube was frozen and stored at -80°C.

### 4.2.2.2. ELECTROCOMPETENT BACTERIA: PRODUCTION AND ELECTROPORATION

---

For the production of electrocompetent bacteria 30 ml LB medium without antibiotics were inoculated with MegaX DH10B<sup>™</sup> T1<sup>R</sup> from the glycerol stock under sterile conditions with the help of a Bunsen burner. The next day four 2 L flasks with 400 ml antibiotic free medium were inoculated with 5 ml of the overnight culture. The bacteria were grown for 2-4 h until an OD<sub>600</sub> of 0.5-0.55 was reached. Then they were transferred into centrifuge bottles and incubated on ice for 15 min. While the cells were pelleted for 10 min at 4000 x g at 4°C four to five dialysis tubings were prepared. They were cut to a length of 20-25 cm and were put into a beaker containing enough water to cover them. They were cooked in the microwave until the water started boiling and were allowed to cool down again. After centrifugation the cell pellets were resuspended in a total volume

of 60 ml sterile water which was pre-cooled. The bacteria were transferred into the dialysis tubings, closed with clips and dialyzed overnight in pre-cooled water at 4°C slightly stirring. The next day the cells were transferred into 50 ml tubes and spun down for 10 min at 4000 x g at 4°C. All cells were resuspended in 900 µl 10% (v/v) glycerol and the OD<sub>600</sub> was determined. The bacteria were diluted with 10% (v/v) glycerol until a theoretical OD<sub>600</sub> of one was reached by calculating OD<sub>600 measured</sub> x 900 µl equals the end volume. The cells were aliquoted and snap-frozen in liquid nitrogen. To test the efficiency 1.5 µl pUC19 vector were electroporated according to manufacturer's instructions and efficiencies in the range of 5x10<sup>8</sup> colony forming units per µg DNA and higher were accepted for library productions.

For one electroporation, 30 µl of electrocompetent bacteria were thawed on ice and 2 µl of the ligation reaction were added. This was then transferred into a pre-cooled electroporation cuvette and inserted into the device. A pulse was initiated with 1800 V, 25 µF and 200 Ω. For recovery 1 ml of pre-warmed SOC medium without antibiotic was added to the cuvette, the bacteria were transferred into a 1.5 ml tube and shaken for 1 h at 600 rpm in a heating block. In case of regular cloning procedures, the bacteria were spun for 3 min at 4000 x g, the supernatant was poured off and the pellet was resuspended in the remaining liquid. This was then plated out. If a library was subcloned all electroporations of one library were pooled and 100 µl were plated out from each, undiluted, 1:10 diluted and 1:100 diluted sample, respectively. The remainder was used to inoculate a maxi culture.

#### 4.2.2.3. CHEMOCOMPETENT BACTERIA: PRODUCTION AND TRANSFORMATION

---

The MAX Efficiency DH5α<sup>TM</sup> were streaked out on an LB-plate without antibiotic from the glycerol stock and one single colony was used to inoculate 6 ml LB medium. The day after, three flasks with 200 ml SOC medium were inoculated with 1 ml of the overnight culture and were grown until an OD<sub>600</sub> of 0.5. The bacteria were pelleted for 15 min with 1800 x g at 4°C and were resuspended in 40 ml TFB I buffer. After 10 min incubation on ice the cells were pelleted again and were resuspended in 10 ml TFB II buffer. Again, the bacteria were incubated on ice for 10 min and were then aliquoted into 50 µl portions, snap frozen in liquid nitrogen and stored at -80°C.

For standard cloning purposes chemocompetent bacteria were thawed on ice and 50 µl bacteria were added to 10 µl of the ligation reaction. This mixture was left on ice for 20 min followed by a heat shock at 42°C for 45 s and an incubation on ice for two more minutes. The complete reaction was plated out. If a plasmid containing a ccdB cassette

was to be transformed, the *ccdb* competent bacteria were used. In all other cases it was DH5 $\alpha$ .

### 4.2.3. CLONING OF NEW CONSTRUCTS

#### 4.2.3.1. Generation of *ccdb* containing plasmids

To facilitate the transfer from a chimeric capsid from the replication competent backbone pSSV9 (Figure 4-1 A) into the helper context (Figure 4-1 B) a helper plasmid with *PacI* and *AsclI* sites was generated. For this a chimeric variant was transferred via *HindIII* and *SpeI* from the pSSV9 backbone into a helper construct. As the *PacI* and *AsclI* site flanked the chimeric *cap* they were transferred as well (Figure 4-1 C). In order to improve transformation, the chimeric *cap* gene was replaced by the suicide gene *ccdb*. This gene is toxic for regular *E. coli* and can only be propagated in special *E. coli* *ccdb* resistant strains e.g. *E. coli* *ccdb* Survival<sup>TM</sup> T1<sup>R</sup> strain. As a consequence bacteria who took up undigested or partially digested backbone will die. The *ccdb* cassette was PCR amplified using primers #1512 and #1513 on the template plasmid #641 and replaced the chimeric *cap* by cloning via *PacI* and *AsclI*.

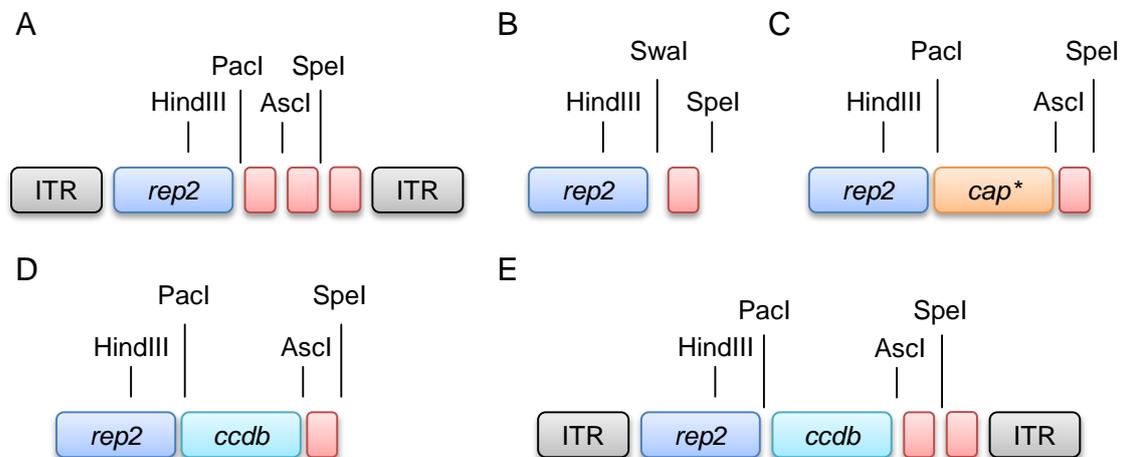


FIGURE 4-1: CCDB CONSTRUCTS

(A) The replication competent construct #778 is shown being characterized by the inverted terminal repeats (ITR) in combination with *rep2* and cloning sites for *cap* transfer. (B) The helper construct #1500 is depicted. This backbone is used to subclone a chimeric capsid by *Swal* and *SpeI* restriction sites to package transgenes into a chimeric capsid. (C) To facilitate cloning a chimeric capsid present in the replication competent context was subcloned using *HindIII* and *SpeI* sites thereby introducing *PacI* and *AsclI*. (D, E) To reduce the amount of false positive clones a *ccdb* construct was first introduced into the helper context (D) and was then transferred to #778 (E). Red boxes indicate stuffer between restriction enzymes.

This generated construct #1544 (Figure 4-1 D). The ccdB cassette was transferred to the replication competent background as well by PacI and AsclI digestion of both, the donor plasmid #1544 and the acceptor plasmid #778, thereby yielding construct #1608 (Figure 4-1 E).

#### 4.2.3.2. CLONING OF AAP EXPRESSION PLASMIDS

Usually, the AAP was PCR amplified from the capsid of interest by matching oligonucleotides. Therefore, the N-terminal and C-terminal part was tested *in silico* for primer binding depending on the serotype origin. These regular AAP expression constructs applied primers #983 - #992, #1292 - #1294 and 2/indet\* in the case were none of the primers fit. The resulting PCR product was subcloned via EcoRI and NotI restriction sites into the acceptor plasmid #68 (Figure 4-2 A). Note that the endogenous start codon "CTG" was changed to "ATG". To confirm the successful knockout of AAP within the knockout constructs, mutated AAP was cloned out using primers AAP2ko\_for / #984, AAP8ko\_for / #990 and AAP9ko\_for / #992 on plasmid #1134, #1326 and #1327, respectively (Figure 4-2 B). The PCR product was subcloned as described above. The knockout is based on a mutated start codon (CTG → CCG) and the introduction of an early stop codon (TGG → TAG).

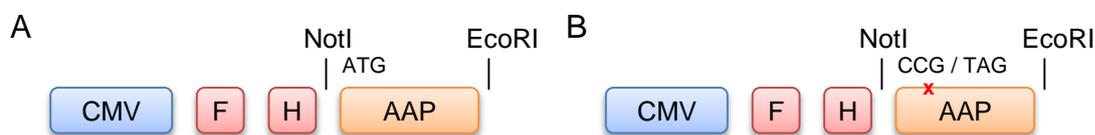


FIGURE 4-2: AAP EXPRESSION CONSTRUCT DESIGN

AAP can be cloned via NotI / EcoRI restriction site under the control of a CMV promoter. In addition, AAP bears a FLAG (F) / HA (H) tag. (A) The endogenous CTG start codon was altered to the conventional ATG. (B) In the AAP knockout construct the start codon was changed to CCG and a stop codon (TAG) was introduced as indicated by the red cross.

#### 4.2.3.3. BARCODE RESTRICTION DURING SELECTION

The aim is to introduce so called barcodes into the replication competent backbone pSSV9 (#778) at two positions thereby being able to track a library and to optimize the PCR rescue further. The barcodes are 20 nucleotide stretches designed to already bear the corresponding sticky ends for ligation. At first, two matching oligonucleotides (Table 4-9), 5  $\mu$ l forward (100  $\mu$ M) and 5  $\mu$ l reverse (100  $\mu$ M), were mixed with 40  $\mu$ l H<sub>2</sub>O and incubated at 95°C for 5 min. The heat block was turned off and allowed to slowly cool

down to room temperature. A ligation mixture was set up as described 4.2.1.3 for standard cloning purposes with 1  $\mu$ l digested acceptor backbone and 2  $\mu$ l annealed oligos. Transformation occurred into chemocompetent bacteria (4.2.2.3). Clones were picked for “dirty” mini prep and assessed for ITR integrity (4.2.1.4, 4.2.1.2.). Positive clones were validated by sequencing using primer #822.

TABLE 4-9: BARCODE ANNEALING

Name	For oligo 5' to 3'	Rev oligo 5' to 3'
introNdeI	<u>cgcgcc</u> tactataagc <u>atatg</u> agatagtactacta	<u>ctagtag</u> taagtactatct <u>atatg</u> cttatagtacgg
BC289	<u>cgcgcc</u> CCTAGAGGTGTGCCGGCTAA <u>ca</u>	<u>tatg</u> TTAGCCGGCACACCTCTAGG <u>gg</u>
BCori	<u>cgcgcc</u> CTAGAAGCCTCCACGGGCTT <u>ca</u>	<u>tatg</u> AAGCCCGTGGAGGCTTCTAG <u>gg</u>
BCnew	<u>cgcgcc</u> GGTTGCGGTGCGATTATGCG <u>ca</u>	<u>tatg</u> CGCATAATGCGACCGCAACC <u>gg</u>
BCred	<u>cgcgcc</u> GGCAATACTCGCAAGCCTGG <u>ca</u>	<u>tatg</u> CCAGGCTTGGGAGTATTGCC <u>gg</u>
BC1	<u>tatg</u> GCTCCACATCGACGTCGTGA <u>a</u>	<u>ctagt</u> TCACGACGTCGATGTGGAGC <u>ca</u>

Underlined are sites for restriction enzymes, small letters indicate stuffers/ overhangs for enzymes, CAPITAL letters indicate barcode sequence

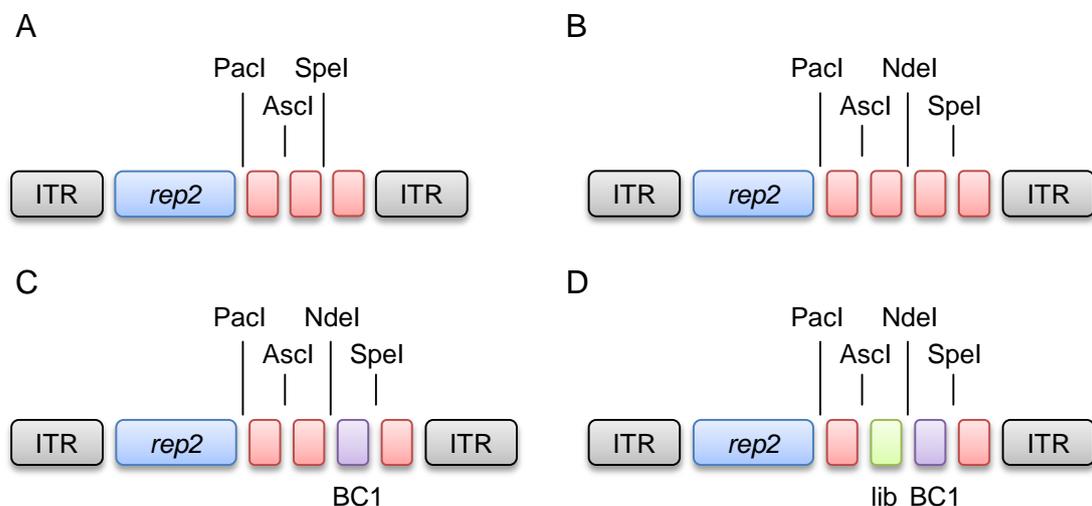


FIGURE 4-3: CONSTRUCTS FOR BARCODE SELECTION

The schematic depicts the construct design for replication competent backbones where libraries as well as selection rounds can be tagged with a barcode. (A, B) Into the initial #778 backbone (A) an NdeI site was integrated (B). (C) The stuffer between NdeI and Spel can then be replaced by a barcode tagging the selection round. (D) This construct can then be used to insert barcodes specific for a given library. Red squares indicate stuffers to separate restriction sites, the purple square (BC1) indicates the selection round barcode, the green square (lib) barcode. ITR, inverted terminal repeats,

The cloning of the barcodes is based on an NdeI site which needed to be introduced into #778 first (Figure 4-3 A). The annealed oligos for “introNdeI” were ligated into Ascl and SpeI digested #778 resulting in the vector depicted in Figure 4-3 B. Next, the annealed oligos for BC1 were introduced via NdeI and SpeI (Figure 4-3 C). This backbone bearing the BC1 barcode was used to insert four different library barcodes by using BC289, BCori, BCred and BCnew (Figure 4-3 D).

#### 4.2.3.4. CONSTRUCTS TO MEDIATE LIVER CELL-TYPE SPECIFICITY

---

The details for all constructs mentioned in this paragraph are summarized in Table 4-10 and are depicted in Figure 4-4 . The cell-type specific promoters used in this study do not fit into the self-complementary vectors due to their size. Therefore, a single-stranded acceptor backbone containing a CMV promoter with eYFP was cloned first. The eYFP was PCR amplified and cloned into the acceptor backbone via NheI and KpnI restriction sites, thus generating the intermediate pSSV9-CMV-eYFP. In a next step, a poly(A) tail was inserted via KpnI and NotI thereby creating the cloning intermediate pSSV9-CMV-eYFP-bgh. At last, a spacer was inserted between the eYFP and the poly(A) tail by using the restriction enzymes Ascl and KpnI bearing additional restriction sites which allow for insertion of miRNA binding sites. This plasmid (#1913) is now the acceptor for different promoters. Two promoters, cd11b and VEC, were PCR amplified and subcloned into #1913. The CMV was replaced by PacI and NheI in case for cd11b and due to an internal NheI site within VEC by PacI and SpeI for VEC, respectively. Note that the same digested backbone was used as the cohesive ends from SpeI and NheI are compatible to ligate to each other but the original NheI site gets thereby destroyed. At the end, a Kupffer cell-specific vector encoding for eYFP (#1914) and a LSEC specific vector encoding eYFP (#1915) were generated. To improve validation the reporter fluorophore was exchanged to eGFP. Complementary to the initial strategy with eYFP, eGFP was PCR amplified and got inserted by NheI and Ascl in case of cd11b resulting in #2010. Since the NheI site got destroyed upon VEC insertion (#1915) AgeI was used as an alternative site thereby creating construct #2009. In addition, a promoter supposedly working well in stellate cells *in vivo* was acquired, PCR amplified and inserted into #2010 by PacI and NheI digestion yielding construct #2022. As an alternative strategy a self-complementary vector was prepared bearing miRNA binding sites to mediate degradation in a cell-type specific manner. As before, four miRNA122 binding sites were introduced into a self-complementary eYFP vector (#552) by PCR amplification and restriction by ClaI and Sall yielding plasmid #2006. For the same reason as above, the fluorophore for analysis should be eGFP. Therefore, the miR122 binding sites were

introduced into #545 by *AvrI* and *SalI* resulting in vector #2007. During that step a *PacI* site was introduced allowing insertion of two miR142-3p binding sites by *PacI* and *SalI* digestion creating construct #2008.

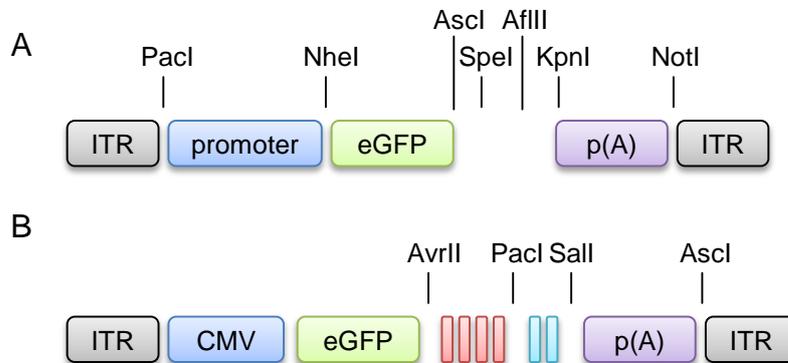


FIGURE 4-4: CELL TYPE SPECIFIC VECTOR DESIGN

The construct design is depicted to mediate cell-type specificity. (A) The promoter of the single-stranded vectors can be exchanged by *PacI* and *NheI* restriction sites. Between the fluorescent reporter eGFP and the poly(A) tail (p(A)) a variety of restriction enzymes is present to allow incorporation of miRNA binding sites. (B) The self-complementary vector bears a CMV driven eGFP. Two different miRNA binding sites namely for miR122 (red) and miR142-3p (light blue) were incorporated between the fluorophore and the poly(A) indicated by the narrow boxes.

