Dissertation

submitted to the

Combined Faculties for the Natural Sciences and Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

presented by

M. Sc. Jonas Weinmann

born in Rottweil, Germany

Oral examination: October 2nd, 2018

Massively parallel *in vivo* characterization of novel adeno-associated viral (AAV) capsids using DNA/RNA barcoding and next generation sequencing

Referees: Prof. Dr. Ana Martin-Villalba

Prof. Dr. Dirk Grimm

ABSTRACT

In recent years, the adeno-associated virus (AAV) gained considerable attention mainly due to the approval of the first AAV-based gene therapy treatment in the Western hemisphere in 2012, named Glybera®. It not only conveyed the feasibility of utilizing this parvovirus to introduce healthy gene copies but simultaneously reinforced further interest in developing more specific and efficient synthetic vectors by capsid engineering approaches such as DNA family shuffling or random peptide display. However, the characterization of lead candidates resulting from these directed evolution strategies is labor-intensive and therefore excludes the possibility to validate multiple promising variants.

Therefore, a comprehensive high-throughput capsid validation pipeline was established in this work adapting a previously reported approach in which a DNA barcode-comprising AAV genome is assigned to a chosen capsid variant during virus production. Thus, the identification of the respective capsid in the complex physiological environment of living animals is enabled by solely detecting the barcode sequence via next generation sequencing. The principle was further improved by placing the barcode into the 3'UTR of a CMV promoter-driven *eyfp* transgene permitting tracking on the DNA and RNA level. Hence, next to information about transduction efficiency, the especially crucial transcriptional activity in a certain tissue was measured. Using this design, three barcoded AAV libraries were generated comprising up to 157 variants including 12 commonly used serotypes, >70 peptide-displaying mutants based on these naturally occurring wild types and several published benchmarks such as AAVDJ, AAV9_PHP.B and AAVAnc80L65. After intravenously injecting the library into C57BL/6J mice and analyzing the RNA and DNA data from >20 collected tissues, prior observations for the literature variants could be confirmed thus validating the workflow. Most impressively, a peptide display mutant previously created in our laboratory exhibited drastically improved efficiencies in the diaphragm, heart and skeletal muscles in comparison to AAV9wt on the cDNA and protein level while in addition demonstrating pronounced muscle specificity.

In conclusion, in the course of this PhD thesis a highly robust barcode-based capsid screening pipeline was established that facilitates and accelerates the identification of promising candidates for gene therapies, best exemplified by the discovery of the muscle-tropism of our lead candidate.

ZUSAMMENFASSUNG

In den letzten Jahren erfuhr das Adeno-assoziierte Virus (AAV) viel Aufmerksamkeit, insbesondere im Jahr 2012 durch die Zulassung von Glybera®, der ersten AAV-basierten Gentherapie in der westlichen Hämisphere. Dadurch wurde nicht nur die erfolgreiche Verwendung dieser Parvoviren zur Einführung gesunder Genkopien demonstriert, sondern auch das Interesse an der Entwicklung spezifischerer und effizienterer Vektoren durch Modifizierung des Kapsides verstärkt. Allerdings ist die Charakterisierung von einzelnen aussichtsreichen Kandidaten besonders arbeitsintensiv, was die Validierung mehrerer Viren erschwert.

Demzufolge wurde in dieser Arbeit ein umfassendes Hochdurchsatz-Kapsid-Validierungssystem etabliert, welches einen zuvor beschriebenen Ansatz adaptiert, bei dem ein DNA-barcodiertes AAV-Genom während der Virusproduktion einem ausgewählten Kapsid zugewiesen wird. Somit wird die Identifizierung des jeweiligen Kapsides in der komplexen physiologischen Umgebung lebender Tiere ermöglicht, indem lediglich die Barcode-Sequenz über Next-Generation Sequenzierung detektiert wird. Durch die Integrierung des Barcodes in die 3'-UTR eines CMV-Promotor-gesteuerten *eyfp* -Transgens wurde der Ansatz weiter verbessert, was eine Detektion auf DNA- und RNA-Ebene ermöglichte. Neben der Transduktionseffizienz wird dadurch zudem die äußerst wichtige Transkriptionsaktivität in einem bestimmten Gewebe gemessen. Unter Verwendung dieses Designs wurden drei barcodierte AAV-Bibliotheken mit bis zu 157 Varianten generiert einschließlich 12 häufig verwendeter Serotypen, >70 Peptid-präsentierender Mutanten auf Basis dieser natürlich vorkommenden Wildtypen und mehrerer veröffentlichter Viren wie AAVDJ, AAV9_PHP.B und AAVAnc80L65. Nach intravenöser Injektion der Bibliothek in C57BL/6J-Mäuse und Analyse der RNA- und DNA-Daten von >20 isolierten Geweben konnten vorherige Beobachtungen für die Literaturvarianten bestätigt werden, wodurch das System validiert wurde. Eine Peptid-präsentierende Mutante unseres Labors zeigte eindrucksvoll eine drastisch verbesserte Effizienz im Vergleich zu AAV9wt in der Zwerchfell-, Herz- und Skelettmuskulatur auf cDNA- und Proteinebene und gleichzeitig eine ausgeprägte Muskelspezifität.

Zusammenfassend wurde im Rahmen dieser Dissertation ein äußerst robustes Barcode-basierendes Kapsid-Validierungssystem etabliert, welches vielversprechende Kandidaten für Gentherapien identifizieren kann, am besten verdeutlicht durch die Entdeckung unseres Muskelkandidaten.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

%	Percent
°C	Degree Celsius
µF	Microfarad
µg	Microgram
µl	Microliter
µm	Micrometer
µM	Micromolar
A	Alanine
aa	Amino acid
AAP	Assembly-activating protein
AAV	Adeno-associated virus
AAVR	AAV receptor
Ad	Adenovirus
AH	Anne-Kathrin Herrmann
ASGCT	American Society of Gene and Cell Therapy
ATP	Adenosine triphosphate
BC	Barcode
BGH	Bovine growth hormone
BI	Boehringer Ingelheim
bp	Base pair
BSA	Bovine serum albumin
C	Cysteine
$CaCl_2$	Calcium chloride
CAR-T	Chimeric antigen receptor T
CBA	Chicken β-actin
cDNA	Complementary DNA
cm	Centimeter
CMV	Cytomegalovirus
CO_2	Carbon dioxide
CRE	*Cis*-regulatory element
$CsCl_2$	Cesium chloride
D	Aspartic acid
dg	Diploid genomes
DMEM	Dulbecco's Modified Eagle's Medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate

DTT	Dithiothreitol
E	Glutamic acid
E.coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EGFP	Enhanced green fluorescent protein
EtOH	Ethanol
EYFP	Enhanced yellow fluorescent protein
F	Phenylalanine
Fab	Fragment antigen binding
g	Gram
G	Glycine
gDNA	Genomic DNA
h	hour
H	Histidine
H_2O	Water
HCl	Hydrochloric acid
hFIX	Human factor IX
HSPG	Heparan sulfate proteoglycan
I	Isoleucine
i.v.	Intravenously
ITR	Inverted terminal repeats
JEA	Jihad El Andari
K	Lysine
kb	Kilobases
KCl	Potassium chloride
kDa	Kilodalton
KH_2PO_4	Monopotassium phosphate
L	Leucine
LB	Lysogeny broth
LD	Liver-detargeted
LSEC	Liver sinusoidal endothelial cells
M	Molar or methionine
MACS	Magnetic-activated cell sorting
$MgCl_2$	Magnesium chloride
$MgSO_4$	Magnesium sulfate
min	Minute
ml	Milliliter
mM	Millimolar
mRNA	Messenger RNA
Mut	Mutant
N	Asparagine