TABLE 4-10: CLONING STRATEGY FOR CELL-TYPE SPECIFIC CONSTRUCTS

Final vector	Insert source	Acceptor backbone
pSSV9-CMV-eYFP	PCR on plasmid #552 with primers #1295/ #1296	#1183
pSSV9-CMV-eYFP- bgh	PCR on plasmid #552 with primers #1297/ #1298	pSSV9_CMV_eYFP
#1913	Spacer was created by oligo annealing of primers #1314/ #1315	pSSV9-CMV-eYFP- bgh
#1914	PCR on VEC vector with primers #1539/ #1540	#1913
#1915	PCR on Cd11b vector with primers #1543/ #1544	#1913
#2009	PCR on #545 with primers #835/ #1296	#1915
#2010	PCR on #545 with primers #1295/ #1296	#1914
#2022	PCR on 1.2-kb peri-lacZ with primers #1784/ #1538	#2010
#2006	PCR on pBK06 with primers #1515/ #1516	#552
#2007	PCR on #2006 with primers #1789/ #1790	#545
#2008	PCR on pBK76 with primers #1791/ #1792	#2006

---

#### 4.2.4. LIBRARY GENERATION BY DNA FAMILY SHUFFLING

---

##### 4.2.4.1. AMPLIFICATION OF CAPSID GENES

---

In a very first step, the *cap* genes of the required serotypes were PCR amplified from a plasmid backbone. For this, 10  $\mu$ l 5x Hifi buffer was mixed with 1  $\mu$ l M13For primer (10  $\mu$ M), 1  $\mu$ l M13Rev primer (10  $\mu$ M), 2  $\mu$ l Hifi polymerase, 200 ng template and water to a final volume of 50  $\mu$ l. After an initial step of 5 min at 95°C, 40 cycles of heating for 15 s at 94°C, annealing for 30 s at 57°C and elongation for 3 min at 68°C followed with a final elongation step at 72°C for 10 min. Finally, the cycler kept the samples at indefinite hold at 4°C. Usually, two to four reactions per *cap* gene were set up in parallel and were separated together in one well on a 1% (w/v) agarose gel and the resulting 2.5 kb band was excised, purified as described in 4.2.1.4 and eluted in 60  $\mu$ l water.

##### 4.2.4.2. FRAGMENTATION BY DNASE I

---

The DNA concentration was determined by the QuBit system which is based on a fluorophore intercalating into double-stranded DNA making it more accurate for this method compared to determination by spectrophotometric devices such as the nanodrop. For one reaction, 4  $\mu$ g total DNA was required and equal amounts of each capsid DNA were mixed with 6  $\mu$ l 10x DNase I buffer and water to a final volume of 59.7  $\mu$ l. Two heating blocks with 25°C and 75°C were prepared. After addition of 0.3  $\mu$ l DNase I to the sample the tube was flicked until no glycerol smear was visible anymore followed by incubation at 25°C for 30 s to 2 min. The incubation time is dependent on the activity of the DNase I and was experimentally determined. In order to inactivate the enzyme, 6  $\mu$ l EDTA (25 mM) was added, the sample was vigorously vortexed and incubated at 75°C for 10 min. After addition of loading dye, the sample was separated on a 1% (w/v) agarose gel and the resulting smear from 100 bp to 1 kb was excised, purified as described in 4.2.1.4. and eluted in 60  $\mu$ l water.

##### 4.2.4.3. FRAGMENTATION BY COVARIS

---

Similarly to DNase I-based fragmentation, 4  $\mu$ g of capsid DNA was mixed and filled up with water to 60  $\mu$ l. The DNA was transferred into a Covaris reaction tube and inserted into the Covaris focused-ultrasonicator S2. A program described previously ((Große 2016), Table 4-11) was chosen to generate fragments of 150 bp, 300 bp and 800 bp in

size. The samples were separated on a 1% (w/v) agarose gel and corresponding sizes were excised and processed as described in the section above.

TABLE 4-11: COVARIS FRAGMENTATION CONDITIONS

Fragment size [bp]	150	300	800
Time [s]	500	120	80
Cycles per burst	200	200	200
Duty cycle [%]	10	10	5
Intensity	5	4	3

#### 4.2.4.4. RE-ASSEMBLY AND NESTED PCR

In the first PCR 500 ng fragmented *cap* DNA was mixed with 10  $\mu$ l 5x HF buffer, 1  $\mu$ l dNTPs (10 nM), 0.5  $\mu$ l phusion HS II polymerase, 1.5  $\mu$ l DMSO and water to a final volume of 50  $\mu$ l. This PCR is primerless as the *cap* fragments will partially re-anneal due to homology which can then be filled up by the polymerase. The PCR starts with an initial step at 98°C for 30 s which is then followed by 40 cycles of heating at 98°C for 10 s, annealing at 42°C for 30 s and elongation at 72°C for 45 s. Finally, an elongation step for 10 min at 72°C occurs followed by indefinite hold at 4°C. For the second PCR nested primers were used. Here, 2  $\mu$ l of the first PCR is used as input material together with 1  $\mu$ l SaFor (10  $\mu$ M), 1  $\mu$ l SaRev (10  $\mu$ M), 0.5  $\mu$ l 25 mM MgSO<sub>4</sub>, 2  $\mu$ l Hifi polymerase, 10  $\mu$ l 5x Hifibuffer and 33.5  $\mu$ l water. The PCR was run at 95°C for 5 min followed by 40 cycles of a two-step protocol with 94°C heating for 15s and 68°C annealing/ elongation for 3 min finalized by 10 min 72°C and indefinite hold at 4°C. Usually, one test reaction was being run to see if reassembly was successful. If this was the case, the reaction was upscaled to 20 reactions which were either separated by a 1% (w/v) agarose gel followed by subsequent purification from the gel using one column only or by direct column purification. In either way, the samples were eluted in 30  $\mu$ l water and 4  $\mu$ l 10x CutSmart buffer, 2  $\mu$ l H<sub>2</sub>O, 2  $\mu$ l PacI and 2  $\mu$ l Ascl were added for overnight digestion at 37°C. The next day, the digest was purified over a 1% (w/v) agarose gel as described earlier. In parallel, the replication competent acceptor backbone #778 or #1608 was digested along. In case of the comparison of different fragmentation methods the taqmix was used for the second PCR using 25  $\mu$ l 2x taqmix, primers as above and filled up to a total volume of 50  $\mu$ l. The cycling program run for 5 min at 95°C followed by 40 cycles of

heating for 15 s at 95°C, annealing for 15 s at 55°C and elongation for 3 min at 68°C. After a final elongation step for 10 min at 68°C the samples were kept on 4°C.

#### 4.2.4.5. GENERATION OF THE PLASMID LIBRARY

---

A ligation mixture in a total volume of 40 µl with 10x ligase buffer, 4 µl T4 ligase, 861 ng vector and 1139 ng insert representing a 1:3 ratio was prepared and incubated overnight at 16°C. Electrocompetent bacteria were thawed on ice and electroporated as described in 4.2.2.2. Per library, 20 electroporations were performed. The next day, a glycerol stock was prepared and the remaining culture was pelleted for maxi prep. The number of colonies was counted on the plates bearing the 1:10 and 1:100 dilutions thereby determining the efficiency which gives an idea about the theoretical diversity. For this, the formula “number of colonies x dilution x number of electroporations x 10” was applied. Initial libraries usually yielded efficiencies in the range of  $5 \times 10^5$  to  $5 \times 10^6$ .

#### 4.2.4.6. SEQUENCE ANALYSIS OF THE GENERATED LIBRARY BY “SALANTO”

---

First, single colonies were screened by colony PCR using primers #822 and #823 to test for successful integration. If more than half of the clones had no integration the library would be repeated. Positive clones were then inoculated for a “dirty” mini prep and the resulting plasmid was sent for sequencing by mixing 500 ng DNA with 5 µM primer in a total volume of 10 µl. The sequencing of classical Sanger sequencing as well as PacBio sequencing was performed at GATC biotech (Konstanz, Germany). Due to the length of the *cap* gene of approximately 2.2 kb three sequencing reactions were prepared to cover the complete sequence: the forward part was covered by primer #822, the end by primer #823 and the middle by either primer #36 or primer #37 depending of the serotype origin at that position. With the help of either AlignX or SerialCloner the complete sequence was put together manually and inserted into a text file already containing the respective sequences of the parental serotypes with the formatting “>name” and the sequence in the line below. The document was saved as “.fasta” file and was opened either in ClustalX or AlignX to perform an alignment. The resulting fasta file containing the alignment was analysed in the Shuffling-ALignment ANalysis Tool (Salanto) (Schurmann et al. 2013). Of note, with the help of the chromatograms all sequences were corrected for “N/n” which were indetermined from the sequencing software and potential “MUT”ations resulting from either alignment errors or from sequencing errors in homopolymer stretches. In addition, real mutations were identified by this as well. In

Salanto one can regard the “Type assignment” which assigns sequence stretches to a given parent, a mutation or to “indetermined” if the stretch originates from a parent where multiple possibilities exist. In order to generate the colored bars shown in this work the type assignment was copied into excel and a macro was run ((Große 2016), 6.2.1) to generate shortened and colored stretches which were then inserted into Powerpoint as a picture. Another feature allows regarding the homology between different variants by depicting the percentage of differing positions. This was used to generate the heatmaps showing homology. Furthermore, one can regard the library composition in terms of percentage of each parental sequence and the probability of a given serotype at a given position. The last analysis mentioned was depicted by line graphs.

---

## 4.2.5. CELL CULTURE AND VIRUS PRODUCTION

---

### 4.2.5.1. CELL CULTURE

---

All cell lines were cultured at 37°C with 5% CO<sub>2</sub> in their respective medium and were split upon reaching 80% confluency. For this purpose, the medium was removed and the cells were washed with PBS. To detach cells trypsin was added and the cells were incubated for 1-5 min in the incubator until they detached from the flask. The cells were then resuspended in 10 ml medium and either 1 ml of cell suspension was filled up with medium for further cultivation or cell number was determined automatically by the countess cell counter and the desired amount was seeded. Usually, DMEM Glutamax medium supplemented with 10% FBS and 5% Penicillin/ Streptomycin was used for most cell lines including HEK293T, LX-2, MCF-7 and SF539. Non-essential amino acids were additionally added for Huh7 cells. HUVECs were cultured in Endothelial Cell Growth Medium with supplements.

TABLE 4-12: LIST OF CELL LINES USED

Cell line	Origin	Tissue	Reference
HEK293T	Human	Embryonic kidney	(DuBridge et al. 1987)
Huh7	Human	Hepatic carcinoma	(Nakabayashi et al. 1982)
HUVECs	Human	Umbilical Vein Endothelial Cells	Provided by promocell (Heidelberg, Germany)
LX-2	Human	Hepatic stellate cells	(Xu et al. 2005)
MCF7	Human	Mammary carcinoma	(Soule et al. 1973)
RAW264.7	Mouse	Macrophage, tumour derived	(Raschke et al. 1978; Ralph and Nakoinz 1977)
SF539	Human	Gliosarcoma	(Rutka et al. 1986)

## 4.2.5.2. CRUDE LYSATE PRODUCTION (SMALL-SCALE)

In order to produce AAVs in a small-scale HEK293T cells were seeded in 2 ml medium at a density of  $0.5 \times 10^6$  cells per well of a 6-well plate. On the next day, two transfection mixes using a total of 2.6  $\mu\text{g}$  DNA and 300 mM NaCl were prepared as listed in Table 4-13. It was preferred to use DNA originating from kit isolation instead of the “dirty” mini protocol. For some AAP trans-complementation assays chimeric capsids were transferred from the replication competent context into the helper context (#1544) by HindIII and SpeI digestion beforehand. After combination of the two mixes they were vortexed and incubated at room temperature for 10 minutes. The mix was then added dropwise to the cells with a P200 pipette. Two days later, the medium was removed, the cells were rinsed once with PBS and were then transferred into a 1.5 ml tube by resuspension in 1 ml PBS. The samples were spun for 10 min at 4000 x g, the supernatant was removed and 500  $\mu\text{l}$  PBS was added. In order to free AAVs from the cells they were lysed by five cycles of vortexing, freezing in liquid nitrogen and thawing in a 37°C water bath. Finally, the cell debris was spun down at 16100 x g for 10 min in a table top centrifuge and the supernatant which contained the AAV particles was transferred into a new tube. This lysate is referred to as a “crude lysate” and still contains cellular components and empty particles. It was stored at -20°C.

TABLE 4-13: TRANSFECTION MIX FOR ONE WELL

	H <sub>2</sub> O	NaCl	PEI	DNA (library)	DNA (vectors)	DNA (AAP rescue)
Mix 1	49 $\mu\text{l}$	49 $\mu\text{l}$	-	1.3 $\mu\text{g}$ Adh, 1.3 $\mu\text{g}$ library	866.66 ng of each: AAV helper, Adh, transgene	690 ng of each: Adh, transgene, AAV helper and either 530 ng AAP or stuffer
Mix 2	27 $\mu\text{l}$	49 $\mu\text{l}$	22 $\mu\text{l}$	-	-	-

Adh Adenohelper

## 4.2.5.3. VIRUS PRODUCTION LARGE-SCALE

For large-scale purification, HEK293T cells were expanded at a density of  $7.5 \times 10^6$  cells per T175 flask. Approximately, four flasks were seeded for 30 plates. Two days later, the cells were harvested into one flask and the cell number was determined. Depending on the plate number a cell mix was prepared with  $4 \times 10^6$  cells in 22 ml medium per plate. Then, the cells were seeded into 15 cm dishes and incubated two more days. For transfection, two mixes (Table 4-14) using 300 mM NaCl were prepared, combined, and

vortexed. After 10 min incubation at room temperature 3.2 ml of the transfection mix was added dropwise to a plate. Usually 44 µg DNA were transfected per plate with equal amounts of each construct with the exception of the library production in the presence or absence of an AAP mixture. Here, a total of 32.4 µg DNA was transfected and the constructs (Adh, library, AAP mix/ stuffer) were used in a 1.3:1.3:1 ratio. Within the AAP mixture all AAPs from the respective parental serotypes were present in equal amounts. Three days later, the cells were scraped off the plates, resuspended in their medium and transferred to appropriate vessels for centrifugation at 500 x g for 15 min. The supernatant was removed, the cells were washed in PBS and the cell pellet was resuspended in 0.5 ml benzonase buffer per plate. Comparable to the crude lysates five freeze/ thaw cycles were performed. In order to remove left-over plasmid DNA from transfection 75 U/ ml benzonase were added and incubated at 37°C for 1 h with inversion of the tube every 10 min. The samples were spun at 4000 x g for 15 min to remove cell debris. This is the starting material for the density centrifugation steps.

TABLE 4-14: TRANSFECTION MIX FOR ONE PLATE

	H <sub>2</sub> O	NaCl	PEI	DNA (library)	DNA (library AAP)	DNA (vectors)
Mix 1	* µl	790 µl	-	22 µg Adh, 22 µg library	11.7 Adh, 11.7 µg library and 9 µg AAP mix or stuffer	14.6 µg of each: AAV helper, Adh, transgene
Mix 2	438 µl	790 µl	352 µl	-	-	-

\* water is used to fill up the DNA to 790 µl therefore the amount varied

#### 4.2.5.4. PURIFICATION BY IODIXANOL DENSITY CENTRIFUGATION

The samples underwent another spinning step at 4000 x g for 15 min to remove more debris. Then, the supernatant was transferred to a new tube and filled up to 7.5 ml with benzonase buffer. A Pasteur pipette was inserted into an ultracentrifuge tube (16x76 mm) and the sample was applied. Now, the iodixanol phases in increasing density were loaded: the clear 15% phase followed by the red 25% phase, the clear 40% phase and finally the yellow 60% phase. Carefully, the Pasteur pipette was removed, air bubbles were removed and the tube was filled up to half of the neck with benzonase buffer. Next, the tubes were sealed and balanced with the help of different caps. The samples were inserted into the type 70.1 Ti rotor and they were run at 4°C for 2 h at 50000 rpm in a Optima™ L-90K ultracentrifuge. After the run was completed, the vacuum was removed and the samples were extracted from the rotor. The tubes were fixed and first, a needle

was inserted at the top. Then, a needle was attached to a syringe and used to push through the 60% phase positioning the needle at the border of the 40% and 60% phase. The AAV containing 40% phase was pulled while taking care to avoid contamination from the 25% phase. The virus was aliquoted as required and stored at -80°C.

#### 4.2.5.5. PURIFICATION BY CESIUM CHLORIDE DENSITY CENTRIFUGATION

A first precipitation step of protein debris was performed with 1/39 volume of 1 M CaCl<sub>2</sub> on ice for 1 h. The sample was spun for 15 min at 4°C at 10000 x g and the supernatant was transferred into a new tube. Now, ¼ volume 40 % PEG solution was added and incubated overnight on ice to precipitate the AAV particles. The particles were pelleted at 2500 x g for 30 min at 4°C and the pellet was resuspended stepwise in a total of 10 ml NaHEPES buffer. The sample was spun again as before to clear the solution and the resulting supernatant was added to a prepared CsCl solution. Therefore, 13.2 g CsCl were mixed with 14 ml NaHEPES buffer. As this is an endothermic reaction the solution will turn cold and needs to be brought back to room temperature prior determination of the refractory index. With the help of a refractometer the refractory indices were determined and were adjusted to 1.3710 with NaHEPES buffer to lower the number or with CsCl to increase it again. The sample was transferred into Optima™ ultracentrifuge tubes (26x77 mm) and filled up with topping solution bearing the same refractory index. A cap was used to close the tube and with the help of larger caps the samples were balanced. The gradients were run under vacuum in a type 70 Ti rotor at 45000 rpm for 21-23 h at 20°C in Optima™ L-90K ultracentrifuge. After the run was completed, the vacuum was removed, the tubes extracted and fixed. First, a needle was used to push through the tube at the bottom and then a second needle was inserted at the top. With the help of the upper needle the flow speed of the gradient could be adjusted to a slow, dropwise manner and several fractions were collected: 3 ml, 3ml, 0.5 ml, 0.5 ml, 0.5 ml, 5 ml, 0.5 ml, 0.5 ml, 0.5 ml, 3 ml. The refractory indices of several fractions were determined and the ones bearing an index between 1.3766 and 1.3711 were pooled. The virus containing fractions were inserted into a dialysis cassette which was then incubated standing for 30 min at room temperature in 1x PBS. After this time PBS was exchanged and then dialysis occurred at 4°C with continuous stirring. More buffer exchanges occurred after 1 h, overnight, 2 h and again 2 h. Per sample one amicon Ultra-15 centrifugal filter unit was washed twice with 15 ml PBS for 2 min at 1000 x g. Then the sample was applied and was carefully concentrated to 500-1000 µl at 400 x g. The virus was aliquoted as required and stored at -80°C.