Na_2HPO_4	Disodium phosphate
NaCl	Sodium chloride
NaOH	Sodium hydroxide
ng	Nanogram
NGS	Next generation sequencing
NHEJ	Non-homologous end joining
NIS	New insertion site
NLS	Nuclear localization signal
nM	Nanomolar
NPC	Nuclear pore complex
NSC	Neural stem cells
nt	Nucleotides
ORF	Open reading frame
P	Proline
PBS	Phosphate-buffered saline
PCR	Polymerase chain reaction
PEG	Polyethylene glycol
PEI	Polyethylenimine
PFA	Paraformaldehyde
PLA2	Phospholipase A2 domain
pM	Picomolar
po	porcine
Q	Glutamine
qPCR	Quantitative PCR
R	Arginine
rAAV	Recombinant AAV
RBE	Rep binding element
RBS	Rep binding site
rcf	Relative centrifugal force
rh	rhesus
RI	Refractive index
RNA	Ribonucleic acid
rpm	Rounds per minute
RT	Room temperature
S	Serine
SA	Sialic acid
scAAV	Self-complementary AAV
SD	Standard deviation
sec	Second
SOB	Super Optimal Broth
SOC	Super Optimal broth with Catabolite repression

SVZ	Subventricular zone
T	Threonine
TAE	Tris-acetate-EDTA
TE	Tris-EDTA
trs	Terminal resolution site
U	Units
UTR	Untranslated region
UV	Ultraviolet
V	Volt or valine
v/v	Volume per volume
vg	Viral genomes
VP	Viral protein
VR	Variable region
W	Tryptophan
w/v	Weight per volume
WH	Wilson helper plasmid
wt	Wild type
Y	Tyrosine
β-ME	β-Mercaptoethanol

1 INTRODUCTION

Genetic disorders such as hemophilia, cystic fibrosis and Parkinson's disease pose a serious problem for the affected patients. Unlike other reversible, temporary diseases, e.g. many viral or bacterial infections, the malfunction manifests itself in the genome of our cells making the search for a permanent cure highly challenging. Nevertheless, researchers across the world tackled the issue and started to develop treatment strategies commonly referred to as gene therapy approaches. These methods aim to alter or circumvent the genetic mutation by introducing DNA sequences comprising a healthy gene copy or tools required for genome modifications. Various ways to penetrate the nonpolar cell membrane were tested for the delivery, and each of them showed advantages and disadvantages.

One possibility is the use of cationic liposomes possessing a positively-charged head for interaction with the negatively-charged DNA and a hydrophobic lipid tail inducing the formation of particles[1,2]. Liposomes are then capable of entering the cells by endocytosis[3]. The principle of condensing the DNA can also be applied when using cationic polymers such as poly-L-lysine or polyethylenimine[4,5]. Next to chemical methods, successful DNA delivery to various cell types was shown for procedures relying on physical mechanisms, e.g. electroporation[6,7], sonoporation[8–10], gene guns[11] and hydrodynamic gene transfer[12–14]. In general, all the aforementioned strategies lack efficiency, especially regarding *in vivo* delivery. This is due to the manifold challenges that are faced prior to nuclear transcription, for instance, passing the endothelium and cell membrane, escaping the endosome, trafficking through the cytoplasm and finally entering the nucleus[15–17]. The design of synthetic delivery tools has to be adapted to overcome these roadblocks that hamper overall efficiency.

Next to nonviral approaches, viruses are nowadays the preferred delivery vector since they have already been optimized by nature for successful cell infection and processing of their cargo. Notable examples are altered retroviruses as they are capable of undergoing reverse transcription and DNA integration. Modifications to their genome and the producer cell line were made to generate replication-incompetent vectors that display an increased safety profile[18,19]. Gammaretrovirus is a genus in the *Retroviridae* family and was proved to transduce hematopoietic stem cells[20,21] and primary T-lymphocytes[22,23]. This has eventually led to the approval of an *ex vivo* stem cell gene therapy treatment called Strimvelis®[24]. The latter is aimed at curing the