## 4.2.5.6. TITRATION BY QPCR

In order to determine the genome copy numbers per ml a titration was performed. Therefore, 10 µl virus was mixed with the same amount TE buffer (1 mM Tris-HCl) and for alkaline lysis 20 µl 2 M NaOH was added. After vortexing the sample incubated 30 min at 56°C and was cooled down shortly on ice. For neutralization 38 µl 1 M HCl was added and vortexed. Finally, the sample was filled up to 1 ml with 922 µl water. This alkaline lysis breaks up the virus particles thereby freeing the viral/ vector genomes. Usually, water served as a negative control and a sample with known titer as a positive control. If the virus was purified via an iodixanol gradient the sample was additionally diluted 1:10. The qPCR was measured in triplicates and the mastermix for one triplex contained 17.5 µl 2x Sensimix II probe kit, 1.4 µl of 1:10 diluted forward primer, 1.4 µl of 1:10 diluted reverse primer, 0.35 µl of 1:10 diluted probe and 9.35 µl water. In case of libraries the rep set was used for titration and for packaged eGFP and eYFP reporters the GFP set was applied as seen in Table 4-15. In a separate tube, 30 µl mastermix were mixed with 5 µl sample which was either originating from alkaline lysis or from a 1:10 dilution row of the plasmid serving as standard. The samples were mixed well and 10 µl were filled into one qPCR tube. The final concentration was 0.4 µM for each primer and 0.1 µM for the probe. The qPCR was run with an initial heating step of 10 min at 95°C followed by 40 cycles of heating for 10 s at 95°C and combined annealing/ elongation for 20 s at 60°C. The titer of the sample was determined based on the standard curve which usually covered a range of  $5 \times 10^9$  to  $5 \times 10^2$  plasmid copies per reaction and multiplying the value with x 100 (factor to get from 10 µl virus to 1 ml vector prep) x 100 (correct for dilution during alkaline lysis) x 7 (dilution of sample within tube) x dilution (in case of iodixanol based samples) x 2 (only in case of single-stranded vectors to correct for the double stranded standard).

TABLE 4-15: PRIMER/ PROBE SETS USED

Target	For 5' to 3'	Rev 5' to 3'	Probe 5' to 3'
eGFP	GAGCGCACCATCT TCTTCAAG	TGTCGCCCTCGAAC TTCAC	FAM-ACGACGGCAACTACA-BHQ1
Rep	AAGTCCTCGGCC AGATAGAC	CAATCACGGCGCAC ATGT	FAM-TGATCGTCACCTCCAACA-BHQ1

#### 4.2.5.7. TITRATION BY DDPCR

---

Comparable to qPCR the virus particles were broken by alkaline lysis and the same primer/ probe sets were used. In addition to this the samples were diluted 1:8000 in a dilution row by using 5  $\mu$ l for a 1:100 dilution followed by a 1:20 and a 1:4 dilution. Between dilution steps the samples were vortexed well. In parallel, a mastermix was prepared with 11  $\mu$ l ddPCRmix for probes no dUTP, 1.1  $\mu$ l forward primer, 1.1  $\mu$ l reverse primer, 1.1  $\mu$ l probe and 2.2  $\mu$ l water. Primers were diluted 1:5.55 from the 100  $\mu$ M stock and the probe was diluted 1:20 from the 100  $\mu$ M stock. In a separate tube 16.5  $\mu$ l mastermix were mixed with 5.5  $\mu$ l sample. The final concentration for each primer was 900 nM and for the probe 250 nM. After vortexing 20  $\mu$ l of the reaction was applied to a cartridge into the appropriate well which already contained 70  $\mu$ l oil for probes in its respective well. The cartridge was sealed with the casket and applied to the droplet generator. The droplets were transferred to a 96-well PCR plate and the plate was sealed at 105°C with tin foil. The PCR was run in a biorad cycler with an initial heating step of 10 min at 94°C followed by 40 cycles of heating 94°C 30 s and 60°C 1 min. After a final step for 10 min at 98°C the samples were kept at indefinite hold at 12°C. The droplets were evaluated with the droplet reader which gave absolute copy numbers per  $\mu$ l as an output. This value was then multiplied by x 100 (alkaline lysis) x 100 (correcting 10  $\mu$ l to 1 ml) x 5 (dilution in tube) x dilution (from dilution series).

---

#### 4.2.6. VIRUS VALIDATION: TRANSDUCTION ASSAYS AND CAPSID FORMATION

---

##### 4.2.6.1. TRANSDUCTION ASSAY

---

The transduction efficiency of different capsids was assessed by packaging a fluorescent marker such as eYFP or eGFP as a surrogate marker for infection. The cell line of interest was seeded into 96-well plates in a volume of 100  $\mu$ l per well. HEK293T cells were seeded at a density of  $1.5 \times 10^6$  cells per 96-well plate and  $0.5 \times 10^6$  cells per 96-well plate were seeded for MCF7 cells and SF539. The next day 10  $\mu$ l virus was added and a serial dilution was performed by mixing a row and transferring 10  $\mu$ l into the next one. In case of crude lysates, the mentioned volume was used and in case of purified virus a specific multiplicity of infection (MOI) as indicated was used. For the latter the required virus volume was mixed with medium to a final volume of 10  $\mu$ l. Two days later the cells were prepared for FACS analysis by removal of the medium, washing with PBS and addition of 30  $\mu$ l trypsin. The cells were incubated at 37°C for 5-10 min until the cells

detached from the plate and were resuspended in 170  $\mu$ l 1% (w/v) BSA in PBS. The amount of positive cells was assessed in a flow cytometer using channels FL3 (mCherry) against FL4 (eGFP) or FL5 (eYFP). Debris was excluded by gating and only the cell population was measured until either 10000 events were detected or 60 s have passed. The results were exported into an Excel sheet by the accompanying MXP software and the values were sorted by a makro (6.2.2.) For evaluation the dilution was chosen were values were below 95% to avoid over saturation. Usually, a ratio to an appropriate control which was set to 1 was built in order to compare between experiments.

#### 4.2.6.2. WESTERN BLOT

Protein samples were separated by reducing Tris-glycine sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Therefore, self-casted SDS gels were prepared by pouring first the resolving gel and putting a layer of isopropanol on top of it until it polymerized. In a next step, the isopropanol was removed, the stacking gel was poured on top and a comb was inserted (Table 4-16). For analysis of AAP expression HEK293T cells were transfected with 2.6  $\mu$ g of the AAP expression plasmid as described in 4.2.5.2. Likewise, cells were harvested two days later and the cell pellet was taken up in 100  $\mu$ l PBS. The samples were mixed with 2xSDS buffer and boiled for 5 min at 95°C. They were cooled down on ice, spun down shortly and 15  $\mu$ l of a 1:5 dilution was loaded per well. For protein size determination, 5  $\mu$ l of a protein ladder was loaded. The SDS gel was run at 90 V until the running front has entered the resolving gel and was then set to 120 V until the running front has just run out. The proteins were transferred onto a nitrocellulose membrane by semi-dry blotting. Therefore, a stack of three Whatman paper sheets cut to size were soaked in blotting buffer and placed onto the blotting device. Soaking and placing was then performed for the membrane, the gel and three more Whatman paper sheets. Air bubbles were removed by rolling over the pile and blotting occurred at 4°C for 1 h at 100 mA per membrane. Then, the device was disassembled, the membrane was blocked in 5% milk in TBST for 1 h at room temperature and the primary antibody was incubated shaking at 4°C overnight. The next day the membrane was washed three times with TBST for 5 min each and the secondary HRP coupled antibody was added for 1 h at room temperature. As before, three washing steps were performed and the blot was developed by addition of the ECL developing kit reagent. Luminescence was detected by film and developed. AAP expression was detected by its HA-tag and for comparison actin was stained as a housekeeper (Table 4 17).

TABLE 4-16: SDS GEL FOR WESTERN BLOT

Resolving gel (5ml)	12%	Stacking gel (2ml)	5%
ddH <sub>2</sub> O	2.15 ml	ddH <sub>2</sub> O	1.46 ml
Acrylamide (19:1)	1.5 ml	Acrylamide (19:1)	250 µl
1.5 M Tris (pH 8.8)/ 0.4% SDS	1.25 ml	0.5 M Tris (pH 6.8)/ 0.4% SDS	250 µl
10% SDS	50 µl	10% SDS	20 µl
1% APS	50 µl	1% APS	20 µl
TEMED	2 µl	TEMED	2 µl

## 4.2.6.3. DOT BLOT

In order to assess the capability of building intact particles a dot blot under native conditions was performed. A nitrocellulose membrane was cut to size, soaked 10 min in PBS and the Bio-Dot® SF Microfiltration Apparatus was assembled according to manufacturer's instruction. Then, 15 µl of untreated crude lysate or 15 µl of heated crude lysate were applied to the membrane. The native crude lysate contained assembled particles which were thought to disassemble upon heating 100 µl for 5 min at 95°C. A vacuum was applied and the samples were soaked through the membrane. After two washing steps with 100 µl PBS the device was disassembled and the membrane was blocked for 1 h at room temperature with 5% milk in TBST. The primary antibody was incubated for 2-3 h at room temperature. Comparable to western blot the membrane was washed and incubated with the secondary antibody. The chemo luminescence was either detected by development with film or by digital imaging with the intas system.

TABLE 4-17: LIST OF ANTIBODIES USED

Antibody	Species origin	Dilution	source	Secondary antibody
Adk1	mouse	1:10	(Kuck, Kern, and Kleinschmidt 2007)	Goat anti-mouse
A20	mouse	1:10	(Wistuba et al. 1995)	Goat anti-mouse
Adk4	mouse	1:10	(Kuck, Kern, and Kleinschmidt 2007)	Goat anti-mouse
Adk5	mouse	1:10	(Kuck, Kern, and Kleinschmidt 2007)	Goat anti-mouse
Adk6	mouse	1:10	(Sonntag et al. 2011)	Goat anti-mouse
Adk8 (3)	mouse	1:10	(Sonntag et al. 2011)	Goat anti-mouse
Adk9	mouse	1:10	(Sonntag et al. 2011)	Goat anti-mouse
Anti-VP	rabbit	1:200	61084, Progen Biotechnik GmbH (Heidelberg, Germany)	Donkey anti-rabbit
B1	mouse	1:10	(Wistuba et al. 1995)	Goat anti-mouse

anti-HA	mouse	1:200	sc-7392, Santa Cruz Biotechnology (Heidelberg, Germany)	Goat anti-mouse
anti- $\beta$ -actin	mouse	1:200	sc-47778, Santa Cruz Biotechnology (Heidelberg, Germany)	Goat anti-mouse
HRP coupled goat anti- mouse	goat	1:10000	115-035-003, Jackson ImmunoResearch Laboratories (West Grove, USA)	-
HRP coupled donkey anti- rabbit	donkey	1:10000	GE Healthcare (Chalfont St. Giles, UK)	-

All hybridoma supernatants detecting free VPs (B1) and assembled particles (Adk, A20) were a kind gift by Jürgen A. Kleinschmidt.

## 4.2.7. IN VIVO SELECTION AND VALIDATION

### 4.2.7.1. *IN VIVO* SELECTION ON STELLATE CELLS

A viral capsid library encompassing serotypes 1, 2, 3b, 4SO, 5SO, 6, 7, 8, 9 and rh10 was generated by DNA family shuffling. The plasmid library was transfected into 80 plates and the virus was produced by CsCl gradient centrifugation (4.2.5.5). In collaboration with the laboratory of Holger Willenbring (University of California, San Francisco, USA)  $5 \times 10^{11}$  vector genomes were injected into one Lrat-Cre; ZsGreen mouse (Mederacke et al. 2013; Madisen et al. 2010). In this mouse model, the Cre recombinase is expressed under the stellate cell-specific Irat promoter. As a consequence, stellate cells will be ZsGreen positive. After 24 h, the mouse was perfused (Mederacke et al. 2015) and stellate cells were isolated based on FACS sorting. The cells which were double positive for ZsGreen and vitamin A were collected. Stellate cells store vitamin A which in turn can be excited by a violet laser. The sorted cells were processed for DNA isolation which then could be used to perform a PCR based rescue of the capsid genes. For this a test PCR was run initially to identify the optimal template volume. Therefore, 0.25  $\mu$ l phusion HS II polymerase were mixed with 5  $\mu$ l GC buffer, 0.75  $\mu$ l #822 primer (10  $\mu$ M), 0.75  $\mu$ l #823 primer (10  $\mu$ M), 0.75  $\mu$ l DMSO and 0.5  $\mu$ l dNTPs (10 mM). The mix was filled up with water to 22.5  $\mu$ l and 2.5  $\mu$ l undiluted or diluted template were added. The condition which yielded the most prominent band was used for a large-scale PCR using five to ten times the mix described above. In case that the PCR product did not yield enough material a nested PCR under the same conditions was performed using primers #824 and #826 and a 1:10 dilution of the undigested PCR product as template. The samples were then purified, digested by PaeI and AseI and cloned back into a replication competent backbone as seen in 4.2.4.5. In contrast to the

generation of an initial library only ten instead of 20 electroporations were performed. Accordingly, a ligation reaction with a total volume of 20  $\mu$ l was set up with 569.5 ng insert and 430.5 ng insert, respectively. A secondary viral library was produced, injected and recovered. In total four rounds of selection were performed. After each round, single clones were sequenced and analysed with Salanto as described in 4.2.4.6. As the C-terminal enrichment differed from round three to round four the fourth round was repeated with two more mice. For validation, the sequences were grouped and two chimeras were randomly chosen from each group for further subcloning into the helper context #1544 via PacI and AscI digestion. This allowed for packaging of an eYFP (#552) reporter by transfection of 15 plates per construct and purification via iodixanol gradient centrifugation (4.2.5.4). For validation  $1 \times 10^{11}$  vector genomes (vg) were injected per mouse with three mice per variant. Ten days later, the liver tissue was fixed in 10% formalin, slices were generated and stained for indicated markers. The animal work and stainings were performed in the Willenbring laboratory and PCR rescues, sequence analysis and virus production in the Grimm laboratory.

The lead candidate AH3-5 was further validated in collaboration with the laboratory of Steven Dooley (University Hospital Mannheim, Germany). Four female BALB/cJRj mice 18 weeks old were injected with  $1 \times 10^{11}$  and ten days later the livers were perfused. The four indicated cell types were isolated from each liver, cultivated and stained for YFP.

#### 4.2.7.2. BULK VALIDATION BY NEXT-GENERATION SEQUENCING

---

In order to test more chimeric variants in parallel an approach based on next-generation sequencing (NGS) was chosen. This was done in close collaboration with a fellow PhD student in the Grimm Laboratory, Jonas Weinmann. In his work he already generated a library encompassing the twelve serotypes as well as various peptide insertion variants where each variant packaged an individual barcode. This library was injected into mice and organs of interest were extracted followed by DNA and RNA isolation. He established the protocols and scripts required for the NGS workflow and the subsequent analysis. In the screen presented in this study an extended library was generated encompassing published benchmarks, 28 chimeric variants selected on stellate cells and presented in this work (named AH) as well as chimeric variants selected on different muscle types (named JEA) from another colleague. Therefore, a barcoded YFP variant was packaged into 28 randomly chosen chimeric variants as described in section 4.2.5.3. The barcode consisted of a 15 nucleotide stretch in the 3' untranslated region of the YFP reporter. All variants were produced separately and at the end equal virus

amounts were pooled, concentrated and buffered to PBS. After titration of both, the initial library mentioned earlier, which already contained AH3-5, and the extended one, they were combined in such a manner that equal capsid amounts are represented. Six female C57BL/6J mice six weeks old were injected with  $1.57 \times 10^{12}$  vector genomes per mouse representing  $1 \times 10^{10}$  particles per variant with a total of 157 variants. These mice were used to regard the biodistribution. In addition, four female BALB/c mice 30 weeks old were injected with a comparable dose in collaboration with the Dooley laboratory. One week later the livers were perfused and different liver cell types were isolated (section 4.2.7.3). The cell pellets were fed into the pipeline for NGS sequencing by J. Weinmann and the details about processing and normalization can be found in his PhD thesis.

#### 4.2.7.3. SINGLE CELL TYPE ISOLATION FROM ONE MOUSE LIVER

The mice were anesthetized and perfused with a two-step collagenase perfusion technique through the vena cava (Mederacke et al. 2015). Hepatocytes were pelleted by a short centrifugation step at  $50 \times g$  for 2 min. The hepatocytes were loaded on a percoll gradient to separate live and dead cells. Live cells were recovered, counted and snap-frozen. The non-parenchymal cells (NPC) were contained in the supernatant of the first spin to pellet hepatocytes. The initial slow spinning step was repeated to remove further hepatocytes. Then the NPC fraction was pelleted at  $300 \times g$  for 10 min and resuspended in 13 ml DMEM w/o phenol red. The cells were transferred into a 15 ml tube and pelleted as before. The pellet was then filled up to 1.2 ml with DMEM w/o phenol red and mixed with 4.8 ml 30% histodenz solution (15 g histodenz in 40 ml DPBS). Carefully, 2 ml PBS were layered on top and the gradient was centrifuged at  $1500 \times g$  for 23 min without break and acceleration 4. The NPCs were collected at the histodenz/ PBS interface and added to 5 ml MACS buffer (0.5% BSA, 2 mM EDTA in PBS). The cells were pelleted as before and resuspended in 365  $\mu$ l MACS buffer. After addition of 15  $\mu$ l beads and 25  $\mu$ l Fcblock the cells incubated for 15 min at  $4^\circ\text{C}$  with slight shaking every 5 min. The cells were washed by addition of 4 ml MACS buffer and pelleted as above. The resulting pellet was incubated with 20  $\mu$ l secondary beads and 380  $\mu$ l MACS buffer. After equilibration of the MS columns (Miltenyi Biotech, Germany) the sample was washed, resuspended in 1 ml MACS buffer and was applied onto the column according to manufacturer's instructions. The flow through was stained for the next marker and the eluate contained the targeted cell population. The Fcblock was added for the first staining only. Subsequently, stellate cells, Kupffer cells (CD11b) and LSECs (CD146) were isolated, counted and snap-frozen. The MACS based sorting for stellate cells occurred with an unpublished marker identified by the Dooley laboratory (manuscript in preparation).

---

## 4.2.8. PROOF OF PRINCIPLE FOR BARCODED LIBRARIES

---

### 4.2.8.1. *IN VITRO* VALIDATION

---

Libraries were generated by DNA family shuffling and were transferred into the standard pSSV9 backbone. The generated vectors from 4.2.3.3 were used to transfer three libraries by *PacI* and *Ascl* digestion. The 289 library was tagged with BC289, the 1-10 library was inserted into BCori and 1789rh10 was transferred into BCnew (see 4.2.3.3). The 289 library was previously generated (Herrmann 2014), the 1-10 library was created for the selection (see 4.2.7.1) and the 1789rh10 was generated during this study. All three libraries were transferred in parallel setting up the ligation in 10  $\mu$ l using 215.3 ng backbone and 284.8 ng insert with two electroporations. The acceptor backbone was 778-BC1 (4.2.3.3). The resulting plasmid library was recovered by mini prep using a commercial kit. As described in 4.2.5.2 crude lysates were produced for each individual library. For the *in vitro* validation 50  $\mu$ l crude lysate were used to transduce HEK293T cells in one well of a 6-well plate. All libraries were individually transduced and additionally in combination with equal library amounts within 50  $\mu$ l. After 24 h the cells were washed with PBS, resuspended in 1 ml PBS and transferred to a 1.5 ml reaction tube. The samples were spun at 1700 x g for 10 min, the supernatant was removed and the cell pellet was resuspended in 120  $\mu$ l DirectPCR lysis reagent with proteinaseK being diluted 1:100 from a 20 mg/ml stock solution. The cells were lysed overnight at 56°C shaking in 0.5 ml tubes. The next day, the samples were transferred into PCR tubes and proteinaseK was inactivated at 85°C for 45 min. This mixture was then used to perform a PCR rescue as described in 4.2.7.1 with #833 as forward primer and different combinations of reverse primer to either amplify the whole cap population (rescueBC1) or to recover *caps* from a single library (rescueBC289, rescueBCnew (1789rh10), rescueBCori). The PCR program was comparable except that annealing occurred at 61°C. For subcloning a digestion with *HindIII* and *NdeI* was performed thereby subcloning *cap* with the respective library barcode back into an appropriate replication competent context. Note that a transfer by *PacI* and *NdeI* was unsuccessful due to a possible overlap of the sticky ends. For the ligation a 3:1 ratio was kept with 45 ng insert and 62.1 ng vector in a total volume of 10  $\mu$ l. The next day 2  $\mu$ l were electroporated and after the recovery period the bacteria were spun at 4000 x g for 5 min. The supernatant was poured off, the pellet was resuspended in the remaining liquid and plated out. Positive clones confirmed by colony PCR using primers #822 and #823 were sent for sequencing with #823. The present barcode was identified by sequencing with the help of two scripts (6.3) running on Python 3.5 and clones were analysed by Salanto as well.

#### 4.2.8.2. *IN VIVO* VALIDATION

In order to see if it is possible to track the library origin *in vivo* a total of four barcoded libraries were produced namely 289, 1-10 and two newly generated 1-10 libraries. The two newly generated 1-10 libraries were different as one was generated with the same conditions as the previous library ("new") and the other one had reduced amounts of two serotypes ("red"). As the first 1-10 library had a bias towards AAV1 and AAV6 they were taken as one portion for the DNase I digest therefore every serotype encompassed 1/9 of the total reaction instead of 1/10 as before. The libraries were transferred by PacI and AscI digestion into a backbone already containing BC1 as the general barcode and the respective library barcodes BC289, BCori, BCred and BCnew. For the ligation 645.9 ng vector and 854.4 ng insert in a total volume of 30  $\mu$ l were mixed. In total, 15 electroporations were performed per library. All four barcoded libraries were produced separately via CsCl gradient centrifugation and titrated (4.2.5.5, 4.2.5.7). Equal amounts of each library were mixed and  $2.09 \times 10^{11}$  vg were injected into two female NMRI mice six weeks of age. The mice were acquired from Janvier and all experiments were conducted by operators bearing a FELASA category B certificate according to European regulations at the "Klinisch Experimenteller Bereich (KEB)" at University Heidelberg. The experiments were approved by German authorities (35-9185.81/G-89/16, "Selektion neuer AAV-Kapside in adulten Mäusen"). After 48 h the mice were sacrificed and liver, kidney, muscle, lung and heart were extracted and snap-frozen in liquid nitrogen. The gDNA was isolated with the help of the DNeasy Blood & Tissue kit according to manufacturer's instructions. The tissue was lysed in buffer ATL including the optional addition of 1:100 diluted reagentDX with the tissue lyser. Elution occurred in 150  $\mu$ l elution buffer and different volumes were tested for *cap* recovery with the primer pair #833 and rescueBC1. As before, PCRs were run to rescue *cap*, digestion occurred with HindIII and NdeI and after ligation two electroporations for each organ were performed. For each organ 96 clones were sequenced and the barcode was determined with the help of the scripts (6.3).



## 5. REFERENCES

- Adams, M. J., E. J. Lefkowitz, A. M. King, and E. B. Carstens. 2014. 'Ratification vote on taxonomic proposals to the International Committee on Taxonomy of Viruses (2014)', *Arch Virol*, 159: 2831-41.
- Akache, B., D. Grimm, X. Shen, S. Fuess, S. R. Yant, D. S. Glazer, J. Park, and M. A. Kay. 2007. 'A two-hybrid screen identifies cathepsins B and L as uncoating factors for adeno-associated virus 2 and 8', *Mol Ther*, 15: 330-9.
- Albright, B. H., C. M. Storey, G. Murlidharan, R. M. Castellanos Rivera, G. E. Berry, V. J. Madigan, and A. Asokan. 2018. 'Mapping the Structural Determinants Required for AAVrh.10 Transport across the Blood-Brain Barrier', *Mol Ther*, 26: 510-23.
- Ameri, H. 2018. 'Prospect of retinal gene therapy following commercialization of voretigene neparvovec-rzyl for retinal dystrophy mediated by RPE65 mutation', *J Curr Ophthalmol*, 30: 1-2.
- Asuri, P., M. A. Bartel, T. Vazin, J. H. Jang, T. B. Wong, and D. V. Schaffer. 2012. 'Directed evolution of adeno-associated virus for enhanced gene delivery and gene targeting in human pluripotent stem cells', *Mol Ther*, 20: 329-38.
- Atchison, R. W., B. C. Casto, and W. M. Hammon. 1965. 'Adenovirus-Associated Defective Virus Particles', *Science*, 149: 754-6.
- Bantel-Schaal, U., and H. zur Hausen. 1984. 'Characterization of the DNA of a defective human parvovirus isolated from a genital site', *Virology*, 134: 52-63.
- Bartlett, J. S., J. Kleinschmidt, R. C. Boucher, and R. J. Samulski. 1999. 'Targeted adeno-associated virus vector transduction of nonpermissive cells mediated by a bispecific F(ab'gamma)2 antibody', *Nat Biotechnol*, 17: 181-6.
- Bartlett, J. S., R. Wilcher, and R. J. Samulski. 2000. 'Infectious entry pathway of adeno-associated virus and adeno-associated virus vectors', *J Virol*, 74: 2777-85.
- Bartneck, M., K. T. Warzecha, C. G. Tag, S. Sauer-Lehnen, F. Heymann, C. Trautwein, R. Weiskirchen, and F. Tacke. 2015. 'Isolation and time lapse microscopy of highly pure hepatic stellate cells', *Anal Cell Pathol (Amst)*, 2015: 417023.
- Becerra, S. P., F. Koczot, P. Fabisch, and J. A. Rose. 1988. 'Synthesis of adeno-associated virus structural proteins requires both alternative mRNA splicing and alternative initiations from a single transcript', *J Virol*, 62: 2745-54.
- Becerra, S. P., J. A. Rose, M. Hardy, B. M. Baroudy, and C. W. Anderson. 1985. 'Direct mapping of adeno-associated virus capsid proteins B and C: a possible ACG initiation codon', *Proc Natl Acad Sci U S A*, 82: 7919-23.
- Berns, K. I., and N. Muzyczka. 2017. 'AAV: An Overview of Unanswered Questions', *Hum Gene Ther*, 28: 308-13.
- Berry, G. E., and A. Asokan. 2016. 'Cellular transduction mechanisms of adeno-associated viral vectors', *Curr Opin Virol*, 21: 54-60.
- Boutin, S., V. Monteilhet, P. Veron, C. Leborgne, O. Benveniste, M. F. Montus, and C. Masurier. 2010. 'Prevalence of serum IgG and neutralizing factors against adeno-associated virus (AAV) types 1, 2, 5, 6, 8, and 9 in the healthy population: implications for gene therapy using AAV vectors', *Hum Gene Ther*, 21: 704-12.
- Bryant, L. M., D. M. Christopher, A. R. Giles, C. Hinderer, J. L. Rodriguez, J. B. Smith, E. A. Traxler, J. Tycko, A. P. Wojno, and J. M. Wilson. 2013. 'Lessons learned from the clinical development and market authorization of Glybera', *Hum Gene Ther Clin Dev*, 24: 55-64.
- Buchholz, C. J., T. Friedel, and H. Buning. 2015. 'Surface-Engineered Viral Vectors for Selective and Cell Type-Specific Gene Delivery', *Trends Biotechnol*, 33: 777-90.
- Buning, H., and M. Schmidt. 2015. 'Adeno-associated Vector Toxicity-To Be or Not to Be?', *Mol Ther*, 23: 1673-75.

- Calcedo, R., L. H. Vandenberghe, G. Gao, J. Lin, and J. M. Wilson. 2009. 'Worldwide epidemiology of neutralizing antibodies to adeno-associated viruses', *J Infect Dis*, 199: 381-90.
- Chandler, M., F. de la Cruz, F. Dyda, A. B. Hickman, G. Moncalian, and B. Ton-Hoang. 2013. 'Breaking and joining single-stranded DNA: the HUH endonuclease superfamily', *Nat Rev Microbiol*, 11: 525-38.
- Chen, C. L., R. L. Jensen, B. C. Schnepf, M. J. Connell, R. Shell, T. J. Sferra, J. S. Bartlett, K. R. Clark, and P. R. Johnson. 2005. 'Molecular characterization of adeno-associated viruses infecting children', *J Virol*, 79: 14781-92.
- Choudhury, S. R., Z. Fitzpatrick, A. F. Harris, S. A. Maitland, J. S. Ferreira, Y. Zhang, S. Ma, R. B. Sharma, H. L. Gray-Edwards, J. A. Johnson, A. K. Johnson, L. C. Alonso, C. Punzo, K. R. Wagner, C. A. Maguire, R. M. Kotin, D. R. Martin, and M. Sena-Esteves. 2016. 'In Vivo Selection Yields AAV-B1 Capsid for Central Nervous System and Muscle Gene Therapy', *Mol Ther*, 24: 1247-57.
- Crudele, J. M., J. D. Finn, J. I. Siner, N. B. Martin, G. P. Niemeyer, S. Zhou, F. Mingozzi, C. D. Lothrop, Jr., and V. R. Arruda. 2015. 'AAV liver expression of FIX-Padua prevents and eradicates FIX inhibitor without increasing thrombogenicity in hemophilia B dogs and mice', *Blood*, 125: 1553-61.
- D'Avola, D., E. Lopez-Franco, B. Sangro, A. Paneda, N. Grossios, I. Gil-Farina, A. Benito, J. Twisk, M. Paz, J. Ruiz, M. Schmidt, H. Petry, P. Harper, R. E. de Salamanca, A. Fontanellas, J. Prieto, and G. Gonzalez-Aseguinolaza. 2016. 'Phase I open label liver-directed gene therapy clinical trial for acute intermittent porphyria', *J Hepatol*, 65: 776-83.
- Dalkara, D., L. C. Byrne, R. R. Klimczak, M. Visel, L. Yin, W. H. Merigan, J. G. Flannery, and D. V. Schaffer. 2013. 'In vivo-directed evolution of a new adeno-associated virus for therapeutic outer retinal gene delivery from the vitreous', *Sci Transl Med*, 5: 189ra76.
- Daya, S., and K. I. Berns. 2008. 'Gene therapy using adeno-associated virus vectors', *Clin Microbiol Rev*, 21: 583-93.
- Deverman, B. E., P. L. Pravdo, B. P. Simpson, S. R. Kumar, K. Y. Chan, A. Banerjee, W. L. Wu, B. Yang, N. Huber, S. P. Pasca, and V. Gradinaru. 2016. 'Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain', *Nat Biotechnol*, 34: 204-9.
- Doherty, D. G. 2016. 'Immunity, tolerance and autoimmunity in the liver: A comprehensive review', *J Autoimmun*, 66: 60-75.
- Duan, D., P. Sharma, J. Yang, Y. Yue, L. Dudus, Y. Zhang, K. J. Fisher, and J. F. Engelhardt. 1998. 'Circular intermediates of recombinant adeno-associated virus have defined structural characteristics responsible for long-term episomal persistence in muscle tissue', *J Virol*, 72: 8568-77.
- DuBridge, R. B., P. Tang, H. C. Hsia, P. M. Leong, J. H. Miller, and M. P. Calos. 1987. 'Analysis of mutation in human cells by using an Epstein-Barr virus shuttle system', *Mol Cell Biol*, 7: 379-87.
- Dudek, A. M., S. Pillay, A. S. Puschnik, C. M. Nagamine, F. Cheng, J. Qiu, J. E. Carette, and L. H. Vandenberghe. 2018. 'An Alternate Route for Adeno-associated Virus (AAV) Entry Independent of AAV Receptor', *J Virol*, 92.
- Earley, L. F., Y. Kawano, K. Adachi, X. X. Sun, M. S. Dai, and H. Nakai. 2015. 'Identification and characterization of nuclear and nucleolar localization signals in the adeno-associated virus serotype 2 assembly-activating protein', *J Virol*, 89: 3038-48.
- Earley, L. F., J. M. Powers, K. Adachi, J. T. Baumgart, N. L. Meyer, Q. Xie, M. S. Chapman, and H. Nakai. 2017. 'Adeno-associated Virus (AAV) Assembly-Activating Protein Is Not an Essential Requirement for Capsid Assembly of AAV Serotypes 4, 5, and 11', *J Virol*, 91.
- Excoffon, K. J., J. T. Koerber, D. D. Dickey, M. Murtha, S. Keshavjee, B. K. Kaspar, J. Zabner, and D. V. Schaffer. 2009. 'Directed evolution of adeno-associated virus to an infectious respiratory virus', *Proc Natl Acad Sci U S A*, 106: 3865-70.

- Farkas, S. L., Z. Zadori, M. Benko, S. Essbauer, B. Harrach, and P. Tijssen. 2004. 'A parvovirus isolated from royal python (*Python regius*) is a member of the genus Dependovirus', *J Gen Virol*, 85: 555-61.
- Fernandez-Iglesias, A., and J. Gracia-Sancho. 2017. 'How to Face Chronic Liver Disease: The Sinusoidal Perspective', *Front Med (Lausanne)*, 4: 7.
- Ferrari, F. K., T. Samulski, T. Shenk, and R. J. Samulski. 1996. 'Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors', *J Virol*, 70: 3227-34.
- Fisher, K. J., G. P. Gao, M. D. Weitzman, R. DeMatteo, J. F. Burda, and J. M. Wilson. 1996. 'Transduction with recombinant adeno-associated virus for gene therapy is limited by leading-strand synthesis', *J Virol*, 70: 520-32.
- Flotte, T. R., and K. I. Berns. 2005. 'Adeno-associated virus: a ubiquitous commensal of mammals', *Hum Gene Ther*, 16: 401-7.
- Gao, G., M. R. Alvira, S. Somanathan, Y. Lu, L. H. Vandenberghe, J. J. Rux, R. Calcedo, J. Sanmiguel, Z. Abbas, and J. M. Wilson. 2003. 'Adeno-associated viruses undergo substantial evolution in primates during natural infections', *Proc Natl Acad Sci U S A*, 100: 6081-6.
- Gao, G. P., M. R. Alvira, L. Wang, R. Calcedo, J. Johnston, and J. M. Wilson. 2002. 'Novel adeno-associated viruses from rhesus monkeys as vectors for human gene therapy', *Proc Natl Acad Sci U S A*, 99: 11854-9.
- Gao, G., L. H. Vandenberghe, M. R. Alvira, Y. Lu, R. Calcedo, X. Zhou, and J. M. Wilson. 2004. 'Clades of Adeno-associated viruses are widely disseminated in human tissues', *J Virol*, 78: 6381-8.
- Geisler, A., A. Jungmann, J. Kurreck, W. Poller, H. A. Katus, R. Vetter, H. Fechner, and O. J. Muller. 2011. 'microRNA122-regulated transgene expression increases specificity of cardiac gene transfer upon intravenous delivery of AAV9 vectors', *Gene Ther*, 18: 199-209.
- Geoffroy, M. C., and A. Salvetti. 2005. 'Helper functions required for wild type and recombinant adeno-associated virus growth', *Curr Gene Ther*, 5: 265-71.
- Girod, A., M. Ried, C. Wobus, H. Lahm, K. Leike, J. Kleinschmidt, G. Deleage, and M. Hallek. 1999. 'Genetic capsid modifications allow efficient re-targeting of adeno-associated virus type 2', *Nat Med*, 5: 1438.
- Girod, A., C. E. Wobus, Z. Zadori, M. Ried, K. Leike, P. Tijssen, J. A. Kleinschmidt, and M. Hallek. 2002. 'The VP1 capsid protein of adeno-associated virus type 2 is carrying a phospholipase A2 domain required for virus infectivity', *J Gen Virol*, 83: 973-8.
- Godoy, P., N. J. Hewitt, U. Albrecht, M. E. Andersen, N. Ansari, S. Bhattacharya, J. G. Bode, J. Bolleyn, C. Borner, J. Bottger, A. Braeuning, R. A. Budinsky, B. Burkhardt, N. R. Cameron, G. Camussi, C. S. Cho, Y. J. Choi, J. Craig Rowlands, U. Dahmen, G. Damm, O. Dirsch, M. T. Donato, J. Dong, S. Dooley, D. Drasdo, R. Eakins, K. S. Ferreira, V. Fonsato, J. Fraczek, R. Gebhardt, A. Gibson, M. Glanemann, C. E. Goldring, M. J. Gomez-Lechon, G. M. Groothuis, L. Gustavsson, C. Guyot, D. Hallifax, S. Hammad, A. Hayward, D. Haussinger, C. Hellerbrand, P. Hewitt, S. Hoehme, H. G. Holzutter, J. B. Houston, J. Hrach, K. Ito, H. Jaeschke, V. Keitel, J. M. Kelm, B. Kevin Park, C. Kordes, G. A. Kullak-Ublick, E. L. LeCluyse, P. Lu, J. Luebke-Wheeler, A. Lutz, D. J. Maltman, M. Matz-Soja, P. McMullen, I. Merfort, S. Messner, C. Meyer, J. Mwinyi, D. J. Naisbitt, A. K. Nussler, P. Olinga, F. Pampaloni, J. Pi, L. Pluta, S. A. Przyborski, A. Ramachandran, V. Rogiers, C. Rowe, C. Schelcher, K. Schmich, M. Schwarz, B. Singh, E. H. Stelzer, B. Stieger, R. Stober, Y. Sugiyama, C. Tetta, W. E. Thasler, T. Vanhaecke, M. Vinken, T. S. Weiss, A. Widera, C. G. Woods, J. J. Xu, K. M. Yarborough, and J. G. Hengstler. 2013. 'Recent advances in 2D and 3D in vitro systems using primary hepatocytes, alternative hepatocyte sources and non-parenchymal liver cells and their use in investigating mechanisms of hepatotoxicity, cell signaling and ADME', *Arch Toxicol*, 87: 1315-530.