very rare disease Severe Combined Immunodeficiency due to Adenosine Deaminase deficiency which is impairing the development of the immune system in children. Another member of the *Retroviridae* family, the lentivirus, is able to carry larger gene cassettes and integrates into coding regions of genes in contrast to the gammaretrovirus which is targeting the 5′-untranslated region[25]. The latter poses a greater risk of causing oncogenic mutagenesis in respective cells[26,27]. Hence, lentiviral vectors have increasingly attracted considerable attention best exemplified by the development of chimeric antigen receptor T-cell therapies (CAR-T). To this end, lentiviruses and, to a lesser extent, gammaretroviruses are used to stably express receptors on T cells directed against antigens on the surface of cancer cells[28–31]. One of the most studied viral vectors are adenoviruses (Ad) owing to their robust transduction profile, especially in the liver. However, gene correction with Ad led to the tragic death of an 18-year-old patient after systemic inflammatory response syndrome, triggered by the virus itself[32]. Although further modifications were performed helping to target other tissues next to the liver and evading host immune system responses[33], arguably the most promising virus for gene therapy is the adeno-associated virus (AAV). Due to its dependence on a helper virus for replication, AAV's safety profile is already naturally advantageous and can be further enhanced by omitting the encapsidation of wild type AAV genes preventing genome integrations. On top, AAV possesses the ability to transduce multiple cell and tissue types, which makes it highly interesting for the therapy of many diseases. The first AAV-based treatment, Glybera®, developed by uniQure, was approved in 2012 for the European market and aimed to restore the rare genetic disease lipoprotein lipase deficiency[34]. Yet, in October 2017, after injecting only one patient, the license of Glybera® was not renewed since the treatment expenses per person amount to one million euros making the enterprise unprofitable. Nevertheless, uniQure has led the way by proving the feasibility of an AAV gene therapy. In December 2017, Spark Therapeutics released LUXTURNA™ to treat *RPE65* mutation-associated retinal dystrophy[35] and despite the again striking price tag of ~$450,000 per eye, a steadily growing number of clinical trials involving AAV-based solutions have been registered increasing the chances for the approval of treatment options for genetic diseases in the future.

1.1 ADENO-ASSOCIATED VIRUS (AAV)

The discovery of the non-enveloped adeno-associated virus dates back over 50 years to 1965 when it was first described as a contamination of adenovirus stocks. The publication of Atchison *et al.* moreover reported a replication

deficiency when adenovirus was absent[36]. Today the replication dependency of AAV on helper viruses such as the already mentioned adenovirus, herpes simplex virus[37] or human papilloma virus[38] is commonly known, justifying the assignment of AAV to the dependoparvovirus genus within the *Parvoviridae* family. AAV is one of the smallest known viruses with a capsid diameter of only ~22 nm, sterically limiting its genome size to ~4.7 kb. The genome itself was identified as single-stranded DNA back in 1969[39]. Important for gene therapy applications, AAV is to date considered to be in principle non-pathogenic, although debatable evidence has been found suggesting AAV integration to cause hepatocellular carcinoma[40–42]. Over the years, hundreds of isolates could be identified in various species, and some of them were classified as novel serotypes. AAV1 and the very extensively studied AAV2 were the first to be discovered[43] followed by AAV4[44], AAV5[45], AAV6 and AAV3b[46], AAV7 and AAV8[47], AAV9 and AAVrh10[48], AAV12[49] and finally AAVpo1[50]. Regarding the amino acid sequence of their respective capsid proteins, AAV serotypes are largely homologous to each other (Figure 1).