- Goncalves, M. A. 2005. 'Adeno-associated virus: from defective virus to effective vector', *Virology*, 2: 43.
- Gory, S., M. Vernet, M. Laurent, E. Dejana, J. Dalmon, and P. Huber. 1999. 'The vascular endothelial-cadherin promoter directs endothelial-specific expression in transgenic mice', *Blood*, 93: 184-92.
- Gray, S. J., B. L. Blake, H. E. Criswell, S. C. Nicolson, R. J. Samulski, T. J. McCown, and W. Li. 2010. 'Directed evolution of a novel adeno-associated virus (AAV) vector that crosses the seizure-compromised blood-brain barrier (BBB)', *Mol Ther*, 18: 570-8.
- Grimm, D., A. Kern, K. Rittner, and J. A. Kleinschmidt. 1998. 'Novel tools for production and purification of recombinant adenoassociated virus vectors', *Hum Gene Ther*, 9: 2745-60.
- Grimm, D., J. S. Lee, L. Wang, T. Desai, B. Akache, T. A. Storm, and M. A. Kay. 2008. 'In vitro and in vivo gene therapy vector evolution via multispecies interbreeding and retargeting of adeno-associated viruses', *J Virol*, 82: 5887-911.
- Grimm, D., and S. Zolotukhin. 2015. 'E Pluribus Unum: 50 Years of Research, Millions of Viruses, and One Goal--Tailored Acceleration of AAV Evolution', *Mol Ther*, 23: 1819-31.
- Grosse, S., M. Penaud-Budloo, A. K. Herrmann, K. Borner, J. Fakhiri, V. Laketa, C. Kramer, E. Wiedtke, M. Gunkel, L. Menard, E. Ayuso, and D. Grimm. 2017. 'Relevance of Assembly-Activating Protein for Adeno-associated Virus Vector Production and Capsid Protein Stability in Mammalian and Insect Cells', *J Virol*, 91.
- Große, Stefanie. 2016. 'Small but increasingly mighty: New insights into Adeno-associated virus (AAV) capsid biology and implications for AAV vector optimization', for the degree of Doctor of Natural Science, Ruperto-Carola University of Heidelberg.
- Hentzschel, F., C. Hammerschmidt-Kamper, K. Borner, K. Heiss, B. Knapp, J. M. Sattler, L. Kaderali, M. Castoldi, J. G. Bindman, Y. Malato, H. Willenbring, A. K. Mueller, and D. Grimm. 2014. 'AAV8-mediated in vivo overexpression of miR-155 enhances the protective capacity of genetically attenuated malarial parasites', *Mol Ther*, 22: 2130-41.
- Herrmann, A. K., and D. Grimm. 2018. 'High-Throughput Dissection of AAV-Host Interactions: The Fast and the Curious', *J Mol Biol*.
- Herrmann, A. K., S. Grosse, C. Bender, E. Wiedtke, J. El Andari, S. Kreuz, E. Kienle, N. Schürmann, and D. Grimm. manuscript in preparation. 'A robust and comprehensive pipeline for directed molecular evolution of shuffled Adeno-associated viral (AAV) gene transfer vectors'.
- Herrmann, A. K., S. Grosse, K. Borner, C. Kramer, E. Wiedtke, M. Gunkel, and D. Grimm. 2018. 'Impact of the assembly-activating protein (AAP) on molecular evolution of synthetic Adeno-associated virus (AAV) capsids', *Hum Gene Ther*.
- Herrmann, A. K., S. Große, K. Börner, C. Krämer, E. Wiedtke, M. Gunkel, and D. Grimm. 2018. 'Impact of the assembly-activating protein (AAP) on molecular evolution of synthetic Adeno-associated virus (AAV) capsids', *Human Gene Therapy*, *accepted*.
- Herrmann, Anne-Kathrin. 2014. 'Impact of different selection approaches on Adeno-associated virus vector evolution', Master Thesis, Ruperto-Carola University of Heidelberg.
- Hoggan, M. D., N. R. Blacklow, and W. P. Rowe. 1966. 'Studies of small DNA viruses found in various adenovirus preparations: physical, biological, and immunological characteristics', *Proc Natl Acad Sci U S A*, 55: 1467-74.
- Hosel, M., A. Huber, S. Bohlen, J. Lucifora, G. Ronzitti, F. Puzzo, F. Boisgerault, U. T. Hacker, W. J. Kwanten, N. Kloting, M. Blüher, A. Gluschko, M. Schramm, O. Utermohlen, W. Bloch, F. Mingozzi, O. Krut, and H. Buning. 2017. 'Autophagy determines efficiency of liver-directed gene therapy with adeno-associated viral vectors', *Hepatology*, 66: 252-65.
- Huang, L. Y., S. Halder, and M. Agbandje-McKenna. 2014. 'Parvovirus glycan interactions', *Curr Opin Virol*, 7: 108-18.
- Johnson, F. B., H. L. Ozer, and M. D. Hoggan. 1971. 'Structural proteins of adenovirus-associated virus type 3', *J Virol*, 8: 860-63.

- Johnson, J. S., C. Li, N. DiPrimio, M. S. Weinberg, T. J. McCown, and R. J. Samulski. 2010. 'Mutagenesis of adeno-associated virus type 2 capsid protein VP1 uncovers new roles for basic amino acids in trafficking and cell-specific transduction', *J Virol*, 84: 8888-902.
- Johnson, J. S., and R. J. Samulski. 2009. 'Enhancement of adeno-associated virus infection by mobilizing capsids into and out of the nucleolus', *J Virol*, 83: 2632-44.
- Kaepfel, C., S. G. Beattie, R. Fronza, R. van Logtenstein, F. Salmon, S. Schmidt, S. Wolf, A. Nowrouzi, H. Glimm, C. von Kalle, H. Petry, D. Gaudet, and M. Schmidt. 2013. 'A largely random AAV integration profile after LPLD gene therapy', *Nat Med*, 19: 889-91.
- Kay, M. A. 2011. 'State-of-the-art gene-based therapies: the road ahead', *Nat Rev Genet*, 12: 316-28.
- Kelich, J. M., J. Ma, B. Dong, Q. Wang, M. Chin, C. M. Magura, W. Xiao, and W. Yang. 2015. 'Super-resolution imaging of nuclear import of adeno-associated virus in live cells', *Mol Ther Methods Clin Dev*, 2: 15047.
- Kern, A., K. Schmidt, C. Leder, O. J. Muller, C. E. Wobus, K. Bettinger, C. W. Von der Lieth, J. A. King, and J. A. Kleinschmidt. 2003. 'Identification of a heparin-binding motif on adeno-associated virus type 2 capsids', *J Virol*, 77: 11072-81.
- Khabou, H., M. Desrosiers, C. Winckler, S. Fouquet, G. Auregan, A. P. Bemelmans, J. A. Sahel, and D. Dalkara. 2016. 'Insight into the mechanisms of enhanced retinal transduction by the engineered AAV2 capsid variant -7m8', *Biotechnol Bioeng*, 113: 2712-24.
- Kmiec, Z. 2001. 'Cooperation of liver cells in health and disease', *Adv Anat Embryol Cell Biol*, 161: III-XIII, 1-151.
- Koerber, J. T., J. H. Jang, and D. V. Schaffer. 2008. 'DNA shuffling of adeno-associated virus yields functionally diverse viral progeny', *Mol Ther*, 16: 1703-9.
- Kotin, R. M., M. Siniscalco, R. J. Samulski, X. D. Zhu, L. Hunter, C. A. Laughlin, S. McLaughlin, N. Muzyczka, M. Rocchi, and K. I. Berns. 1990. 'Site-specific integration by adeno-associated virus', *Proc Natl Acad Sci U S A*, 87: 2211-5.
- Kuck, D., A. Kern, and J. A. Kleinschmidt. 2007. 'Development of AAV serotype-specific ELISAs using novel monoclonal antibodies', *J Virol Methods*, 140: 17-24.
- Larkin, M. A., G. Blackshields, N. P. Brown, R. Chenna, P. A. McGettigan, H. McWilliam, F. Valentin, I. M. Wallace, A. Wilm, R. Lopez, J. D. Thompson, T. J. Gibson, and D. G. Higgins. 2007. 'Clustal W and Clustal X version 2.0', *Bioinformatics*, 23: 2947-8.
- Laughlin, C. A., J. D. Tratschin, H. Coon, and B. J. Carter. 1983. 'Cloning of infectious adeno-associated virus genomes in bacterial plasmids', *Gene*, 23: 65-73.
- LeCluyse, E. L., R. P. Witek, M. E. Andersen, and M. J. Powers. 2012. 'Organotypic liver culture models: meeting current challenges in toxicity testing', *Crit Rev Toxicol*, 42: 501-48.
- Li, W., A. Asokan, Z. Wu, T. Van Dyke, N. DiPrimio, J. S. Johnson, L. Govindaswamy, M. Agbandje-McKenna, S. Leichtle, D. Eugene Redmond, Jr., T. J. McCown, K. B. Petermann, N. E. Sharpless, and R. J. Samulski. 2008. 'Engineering and Selection of Shuffled AAV Genomes: A New Strategy for Producing Targeted Biological Nanoparticles', *Mol Ther*, 16: 1252-60.
- Li, W., L. Zhang, J. S. Johnson, W. Zhijian, J. C. Grieger, X. Ping-Jie, L. M. Drouin, M. Agbandje-McKenna, R. J. Pickles, and R. J. Samulski. 2009. 'Generation of novel AAV variants by directed evolution for improved CFTR delivery to human ciliated airway epithelium', *Mol Ther*, 17: 2067-77.
- Lindsley, A., P. Snider, H. Zhou, R. Rogers, J. Wang, M. Olaopa, A. Kruzynska-Frejtag, S. V. Koushik, B. Lilly, J. B. Burch, A. B. Firulli, and S. J. Conway. 2007. 'Identification and characterization of a novel Schwann and outflow tract endocardial cushion lineage-restricted periostin enhancer', *Dev Biol*, 307: 340-55.
- Lisowski, L., A. P. Dane, K. Chu, Y. Zhang, S. C. Cunningham, E. M. Wilson, S. Nygaard, M. Grompe, I. E. Alexander, and M. A. Kay. 2014. 'Selection and evaluation of clinically relevant AAV variants in a xenograft liver model', *Nature*, 506: 382-6.

- Liu, J., and Y. A. Moon. 2016. 'Simple Purification of Adeno-Associated Virus-DJ for Liver-Specific Gene Expression', *Yonsei Med J*, 57: 790-4.
- Lusby, E., K. H. Fife, and K. I. Berns. 1980. 'Nucleotide sequence of the inverted terminal repetition in adeno-associated virus DNA', *J Virol*, 34: 402-9.
- Madigan, V. J., and A. Asokan. 2016. 'Engineering AAV receptor footprints for gene therapy', *Curr Opin Virol*, 18: 89-96.
- Madisen, L., T. A. Zwingman, S. M. Sunkin, S. W. Oh, H. A. Zariwala, H. Gu, L. L. Ng, R. D. Palmiter, M. J. Hawrylycz, A. R. Jones, E. S. Lein, and H. Zeng. 2010. 'A robust and high-throughput Cre reporting and characterization system for the whole mouse brain', *Nat Neurosci*, 13: 133-40.
- Maguire, C. A., D. Gianni, D. H. Meijer, L. A. Shaket, H. Wakimoto, S. D. Rabkin, G. Gao, and M. Sena-Esteves. 2010. 'Directed evolution of adeno-associated virus for glioma cell transduction', *J Neurooncol*, 96: 337-47.
- Maheshri, N., J. T. Koerber, B. K. Kaspar, and D. V. Schaffer. 2006. 'Directed evolution of adeno-associated virus yields enhanced gene delivery vectors', *Nat Biotechnol*, 24: 198-204.
- Matsushita, T., S. Elliger, C. Elliger, G. Podsakoff, L. Villarreal, G. J. Kurtzman, Y. Iwaki, and P. Colosi. 1998. 'Adeno-associated virus vectors can be efficiently produced without helper virus', *Gene Ther*, 5: 938-45.
- McCarty, D. M. 2008. 'Self-complementary AAV vectors; advances and applications', *Mol Ther*, 16: 1648-56.
- McCarty, D. M., H. Fu, P. E. Monahan, C. E. Toulson, P. Naik, and R. J. Samulski. 2003. 'Adeno-associated virus terminal repeat (TR) mutant generates self-complementary vectors to overcome the rate-limiting step to transduction in vivo', *Gene Ther*, 10: 2112-8.
- McCarty, D. M., P. E. Monahan, and R. J. Samulski. 2001. 'Self-complementary recombinant adeno-associated virus (scAAV) vectors promote efficient transduction independently of DNA synthesis', *Gene Ther*, 8: 1248-54.
- McLaughlin, S. K., P. Collis, P. L. Hermonat, and N. Muzyczka. 1988. 'Adeno-associated virus general transduction vectors: analysis of proviral structures', *J Virol*, 62: 1963-73.
- Mederacke, I., D. H. Dapito, S. Affo, H. Uchinami, and R. F. Schwabe. 2015. 'High-yield and high-purity isolation of hepatic stellate cells from normal and fibrotic mouse livers', *Nat Protoc*, 10: 305-15.
- Mederacke, I., C. C. Hsu, J. S. Troeger, P. Huebener, X. Mu, D. H. Dapito, J. P. Pradere, and R. F. Schwabe. 2013. 'Fate tracing reveals hepatic stellate cells as dominant contributors to liver fibrosis independent of its aetiology', *Nat Commun*, 4: 2823.
- Merlin, S., E. S. Cannizzo, E. Borroni, V. Brusca, P. Schinco, W. Tulalamba, M. K. Chuah, V. R. Arruda, T. VandenDriessche, M. Prat, G. Valente, and A. Follenzi. 2017. 'A Novel Platform for Immune Tolerance Induction in Hemophilia A Mice', *Mol Ther*, 25: 1815-30.
- Miller, D. G., G. D. Trobridge, L. M. Petek, M. A. Jacobs, R. Kaul, and D. W. Russell. 2005. 'Large-scale analysis of adeno-associated virus vector integration sites in normal human cells', *J Virol*, 79: 11434-42.
- Mizukami, H., N. S. Young, and K. E. Brown. 1996. 'Adeno-associated virus type 2 binds to a 150-kilodalton cell membrane glycoprotein', *Virology*, 217: 124-30.
- Mori, S., L. Wang, T. Takeuchi, and T. Kanda. 2004. 'Two novel adeno-associated viruses from cynomolgus monkey: pseudotyping characterization of capsid protein', *Virology*, 330: 375-83.
- Muller, O. J., F. Kaul, M. D. Weitzman, R. Pasqualini, W. Arap, J. A. Kleinschmidt, and M. Trepel. 2003. 'Random peptide libraries displayed on adeno-associated virus to select for targeted gene therapy vectors', *Nat Biotechnol*, 21: 1040-6.
- Munch, R. C., H. Janicki, I. Volker, A. Rasbach, M. Hallek, H. Buning, and C. J. Buchholz. 2013. 'Displaying high-affinity ligands on adeno-associated viral vectors enables tumor cell-specific and safe gene transfer', *Mol Ther*, 21: 109-18.

- Munch, R. C., A. Muth, A. Muik, T. Friedel, J. Schmatz, B. Dreier, A. Trkola, A. Pluckthun, H. Buning, and C. J. Buchholz. 2015. 'Off-target-free gene delivery by affinity-purified receptor-targeted viral vectors', *Nat Commun*, 6: 6246.
- Nakabayashi, H., K. Taketa, K. Miyano, T. Yamane, and J. Sato. 1982. 'Growth of human hepatoma cells lines with differentiated functions in chemically defined medium', *Cancer Res*, 42: 3858-63.
- Naumer, M., F. Sonntag, K. Schmidt, K. Nieto, C. Panke, N. E. Davey, R. Popa-Wagner, and J. A. Kleinschmidt. 2012. 'Properties of the adeno-associated virus assembly-activating protein', *J Virol*, 86: 13038-48.
- Nicolson, S. C., and R. J. Samulski. 2014. 'Recombinant adeno-associated virus utilizes host cell nuclear import machinery to enter the nucleus', *J Virol*, 88: 4132-44.
- Nieto, K., and A. Salvetti. 2014. 'AAV Vectors Vaccines Against Infectious Diseases', *Front Immunol*, 5: 5.
- Nonnenmacher, M. E., J. C. Cintrat, D. Gillet, and T. Weber. 2015. 'Syntaxin 5-dependent retrograde transport to the trans-Golgi network is required for adeno-associated virus transduction', *J Virol*, 89: 1673-87.
- Nonnenmacher, M., H. van Bakel, R. J. Hajjar, and T. Weber. 2015. 'High capsid-genome correlation facilitates creation of AAV libraries for directed evolution', *Mol Ther*, 23: 675-82.
- Nonnenmacher, M., and T. Weber. 2012. 'Intracellular transport of recombinant adeno-associated virus vectors', *Gene Ther*, 19: 649-58.
- Ojala, D. S., S. Sun, J. L. Santiago-Ortiz, M. G. Shapiro, P. A. Romero, and D. V. Schaffer. 2018. 'In Vivo Selection of a Computationally Designed SCHEMA AAV Library Yields a Novel Variant for Infection of Adult Neural Stem Cells in the SVZ', *Mol Ther*, 26: 304-19.
- Opie, S. R., K. H. Warrington, Jr., M. Agbandje-McKenna, S. Zolotukhin, and N. Muzyczka. 2003. 'Identification of amino acid residues in the capsid proteins of adeno-associated virus type 2 that contribute to heparan sulfate proteoglycan binding', *J Virol*, 77: 6995-7006.
- Parks, W. P., J. L. Melnick, R. Rongey, and H. D. Mayor. 1967. 'Physical assay and growth cycle studies of a defective adeno-satellite virus', *J Virol*, 1: 171-80.
- Perabo, L., H. Buning, D. M. Kofler, M. U. Ried, A. Girod, C. M. Wendtner, J. Enssle, and M. Hallek. 2003. 'In vitro selection of viral vectors with modified tropism: the adeno-associated virus display', *Mol Ther*, 8: 151-7.
- Perabo, L., J. Endell, S. King, K. Lux, D. Goldnau, M. Hallek, and H. Buning. 2006. 'Combinatorial engineering of a gene therapy vector: directed evolution of adeno-associated virus', *J Gene Med*, 8: 155-62.
- Pillay, S., and J. E. Carette. 2017. 'Host determinants of adeno-associated viral vector entry', *Curr Opin Virol*, 24: 124-31.
- Pillay, S., N. L. Meyer, A. S. Puschnik, O. Davulcu, J. Diep, Y. Ishikawa, L. T. Jae, J. E. Wosen, C. M. Nagamine, M. S. Chapman, and J. E. Carette. 2016. 'An essential receptor for adeno-associated virus infection', *Nature*, 530: 108-12.
- Pillay, S., W. Zou, F. Cheng, A. S. Puschnik, N. L. Meyer, S. S. Ganaie, X. Deng, J. E. Wosen, O. Davulcu, Z. Yan, J. F. Engelhardt, K. E. Brown, M. S. Chapman, J. Qiu, and J. E. Carette. 2017. 'AAV serotypes have distinctive interactions with domains of the cellular receptor AAVR', *J Virol*.
- Ponnazhagan, S., G. Mahendra, S. Kumar, J. A. Thompson, and M. Castillas, Jr. 2002. 'Conjugate-based targeting of recombinant adeno-associated virus type 2 vectors by using avidin-linked ligands', *J Virol*, 76: 12900-7.
- Powell, S. K., N. Khan, C. L. Parker, R. J. Samulski, G. Matsushima, S. J. Gray, and T. J. McCown. 2016. 'Characterization of a novel adeno-associated viral vector with preferential oligodendrocyte tropism', *Gene Ther*, 23: 807-14.

- Protzer, U., M. K. Maini, and P. A. Knolle. 2012. 'Living in the liver: hepatic infections', *Nat Rev Immunol*, 12: 201-13.
- Prudencio, M., A. Rodriguez, and M. M. Mota. 2006. 'The silent path to thousands of merozoites: the Plasmodium liver stage', *Nat Rev Microbiol*, 4: 849-56.
- Rabinowitz, J. E., W. Xiao, and R. J. Samulski. 1999. 'Insertional mutagenesis of AAV2 capsid and the production of recombinant virus', *Virology*, 265: 274-85.
- Ralph, P., and I. Nakoinz. 1977. 'Antibody-dependent killing of erythrocyte and tumor targets by macrophage-related cell lines: enhancement by PPD and LPS', *J Immunol*, 119: 950-54.
- Rangarajan, S., L. Walsh, W. Lester, D. Perry, B. Madan, M. Laffan, H. Yu, C. Vettermann, G. F. Pierce, W. Y. Wong, and K. J. Pasi. 2017. 'AAV5-Factor VIII Gene Transfer in Severe Hemophilia A', *N Engl J Med*, 377: 2519-30.
- Raschke, W. C., S. Baird, P. Ralph, and I. Nakoinz. 1978. 'Functional macrophage cell lines transformed by Abelson leukemia virus', *Cell*, 15: 261-7.
- Rezvani, M., R. Espanol-Suner, Y. Malato, L. Dumont, A. A. Grimm, E. Kienle, J. G. Bindman, E. Wiedtke, B. Y. Hsu, S. J. Naqvi, R. F. Schwabe, C. U. Corvera, D. Grimm, and H. Willenbring. 2016. 'In Vivo Hepatic Reprogramming of Myofibroblasts with AAV Vectors as a Therapeutic Strategy for Liver Fibrosis', *Cell Stem Cell*, 18: 809-16.
- Rhoads, A., and K. F. Au. 2015. 'PacBio Sequencing and Its Applications', *Genomics Proteomics Bioinformatics*, 13: 278-89.
- Ruffing, M., H. Zentgraf, and J. A. Kleinschmidt. 1992. 'Assembly of viruslike particles by recombinant structural proteins of adeno-associated virus type 2 in insect cells', *J Virol*, 66: 6922-30.
- Rutka, J. T., J. R. Giblin, H. K. Hoifodt, D. V. Dougherty, C. W. Bell, J. R. McCulloch, R. L. Davis, C. B. Wilson, and M. L. Rosenblum. 1986. 'Establishment and characterization of a cell line from a human gliosarcoma', *Cancer Res*, 46: 5893-902.
- Rutledge, E. A., C. L. Halbert, and D. W. Russell. 1998. 'Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2', *J Virol*, 72: 309-19.
- Samulski, R. J., K. I. Berns, M. Tan, and N. Muzyczka. 1982. 'Cloning of adeno-associated virus into pBR322: rescue of intact virus from the recombinant plasmid in human cells', *Proc Natl Acad Sci U S A*, 79: 2077-81.
- Samulski, R. J., L. S. Chang, and T. Shenk. 1989. 'Helper-free stocks of recombinant adeno-associated viruses: normal integration does not require viral gene expression', *J Virol*, 63: 3822-8.
- Samulski, R. J., X. Zhu, X. Xiao, J. D. Brook, D. E. Housman, N. Epstein, and L. A. Hunter. 1991. 'Targeted integration of adeno-associated virus (AAV) into human chromosome 19', *EMBO J*, 10: 3941-50.
- Schindelin, J., I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J. Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona. 2012. 'Fiji: an open-source platform for biological-image analysis', *Nat Methods*, 9: 676-82.
- Schmidt, M., L. Govindasamy, S. Afione, N. Kaludov, M. Agbandje-McKenna, and J. A. Chiorini. 2008. 'Molecular characterization of the heparin-dependent transduction domain on the capsid of a novel adeno-associated virus isolate, AAV(VR-942)', *J Virol*, 82: 8911-6.
- Schmidt, M., A. Voutetakis, S. Afione, C. Zheng, D. Mandikian, and J. A. Chiorini. 2008. 'Adeno-associated virus type 12 (AAV12): a novel AAV serotype with sialic acid- and heparan sulfate proteoglycan-independent transduction activity', *J Virol*, 82: 1399-406.
- Schurmann, N., L. G. Trabuco, C. Bender, R. B. Russell, and D. Grimm. 2013. 'Molecular dissection of human Argonaute proteins by DNA shuffling', *Nat Struct Mol Biol*, 20: 818-26.
- Simard, A. R., D. Soulet, G. Gowing, J. P. Julien, and S. Rivest. 2006. 'Bone marrow-derived microglia play a critical role in restricting senile plaque formation in Alzheimer's disease', *Neuron*, 49: 489-502.

- Smith, J., R. Herrero, K. Erles, D. Grimm, N. Munoz, F. X. Bosch, L. Tafur, K. V. Shah, and J. R. Schlehofer. 2001. 'Adeno-associated virus seropositivity and HPV-induced cervical cancer in Spain and Colombia', *Int J Cancer*, 94: 520-6.
- Sonntag, F., S. Bleker, B. Leuchs, R. Fischer, and J. A. Kleinschmidt. 2006. 'Adeno-associated virus type 2 capsids with externalized VP1/VP2 trafficking domains are generated prior to passage through the cytoplasm and are maintained until uncoating occurs in the nucleus', *J Virol*, 80: 11040-54.
- Sonntag, F., K. Kother, K. Schmidt, M. Weghofer, C. Raupp, K. Nieto, A. Kuck, B. Gerlach, B. Bottcher, O. J. Muller, K. Lux, M. Horer, and J. A. Kleinschmidt. 2011. 'The assembly-activating protein promotes capsid assembly of different adeno-associated virus serotypes', *J Virol*, 85: 12686-97.
- Sonntag, F., K. Schmidt, and J. A. Kleinschmidt. 2010. 'A viral assembly factor promotes AAV2 capsid formation in the nucleolus', *Proc Natl Acad Sci U S A*, 107: 10220-5.
- Soule, H. D., J. Vazquez, A. Long, S. Albert, and M. Brennan. 1973. 'A human cell line from a pleural effusion derived from a breast carcinoma', *J Natl Cancer Inst*, 51: 1409-16.
- Srivastava, A. 2016. 'In vivo tissue-tropism of adeno-associated viral vectors', *Curr Opin Virol*, 21: 75-80.
- Srivastava, A., E. W. Lusby, and K. I. Berns. 1983. 'Nucleotide sequence and organization of the adeno-associated virus 2 genome', *J Virol*, 45: 555-64.
- Stahnke, S., K. Lux, S. Uhrig, F. Kreppel, M. Hosel, O. Coutelle, M. Ogris, M. Hallek, and H. Buning. 2011. 'Intrinsic phospholipase A2 activity of adeno-associated virus is involved in endosomal escape of incoming particles', *Virology*, 409: 77-83.
- Stemmer, W. P. 1994a. 'DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution', *Proc Natl Acad Sci U S A*, 91: 10747-51.
- . 1994b. 'Rapid evolution of a protein in vitro by DNA shuffling', *Nature*, 370: 389-91.
- Surosky, R. T., M. Urabe, S. G. Godwin, S. A. McQuiston, G. J. Kurtzman, K. Ozawa, and G. Natsoulis. 1997. 'Adeno-associated virus Rep proteins target DNA sequences to a unique locus in the human genome', *J Virol*, 71: 7951-9.
- Thomas, C. E., T. A. Storm, Z. Huang, and M. A. Kay. 2004. 'Rapid uncoating of vector genomes is the key to efficient liver transduction with pseudotyped adeno-associated virus vectors', *J Virol*, 78: 3110-22.
- Trempe, J. P., and B. J. Carter. 1988. 'Alternate mRNA splicing is required for synthesis of adeno-associated virus VP1 capsid protein', *J Virol*, 62: 3356-63.
- Tse, L. V., S. Moller-Tank, R. M. Meganck, and A. Asokan. 2018. 'Mapping and Engineering Functional Domains of the Assembly Activating Protein of Adeno-Associated Viruses', *J Virol*.
- Wang, Z., H. I. Ma, J. Li, L. Sun, J. Zhang, and X. Xiao. 2003. 'Rapid and highly efficient transduction by double-stranded adeno-associated virus vectors in vitro and in vivo', *Gene Ther*, 10: 2105-11.
- Weitzman, M. D., and R. M. Linden. 2011. 'Adeno-associated virus biology', *Methods Mol Biol*, 807: 1-23.
- Wistuba, A., A. Kern, S. Weger, D. Grimm, and J. A. Kleinschmidt. 1997. 'Subcellular compartmentalization of adeno-associated virus type 2 assembly', *J Virol*, 71: 1341-52.
- Wistuba, A., S. Weger, A. Kern, and J. A. Kleinschmidt. 1995. 'Intermediates of adeno-associated virus type 2 assembly: identification of soluble complexes containing Rep and Cap proteins', *J Virol*, 69: 5311-9.
- Xiao, P. J., C. Li, A. Neumann, and R. J. Samulski. 2012. 'Quantitative 3D tracing of gene-delivery viral vectors in human cells and animal tissues', *Mol Ther*, 20: 317-28.
- Xiao, X., J. Li, and R. J. Samulski. 1998. 'Production of high-titer recombinant adeno-associated virus vectors in the absence of helper adenovirus', *J Virol*, 72: 2224-32.

- Xie, J., Q. Xie, H. Zhang, S. L. Ameres, J. H. Hung, Q. Su, R. He, X. Mu, S. Seher Ahmed, S. Park, H. Kato, C. Li, C. Mueller, C. C. Mello, Z. Weng, T. R. Flotte, P. D. Zamore, and G. Gao. 2011. 'MicroRNA-regulated, systemically delivered rAAV9: a step closer to CNS-restricted transgene expression', *Mol Ther*, 19: 526-35.
- Xu, L., A. Y. Hui, E. Albanis, M. J. Arthur, S. M. O'Byrne, W. S. Blaner, P. Mukherjee, S. L. Friedman, and F. J. Eng. 2005. 'Human hepatic stellate cell lines, LX-1 and LX-2: new tools for analysis of hepatic fibrosis', *Gut*, 54: 142-51.
- Yang, L., J. Jiang, L. M. Drouin, M. Agbandje-McKenna, C. Chen, C. Qiao, D. Pu, X. Hu, D. Z. Wang, J. Li, and X. Xiao. 2009. 'A myocardium tropic adeno-associated virus (AAV) evolved by DNA shuffling and in vivo selection', *Proc Natl Acad Sci U S A*, 106: 3946-51.
- Zhang, H., B. Yang, X. Mu, S. S. Ahmed, Q. Su, R. He, H. Wang, C. Mueller, M. Sena-Esteves, R. Brown, Z. Xu, and G. Gao. 2011. 'Several rAAV vectors efficiently cross the blood-brain barrier and transduce neurons and astrocytes in the neonatal mouse central nervous system', *Mol Ther*, 19: 1440-8.
- Zincarelli, C., S. Soltys, G. Rengo, and J. E. Rabinowitz. 2008. 'Analysis of AAV serotypes 1-9 mediated gene expression and tropism in mice after systemic injection', *Mol Ther*, 16: 1073-80.
- Zinn, E., S. Pacouret, V. Khaychuk, H. T. Turunen, L. S. Carvalho, E. Andres-Mateos, S. Shah, R. Shelke, A. C. Maurer, E. Plovie, R. Xiao, and L. H. Vandenberghe. 2015. 'In Silico Reconstruction of the Viral Evolutionary Lineage Yields a Potent Gene Therapy Vector', *Cell Rep*, 12: 1056-68.

## 6. SUPPLEMENTAL INFORMATION

### 6.1. SUPPLEMENTAL FIGURES

#### 6.1.1. OPTIMIZATION OF DNA FAMILY SHUFFLING

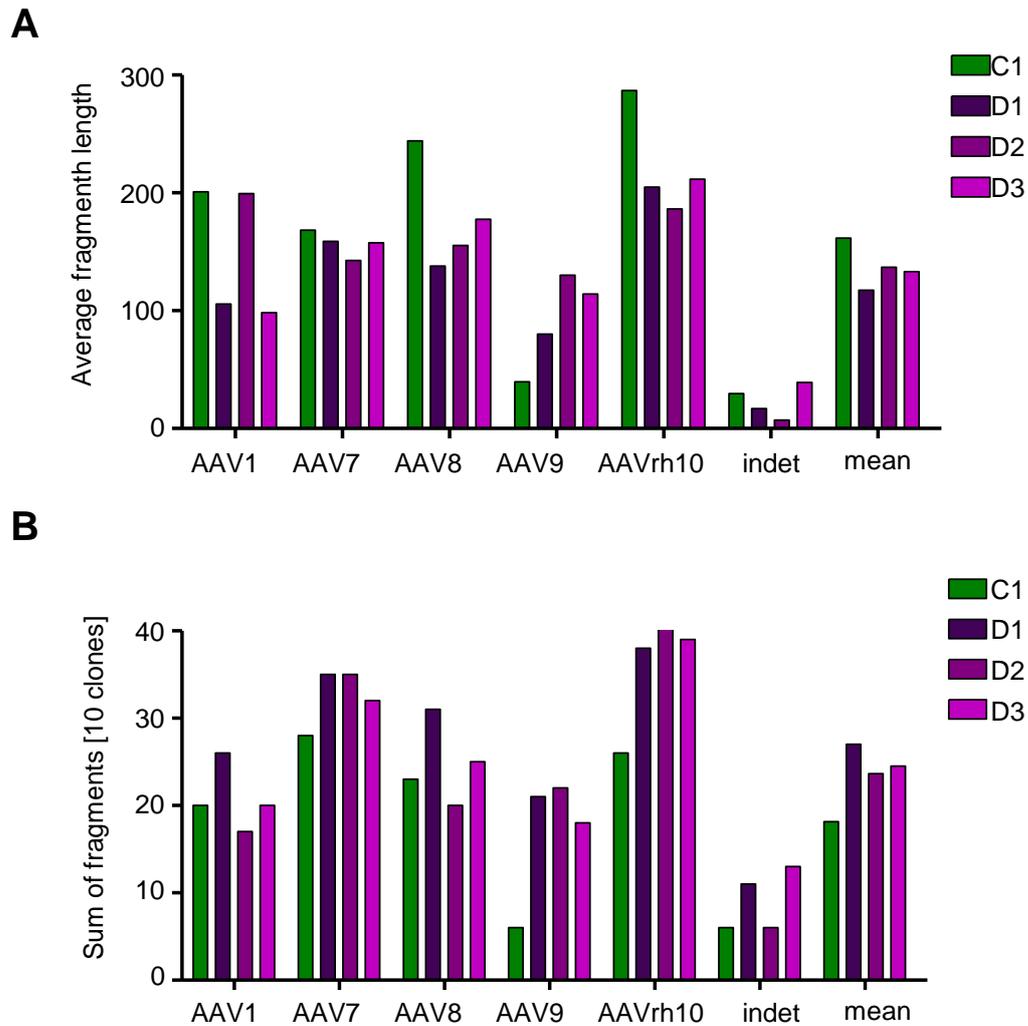


FIGURE 6-1: SALANTO ANALYSIS LIBRARIES C1 AND D1-3

The complete analysis for 10 chimeric clones is depicted. (A) The average fragment length for each parental serotype of each library was determined and presented as a bar graph. (B) Accordingly, the sum of fragments is shown. The data set was adapted from (Herrmann et al. manuscript in preparation).