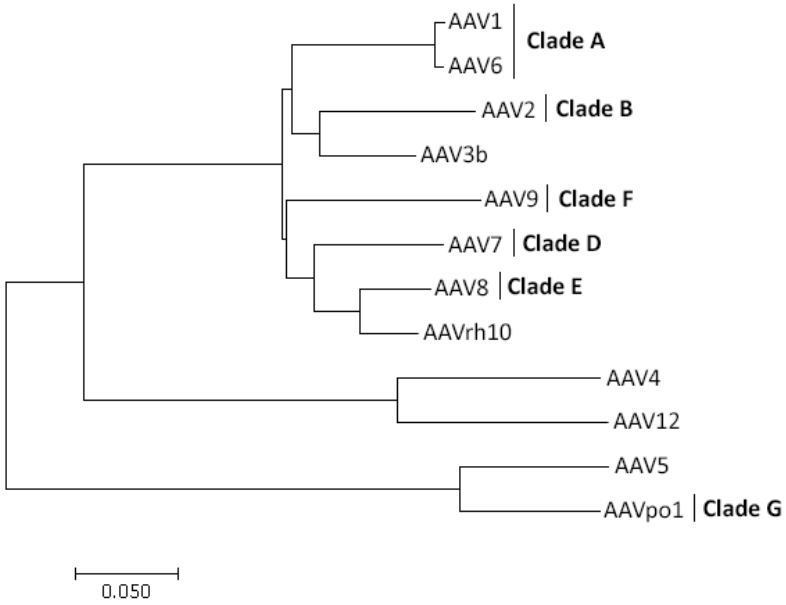

Figure 1: Phylogenetic tree of popular AAV serotypes
Phylogenetic tree of the respective VP1 amino acid sequences of the primarily used serotypes in our laboratory AAV1, AAV2, AAV3b, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAVrh10, AAVpo1 and AAV12. Additional members of the respective groups with a common ancestor (clades) are not shown.

The highest relatedness is observed for the non-human primate isolate AAV1 and the human variant AAV6 that vary in only six amino acids (99% homology). Two pairs in the phylogenetic tree, AAV4 and AAV12 as well as AAV5 and AAVpo1 are substantially more different compared to serotypes

from clade A, B, D, E and F (Table 1). The homology of the remaining AAV1, AAV2, AAV3b, AAV6, AAV7, AAV8, AAV9 and AAVrh10 is around ~85%.

Table 1: Homology[a] of AAV serotypes

AAV	1	2	3b	4	5	6	7	8	9	rh10	po1	12
1	100											
2	83	100										
3b	87	88	100									
4	64	61	64	100								
5	59	58	59	53	100							
6	99	83	87	64	59	100						
7	85	82	85	64	59	85	100					
8	84	83	86	64	58	84	88	100				
9	83	82	84	63	57	82	81	85	100			
rh10	85	84	86	64	58	85	89	94	86	100		
po1	59	58	59	53	86	59	59	58	57	57	100	
12	61	60	62	79	53	61	62	62	60	61	52	100

[a]Describes the VP1 homology in %.

1.1.1 GENOME ORGANIZATION AND TRANSCRIPTOME

The 4.7 kb-long AAV genome harbors two genes, *rep* and *cap*, flanked by 145 bp-long inverted terminal repeats (ITR) (Figure 2). A palindromic sequence within the ITR forms a T-shaped hairpin structure[51]. Furthermore, the ITR contains *cis*-elements required for replication and packaging of the genome [52]. Between the 5′ and 3′ ITR, the internal promoters p5 and p19 govern the transcription of transcripts coding for Rep78 and Rep68 as well as Rep52 and Rep40, respectively[53]. The resulting proteins are involved in various steps of the AAV infection cycle (1.1.3). The third promoter, p40, is driving the transcription of mRNAs encoding the three capsid proteins, VP1, VP2, VP3 and the assembly-activating protein (AAP)[54–56]. All primary transcripts utilize the same polyadenylation signal downstream of *cap* and carry an intron positioned between *rep* and *cap*. The donor site D and the two acceptor sites A1 and A2 surrounding the intron allow alternative splicing of the Rep and VP mRNAs (Figure 2). Unspliced transcripts of the p5 and p19 promoter lead to expression of Rep78 and Rep52, respectively. Splicing with either A1 or A2 as an acceptor site results in Rep68 for the p5 transcript or Rep40 for the p19 mRNA[53]. A1-splicing for p40-driven transcripts enables VP1 expression by using a regular AUG start codon. VP2 and VP3 are both encoded on the same A2-spliced mRNA but differ in their start codon. VP2 translation starts with the unusual

and inefficient ACG codon whereas VP3 utilizes AUG[55,57]. The discrepancy of the start codons and the fact that A1-splicing occurs less frequently than A2-splicing explains the VP1:VP2:VP3 ratio of 1:1:10[58]. AAP expression is enabled by exploiting an alternative open reading frame (ORF) of *cap* and the highly uncommon initiation codon CUG between VP2 and VP3[56]. The resulting protein is required for the formation of the capsid and is currently studied extensively[56,59–63].