## 6.1.2. INADVERTENT SHUFFLING OF AAP

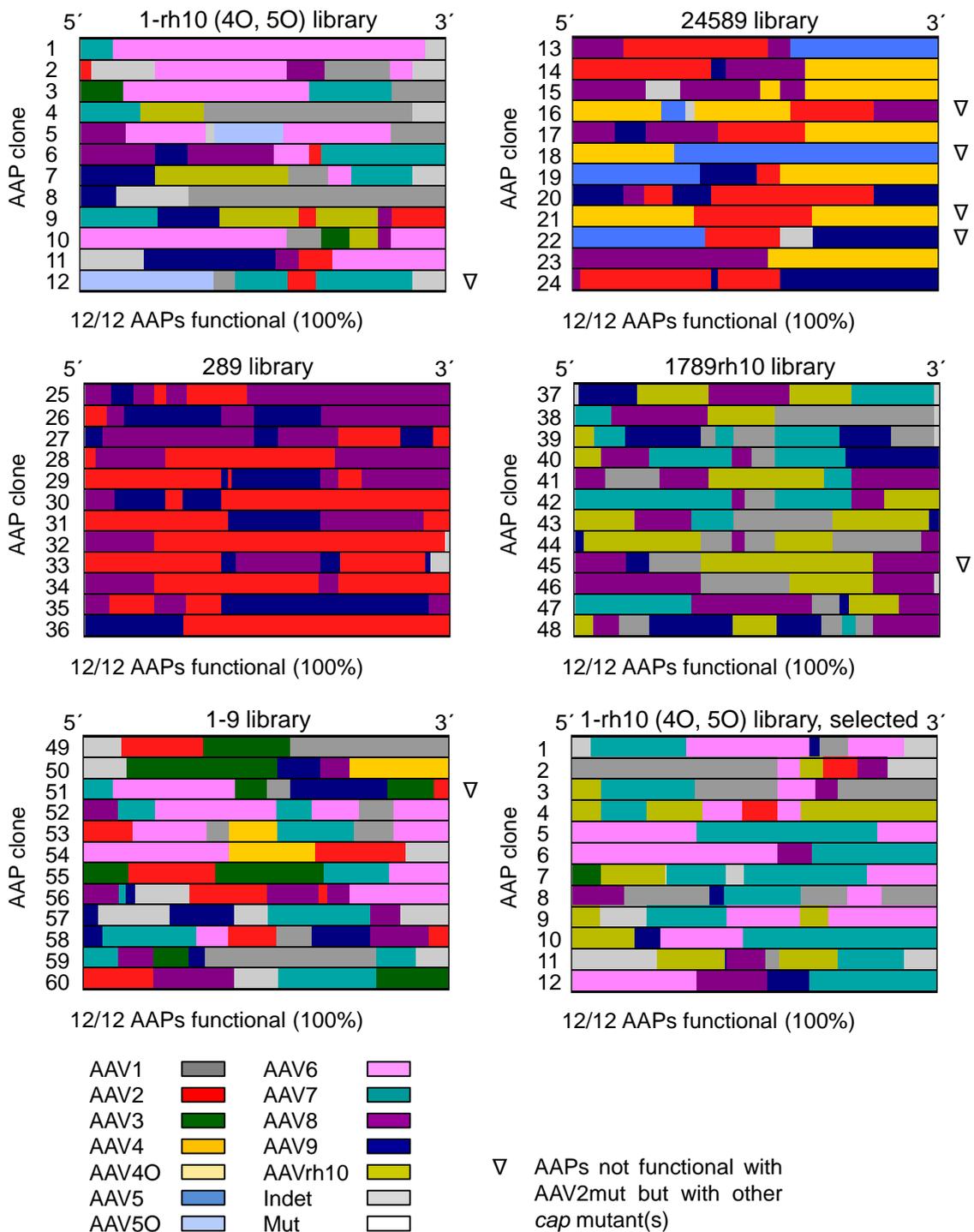


FIGURE 6-2: SEQUENCE ANALYSIS OF THE CHIMERIC AAPs

The distribution of all parental sequences of the chimeric AAPs was analyzed with Salanto. Each bar represents the nucleotide sequence of AAP for one clone and the parental stretches are colored accordingly. If several possibilities exist and no clear parent can be identified the sequence is assigned as indetermined (indet). Mutations (mut) are highlighted in white. The triangles indicate chimeric AAPs which were not compatible with the AAV2 mutant. This data set was adapted (Herrmann, Grosse, et al. 2018).

TABLE 6-1: RESCUE OF DIFFERENT PARENTAL AAP KNOCKOUTS

AAP <sup>a</sup>	AAVmut								
	1	3	4	5	6	7	8	9	rh10
12	82.2±6.9	2.1±0.6	23.0±1.1	83.5±0.9	36.8±2.2	12.5±0.5	37.4±3.2	43.6±1.6	1.75±0.2
16	-	-	26.0±4.2	2.0±0.0	-	-	0.0±0.0	0.0±0.0	-
18	-	-	22.0±8.8	2.5±0.7	-	-	0.0±0.0	0.0±0.0	-
19 <sup>b</sup>	-	-	65.0±12.7	84.7±24.9	-	-	72.7±8.1	87.7±7.1	-
21	-	-	85.0±16.9	15±2.8	-	-	0.0±0.0	0.0±0.0	-
22	-	-	-	61.0±5.7	-	-	5.0±0.7	2.3±0.2	-
39 <sup>b</sup>	26.4±20.3	-	-	-	-	10.0±5.0	40.8±25.6	20.7±23.8	9.5±3.5
45	89.5±2.5	-	-	-	-	11.6±0.8	75.2±3.1	82.4±1.6	11.0±0.8
51 <sup>b</sup>	1.0±0.0	11.0±1.4	13.5±6.4	4.0±0.0	8.5±3.5	4.6±4.7	16.3±9.5	17.3±18.8	-

<sup>a</sup>Crude cell lysates were tested in MCF7 (AAV4) or SF539 cells (AAV1, 3, 5 to 9, rh10) (n ≥ 2). Positive cells were determined by flow cytometry and values (shown plus SD) were normalized to the respective AAV wild-type construct (always set to 100%). Grey, non-functional AAPs (< 5% activity); orange, intermediate performers (5 to 50%); green, high performers (≥ 50%).

<sup>a</sup>Numbers in this column refer to the labels of the chimeric AAP clones, not to AAV serotypes.

<sup>b</sup>These three clones were included since they barely crossed the 5% assay cutoff in Figure 2C."

This table was taken from (Herrmann, Grosse, et al. 2018).

## 6.1.3. AAV SELECTION ON STELLATE CELLS

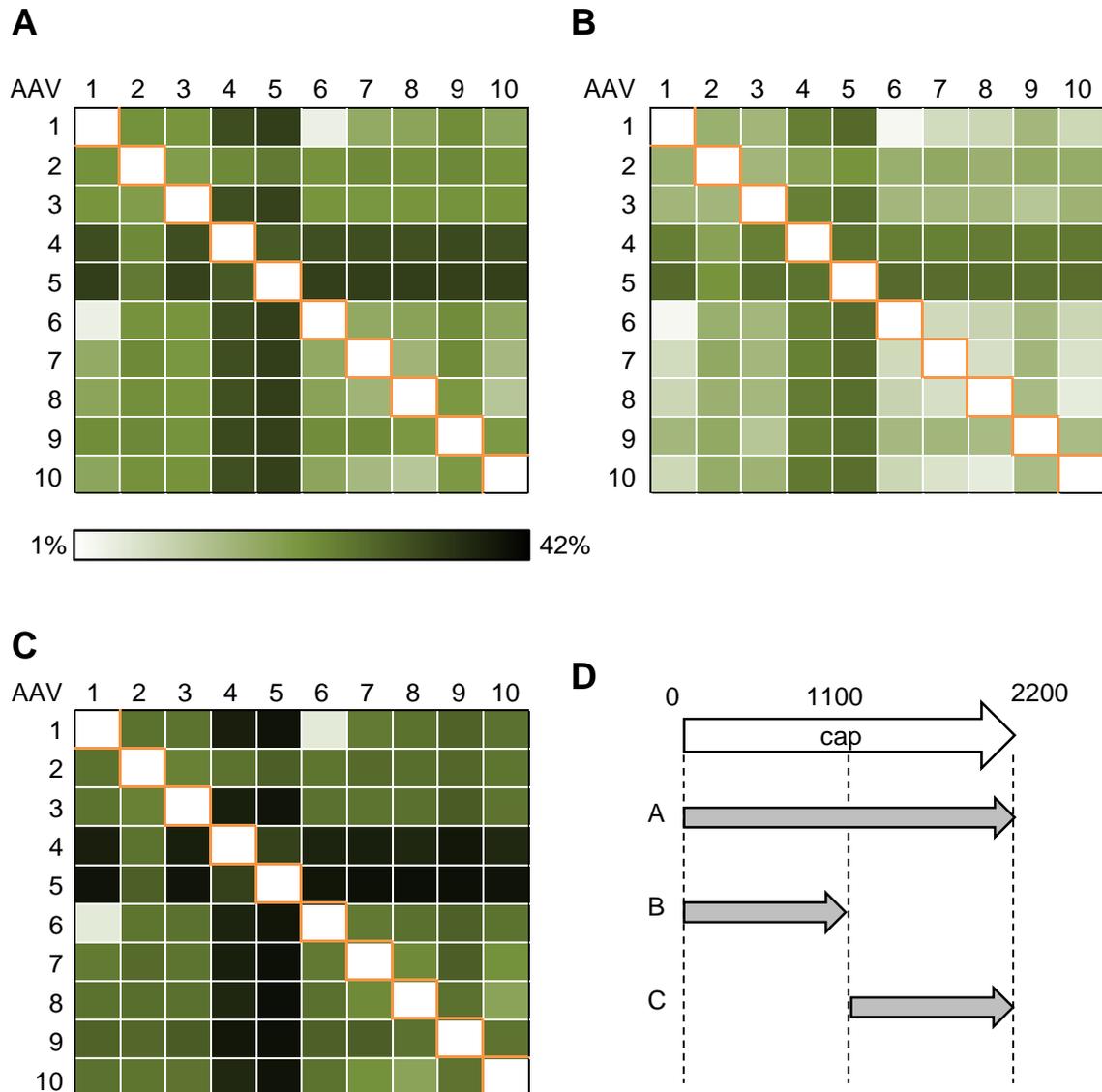


FIGURE 6-3: HOMOLOGY OF THE CAPSIDS

The homology of the nucleotide sequence of ten different AAVs used in this study is depicted. An alignment was analyzed in Salanto and the difference in percentage (%) is depicted, *i.e.* the higher the homology, the brighter the color gets. (A) The alignment used the complete cap gene and the data is depicted as a heatmap. The scale is depicted below and is valid for all graphs. Orange boxes indicate the same combination. (B) The first 1100 bp of the *cap* gene were taken for the alignment. (C) The remaining nucleotides were analyzed likewise. (D) The scheme depicts the party used for the respective alignments.

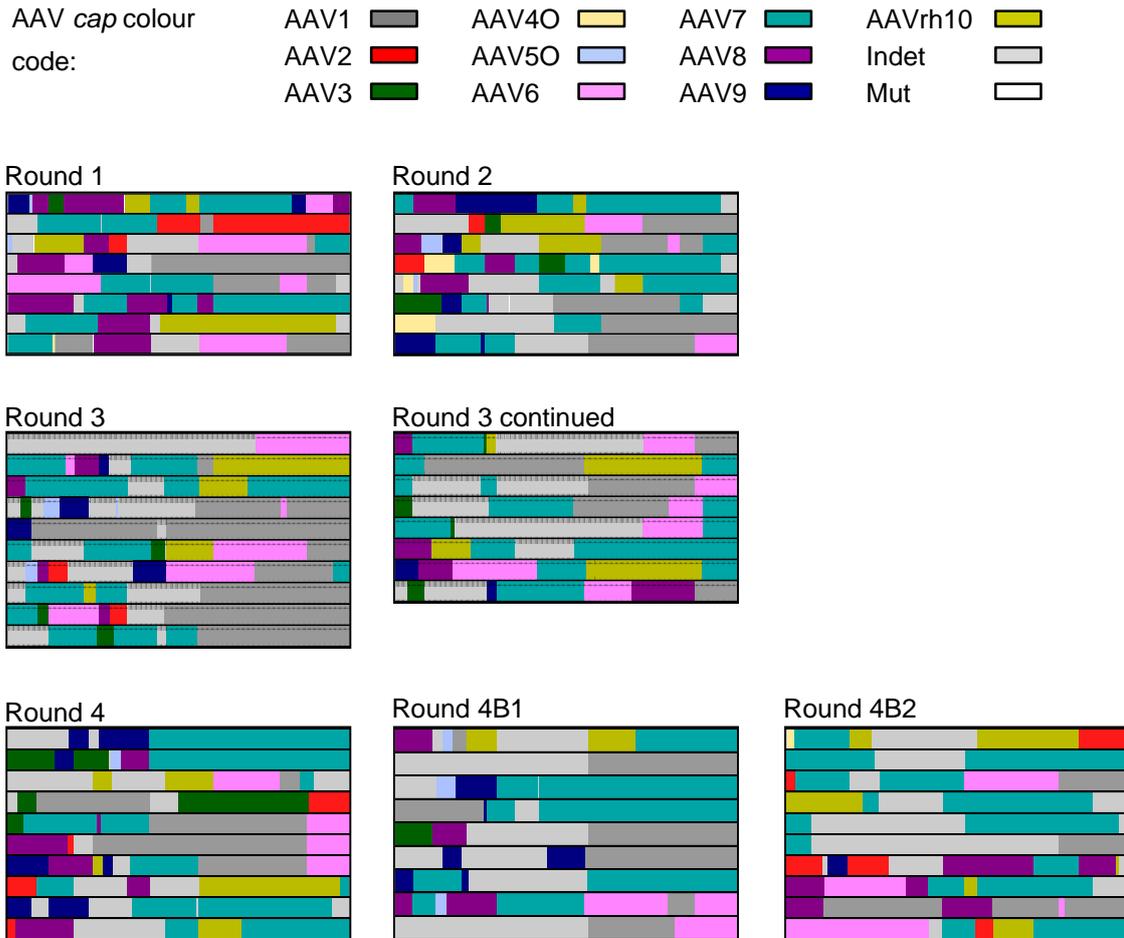


FIGURE 6-4: AMINO ACID SEQUENCES SORTED BY SELECTION ROUND

The amino acid sequences from the chimeras were monitored throughout the course of selection. Each bar represents the sequence of one clone and the parental distribution was colored according to the color scheme at the top. The fourth selection round was repeated with two more mice (B1 and B2).

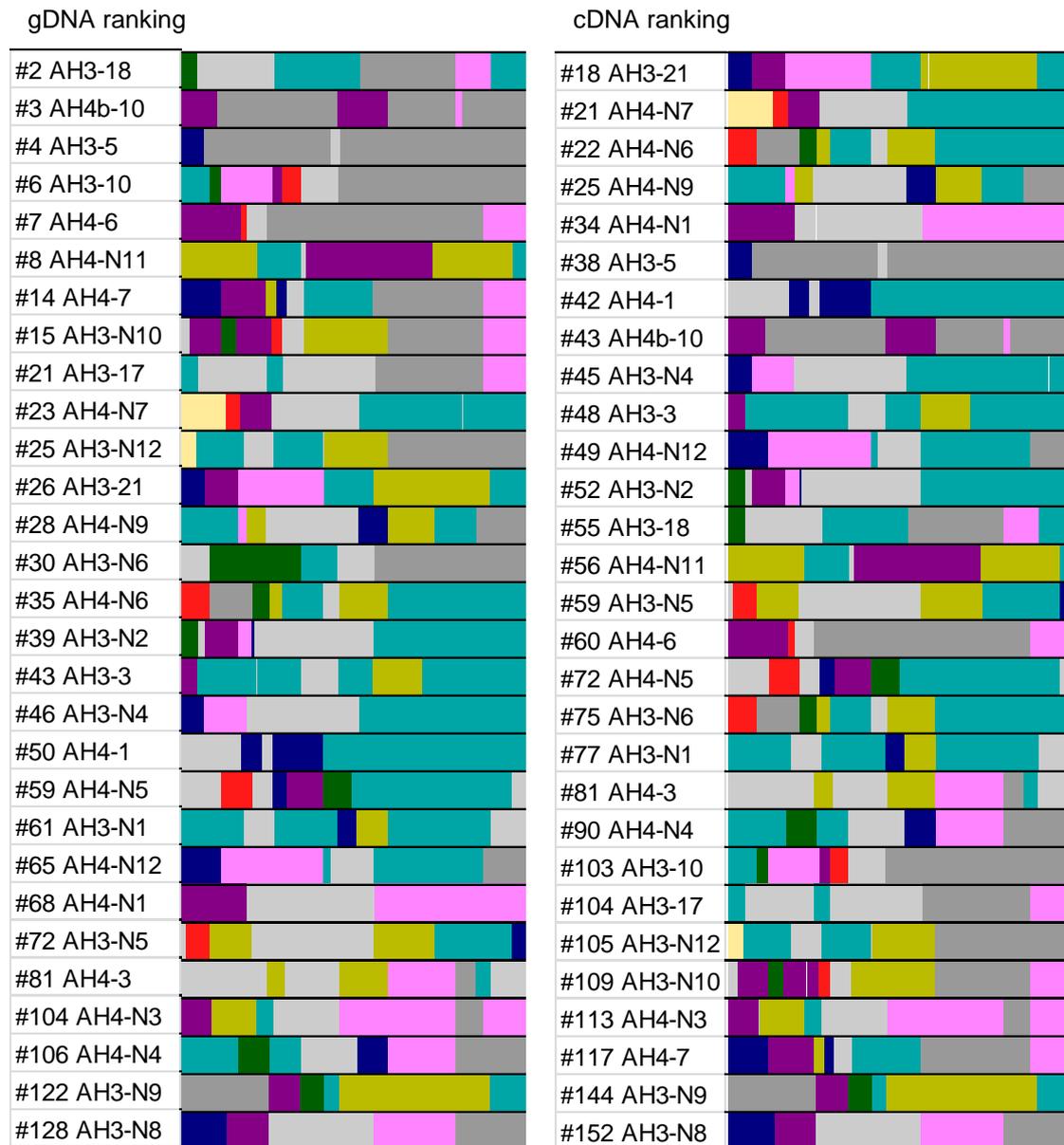


FIGURE 6-5: THE VARIANTS TESTED IN THE NGS SCREEN

The amino acid sequences of the variants tested within the NGS screen were assigned to their parental sequences. The numbers in front of the names indicate the position of the variant achieved in the ranking and the sequences were ranked accordingly. Variants which were newly sequenced to extent the repertoire are denoted with "N".