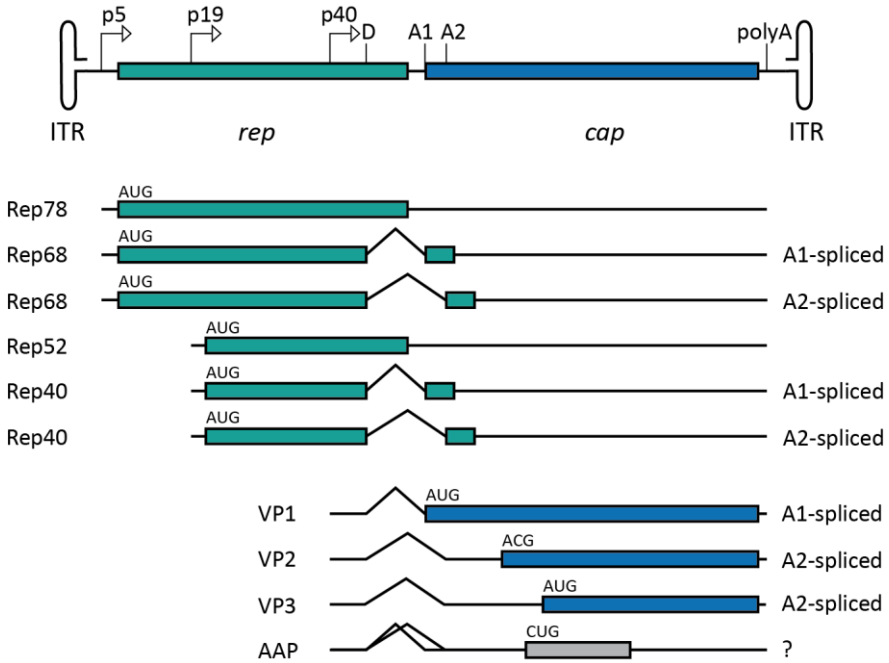

Figure 2: AAV genome organization and transcriptome
Depicted is the AAV genome with *rep* and *cap* genes flanked by ITRs. Promoters (arrows) p5 and p19 are driving the transcription of mRNAs encoding Rep78/Rep68 and Rep52/Rep40, respectively. The mRNA for the capsid proteins VP1, VP2 and VP3 as well as AAP is generated by p40 activity. All intron-containing transcripts can be unspliced or spliced either with the A1 or A2 acceptor site and the common splice donor (D). Translation of capsid proteins is controlled by splicing efficiencies and unconventional start codons in the case of VP2 and AAP. On top, AAP is using an alternative open reading frame in *cap*.

1.1.2 CAPSID STRUCTURE AND RECEPTORS

Over the years, the use of X-ray crystallography and cryo-reconstruction has led to the identification of the ~3900 kDa large capsid structures of AAV1[64], AAV2[65,66], AAV3b[67], AAV4[68,69], AAV5[70], AAV6[71], AAV7[72], AAV8[73] and AAV9[74,75]. Comparison of the individual structures revealed that the VP core contains an eight-stranded anti-parallel β-barrel motif, βB to βI, as well as an α-helix, αA (Figure 3A). The loop structures connecting the β-strands, named after the flanking β-strands, appear on the surface of the assembled capsid and comprise the variable regions, VRI to VRIX[69].