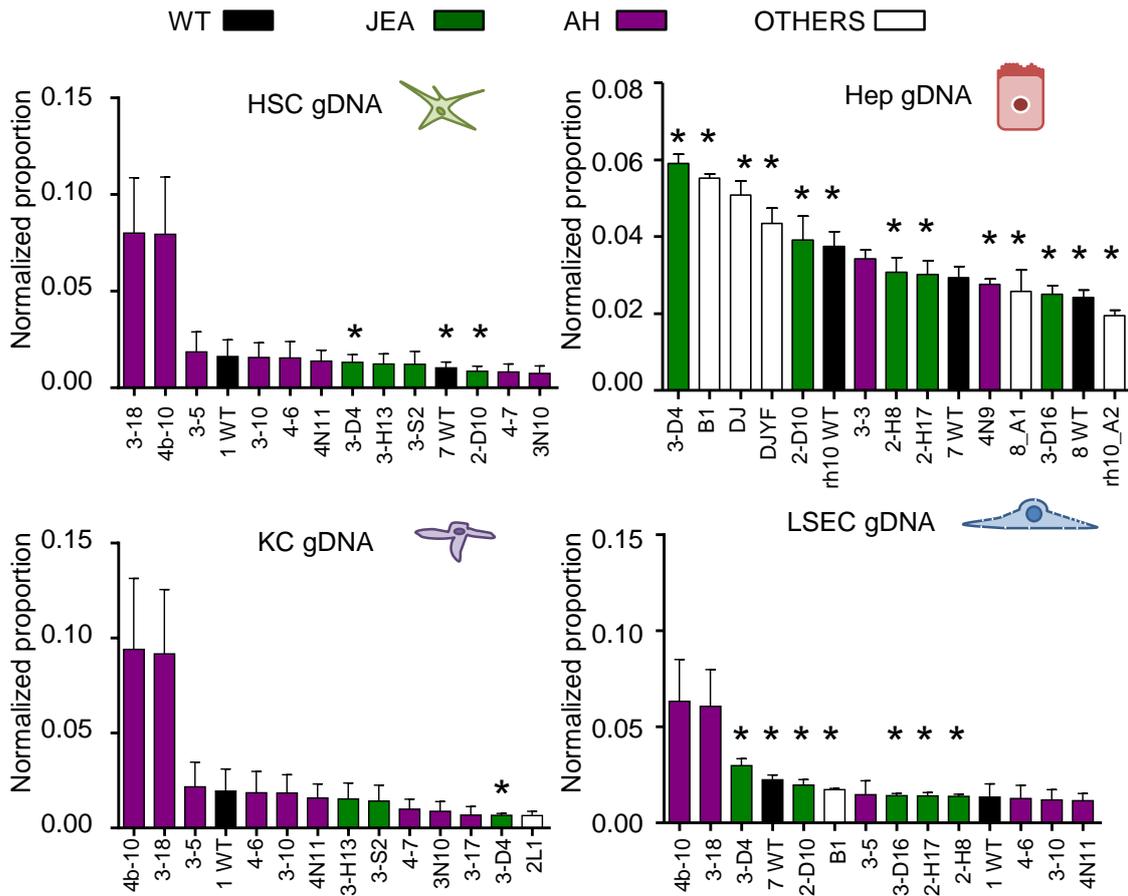


FIGURE 6-6: GDNA PROFILES WITHOUT AAV5

The graphs show the same dataset as in Figure 2-18 but for a better comparison AAV5 wild-type was removed from the analysis.

#### 6.1.4. OPTIMIZATION OF THE SELECTION PROCESS

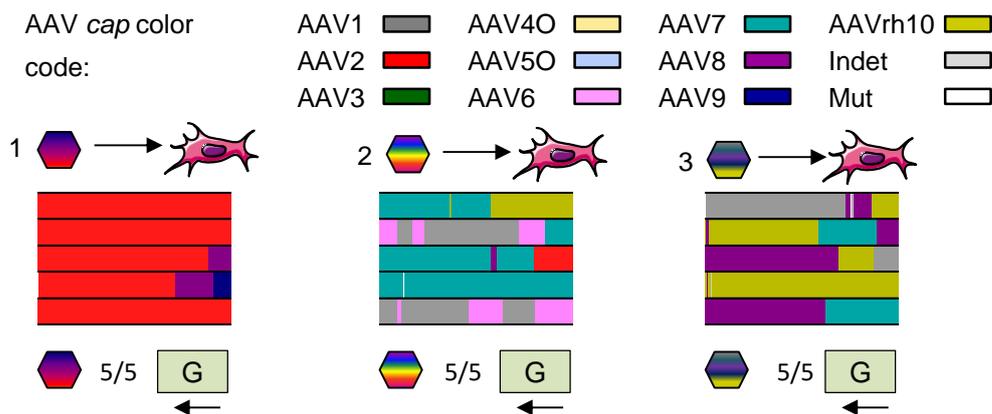


FIGURE 6-7: AMPLIFICATION WITH THE COMMON PRIMER

PCR rescue using a common primer was performed on single-infected samples. Three different libraries were used to separately transduce HEK293T cells and the *cap* genes were recovered for further subcloning. The approximate last 800 bp were sequenced and the parental sequences were assigned according to the color code with Salanto. Adapted from (Herrmann et al. manuscript in preparation).

## 6.2.EXCEL MACROS APPLIED

## 6.2.1. GENERATION OF COLORED TYPE ASSIGNMENTS

After the type assignment was copied into Excel the following macro was run, column width was set to 0.5 and the resulting colored bars were copied and pasted into Powerpoint as a picture. The size was adjusted as seen in the figures.

```

Dim Zelle As Range
For Each Zelle In ActiveSheet.UsedRange
If Zelle.Value = ">AAV1" Then
Zelle.Interior.ColorIndex = 16
End If
If Zelle.Value = ">AAV1" Then
Zelle.Interior.TintAndShade = 0.2
End If
If Zelle.Value = ">AAV2" Then
Zelle.Interior.ColorIndex = 3
End If
If Zelle.Value = ">AAV2" Then
Zelle.Interior.TintAndShade = 0.1
End If
If Zelle.Value = ">AAV3" Then
Zelle.Interior.ColorIndex = 51
End If
If Zelle.Value = ">AAV3" Then
Zelle.Interior.TintAndShade = 0.1
End If
If Zelle.Value = ">AAV4" Then
Zelle.Interior.ColorIndex = 44
End If
If Zelle.Value = ">AAV5" Then
Zelle.Interior.ColorIndex = 41
End If
If Zelle.Value = ">AAV5" Then
Zelle.Interior.TintAndShade = 0.1
End If
If Zelle.Value = ">AAV6" Then
Zelle.Interior.ColorIndex = 21
End If
If Zelle.Value = ">AAV6" Then
Zelle.Interior.TintAndShade = 0.7
End If
If Zelle.Value = ">AAV7" Then
Zelle.Interior.ColorIndex = 31
End If
If Zelle.Value = ">AAV7" Then
Zelle.Interior.TintAndShade = 0.1
End If
If Zelle.Value = ">AAV8" Then
Zelle.Interior.ColorIndex = 29
End If
If Zelle.Value = ">AAV8" Then
Zelle.Interior.TintAndShade = 0.02
End If
If Zelle.Value = ">AAV9" Then
Zelle.Interior.ColorIndex = 11
End If
If Zelle.Value = ">AAV9" Then
Zelle.Interior.TintAndShade = 0.1
End If

If Zelle.Value = ">AAVrh10" Then
Zelle.Interior.ColorIndex = 52
End If
If Zelle.Value = ">AAVrh10" Then
Zelle.Interior.TintAndShade = 0.3
End If
If Zelle.Value = ">AAVpo.1" Then
Zelle.Interior.ColorIndex = 10
End If
If Zelle.Value = ">AAVpo.1" Then
Zelle.Interior.TintAndShade = 0.1
End If
If Zelle.Value = ">AAV12" Then
Zelle.Interior.ColorIndex = 43
End If
If Zelle.Value = ">AAV12" Then
Zelle.Interior.TintAndShade = 0.1
End If
If Zelle.Value = ">AAV4GA" Then
Zelle.Interior.ColorIndex = 44
End If
If Zelle.Value = ">AAV4SO" Then
Zelle.Interior.ColorIndex = 44
End If
If Zelle.Value = ">AAV4SO" Then
Zelle.Interior.TintAndShade = 0.6
End If
If Zelle.Value = ">AAV4GA" Then
Zelle.Interior.TintAndShade = 0.6
End If
If Zelle.Value = ">AAV5GA" Then
Zelle.Interior.ColorIndex = 41
End If
If Zelle.Value = ">AAV5GA" Then
Zelle.Interior.TintAndShade = 0.6
End If
If Zelle.Value = ">AAV5SO" Then
Zelle.Interior.ColorIndex = 41
End If
If Zelle.Value = ">AAV5SO" Then
Zelle.Interior.TintAndShade = 0.6
End If
If Zelle.Value = "indet" Then
Zelle.Interior.ColorIndex = 15
End If
If Zelle.Value = "indet" Then
Zelle.Interior.TintAndShade = 0.2
End If
If Zelle.Value = "MUT" Then
Zelle.Interior.ColorIndex = 2
End If
Next Zelle

End Sub

```

## 6.2.2. EXTRACTION OF TRANSDUCTION DATA

To test for the transduction efficiency of AAVs having packaged a GFP reporter, FACS analysis was performed. After the run was complete the data was exported as an Excel sheet by the accompanying MXP software. In the past, Joel Beaudouin designed a macro to extract different values such as "%gated" on the GFP channel or median intensity to simplify data analysis.

```

Dim i As Integer
Dim j As Integer

Sheets.Add
ActiveSheet.Name = "result"

For i = 1 To 12
For j = 1 To 8
    Worksheets("result").Cells(j, i).Value =
Worksheets("Result 1").Cells(9 + ((i - 1) + (j - 1) * 12) * 7,
8)
Next
Next

Sheets.Add
ActiveSheet.Name = "result_positive cells"

For i = 1 To 96
'For j = 1 To 8
    Worksheets("result_positive cells").Cells(i, 1).Value =
Worksheets("Result 1").Cells(9 + ((i - 1)) * 7, 8)
'Next

Next
Next

Sheets.Add
ActiveSheet.Name = "result_x-mean"

For i = 1 To 96
For j = 1 To 8
    Worksheets("result_x-mean").Cells(i, 1).Value =
Worksheets("Result 1").Cells(9 + ((i - 1)) * 7, 10)
Next
Next

Sheets.Add
ActiveSheet.Name = "result_x-median"

For i = 1 To 96
For j = 1 To 8
    Worksheets("result_x-median").Cells(i, 1).Value =
Worksheets("Result 1").Cells(9 + ((i - 1)) * 7, 9)
Next
Next
End Sub

```

## 6.3. PYTHON SCRIPTS APPLIED

### 6.3.1. REVERSE COMPLEMENTING SEQUENCING DATA

In order to detect the library-specific barcode (section 4.2.3.3) the plasmids were sequenced with a reverse primer. As numerous sequencing files had to be regarded a code was kindly written by Mitja Kleider which extracted all sequences in a given folder to a separately saved text file in the reverse complemented orientation.

```

#!/usr/bin/env python
# coding: utf-8

"""
Merge all .seq files in a directory into one text file, ignore text behind ".seq"
"""

import glob
import re
import os

try:
    # Python2

```

```

import tkinterFileDialog as fd
except:
# Python3
import tkinterFileDialog as fd

# which directory contains .seq files?
directory = fd.askdirectory()

result = []
for filename in glob.glob(os.path.join(directory, "*.seq")):
    with open(filename) as f:
        lines = f.readlines()
        assert len(lines) == 2
        result.append(">{}\n".format(os.path.basename(filename)))
        reverse = lines[1].replace("\n", "")[:-1]
        complement = []
        for letter in reverse:
            position = "ACTGactgNn".find(letter)
            if position == -1:
                raise ValueError("unknown letter: {}".format(letter))
            complement.append("TGACtgacNn"[position])
        complement.append("\n")
        result.append(".join(complement))
# empty line between files
result.append("\n")

# write result to file
with open(os.path.join(directory, "seqdata_reversecomplement.txt"), "w") as f:
    f.writelines(result)

```

### 6.3.2. BARCODE EXTRACTION FROM SEQUENCING DATA

A second code by Mitja Kleider was written to sort the sequences into individual text files according to their barcode. Therefore, in an empty folder the text documents “patterns” and “source” were collected together with the code. “Patterns” contained the barcode name and sequence and “source” supplied the sequences to be analysed. After the script was run separate text files for each barcode were generated and in addition all sequences which could not be assigned were collected in “unknown”. These sequences were analyzed manually with the help of the chromatogram.

```

#!/usr/bin/env python

"""

"""

barcodes = {}

# load patterns from patterns.txt
with open("patterns.txt", "r") as f_patterns:
    for line in f_patterns.readlines():
        # each line is one barcode, <name> <barcode>
        name, code = line.split(" ")
        # TODO check barcode for validity
        barcodes[code.rstrip().upper()] = name

output_files = {}
for name in barcodes.values():
    output_files[name] = open(name+".txt", "w")
output_files["unknown"] = open("unknown.txt", "w")

```

```
# load sequences from input.txt
with open("input.txt", "r") as f_input:
    for line in f_input.readlines():
        if line == "\n":
            continue
        if line[0] == ">":
            sequence_name = line
        else:
            sequence = line.upper()

        # look for patterns in sequence
        num_found = 0
        for code in barcodes.keys():
            if sequence.find(code) != -1:
                # write sequence to <name of pattern>.txt
                name = barcodes[code]
                output_files[name].write(sequence_name)
                output_files[name].write(sequence)
                output_files[name].write("\n")
                num_found += 1
        assert num_found < 2, num_found
        if num_found == 0:
            # write unknown pattern to unknown.txt
            output_files["unknown"].write(sequence_name)
            output_files["unknown"].write(sequence)
            output_files["unknown"].write("\n")
```



---

## ACKNOWLEDGMENTS

---

At last I would like to thank different people. First of all, I would like to thank Prof. Dr. Dirk Grimm for welcoming me in the group and for giving me the opportunity to work on such interesting projects. Thanks a lot for supporting my work with ideas, collaborations and fruitful discussions! I made a lot of valuable experiences over the years. Furthermore, I am very grateful that I was given the chance to visit conferences and most importantly, that part of my work was published. In addition, I was very happy that I was chosen for the open PhD position within the SFB1129 back then from which I gratefully achieved funding.

I would like to thank Prof. Dr. Ralf Bartenschlager for taking the time of being my Erstgutachter and for providing helpful advice in my thesis advisory committees. Additionally, I would like to thank Dr. Petr Chlanda and Dr. Pierre-Yves Lozach that they participate in my defense committee. At this point I would also like to thank Ann-Kristin Mueller that she joined my TAC meetings during the past years. Along these lines I would like to thank HBIGS, especially Martina Galvan and Rolf Lutz, for support and quick replies.

As mentioned above, I had the opportunity to collaborate with different groups and I would like to express my gratitude to them. I would like to thank Holger Willenbring and Regina Espanol-Suner that we performed the selection on stellate cells together. As well, I am really thankful for the support we obtained during validation from the lab of Steven Dooley and especially Christof Dormann as well as Bedair Dewidar who helped with the experiments. Two of the promoters were kindly provided by Antonia Follenzi which I highly appreciate and which will become important in the future. I appreciated the nice contacts and that I was given the chance to learn more about cell isolations.

Life in the lab was made a lot easier by the good infrastructure present in the Bioquant which I would like to acknowledge at this point. Thank you, Frau Graffy, that we could always call you if something was off with the machines. Likewise, thanks a lot Daniel (Browne) that you always came when I had computer issues. Lastly, everyday lab life was facilitated a lot by Herr Schawerna and Yvonne (Hess). Thanks a lot for taking care of autoclaving and supplying us with clean glass ware.

Over the years I had the chance to supervise a number of students and I would like to thank two of them at this point. Thanks a lot, Micha (Rosenkranz) that you helped out with cloning when I was at the near end of my PhD. Niklas (Beumer), I highly respect your bioinformatical skills and thanks for providing me with the sequences.

The working atmosphere in the lab was always productive and I really enjoyed working there. Thanks a lot to all previous and current members of the Grimm lab! You are awesome and I really enjoyed numerous lunch breaks as well as summer parties with the famous Barbed Wire. A huge thank you (!) is reserved for two of our lab angels. Thanks so much Ellen (Wiedtke) that you deal with all of our ordering and that you (nearly) always knew an answer to my numerous lab questions. Chiara (Krämer), thanks a lot for your help! I will remember all of these pipetting parties in the cell culture hoods when we had mass productions for the papers ☺. It was, and still is, lots of fun with you! From the past members I would like to especially thank Stefanie (Große) as she introduced me into the world of DNA family shuffling when I was still a “Baby-PhD” ☺. Franzi (Hentzschel), it was lots of fun being your bench/ office neighbor and thanks for cheering me up when I realized that all of a sudden I grew up to a “Senior-PhD”. One subchapter of this thesis would have not been possible without the help of Jonas (Weinmann) who established the NGS workflow. Thanks a lot for your help! I really enjoyed teaming up with you and all the interesting discussions we had. ☺ Janina (Haar), thanks a lot for having an open ear for me when I started writing the thesis and had lots of questions associated with it. Thanks also to Caro, Julia, Claire, Jad, Kleo and Cinja for creating such a nice atmosphere! I really appreciated your feedback in lab seminars and scientific discussions.

Mein Dank gilt auch meinen Freundinnen, Aysa, Sarah und Emma. Ich finde es so schön, dass wir es seit der Schulzeit schaffen den Kontakt aufrecht zu erhalten und immer noch viel zu Lachen haben. Ronny, es war damals so toll, dass wir beide in Heidelberg für den Master angenommen wurden und wir hatten echt die beste WG. ☺ Ich finde es toll, dass ich regelmäßig Lunch-Dates mit „Dukke“ habe.

Mit Beginn meiner Masterarbeit habe ich auch meine große Liebe kennen gelernt. Danke Mitja, dass du mich zum Lachen bringst und für dein Verständnis, deine Unterstützung und deine Liebe. Ich möchte mich auch nochmal für die von dir programmierten Scripts bedanken, die ersparen mir so unglaublich viel Zeit und frustriertes Klicken am Computer. ☺

Mein größter Dank allerdings geht an meine Eltern. Mama, Papa, ohne euch hätte ich das hier alles nicht geschafft!!! Hättet ihr damals nicht so sehr an mich geglaubt, wäre ich nicht aufs Gymnasium gegangen. Danke für eure Unterstützung bei allen meinen Entscheidungen! Danke für eure Liebe und danke, dass ihr mir geholfen habt zu der Person zu werden, die ich jetzt bin. Ihr seid die Besten!