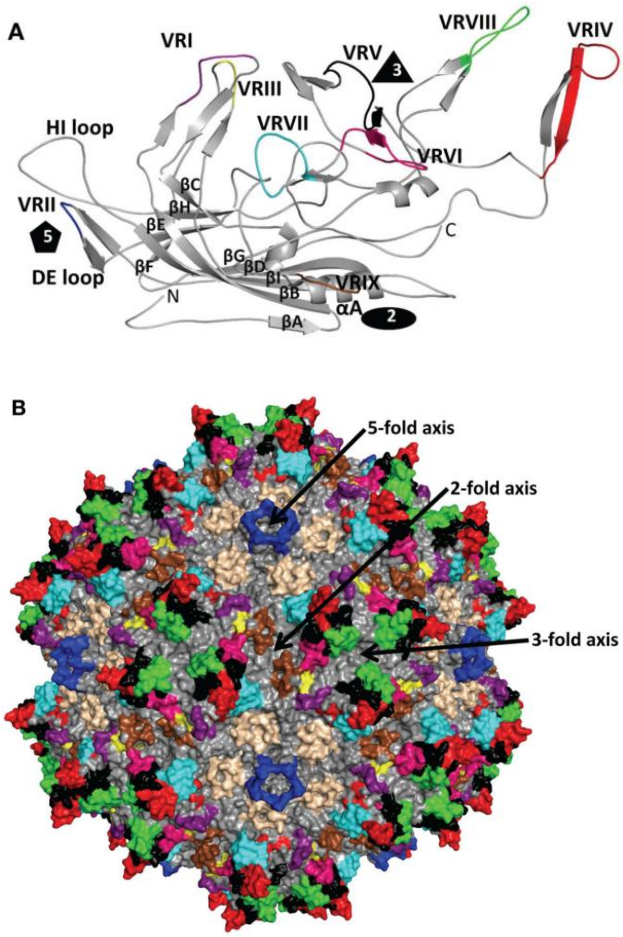

Figure 3: AAV VP3 and capsid structure
(A) VP3 monomer of AAV1 with the variable regions VRI to VRIX as well as the various symmetry axes. The β-barrels (βB to βI), the connecting loops (DE, HI) and the α-helix (αA) form the core of the protein. N, N-terminus; C, C-terminus. **(B)** Depiction of an assembled icosahedral AAV2 particle comprising 60 monomeric VPs. Two-, three- and five-fold axes are indicated. The pore at the five-fold axis is connecting the inside to the outside. The three-fold axis shows the crucial protrusions for receptor binding. Color code as in **(A)**. Taken from Tseng and Mc-Kenna, 2014[76].

Sixty copies of VP proteins finally assemble to form the T = 1 icosahedral capsid via two-, three-, and five-fold symmetry-related interactions (Figure 3B). These interactions form the typical surface area of the particle with cylindrical channels at the five-fold axis surrounded by a depression, protrusions enclosing a depression at the three-fold axis and depressions at the two-fold axis[76]. The DE loop, linking βD and βE, forms the cylindrical channel at the five-fold axis, which is a pore separating the inside from the outside of the capsid. AAV genome-bound Rep proteins can attach to the capsid leading to the encapsidation of the DNA strand[77]. Since VP proteins derive from the same ORF, they share a common 534 aa-long C-terminus. However, start codon usage and alternative splicing lead to shorter N-termini of VP2 and VP3 as compared to VP1, whose additional amino acids encoding a phospholipase A2 (PLA2) domain required for virus infectivity[78] (1.1.3). This unique N-terminus as well as the truncated N-terminus of VP2 are involved in forming globules inside the capsid[79]. Upon conformational change, the VP1 N-terminus is released through the five-fold axis-channels exposing the PLA2 domain[79,80] and additional nuclear localization signals (NLS)[81]. Although VP1 carries important domains for infectivity and intra-cellular trafficking, particle formation is possible with solely VP1[82], VP2[82,83] or VP3[56,59]. To make VP3-only particles,

AAP has to be complemented in *trans* since the VP3-encoding mRNA lacks the start codon and first amino acids of AAP[56].

The capsid surface-exposed variable regions differ predominantly between the serotypes due to the fact that these areas are not involved in the essential core structure of the particle, leaving room for evolutionary adaptation. They play a major role in receptor binding and antibody recognition. The former was first discovered for AAV2, which is naturally able to interact with heparan sulfate proteoglycan (HSPG)[84]. In general, most AAV serotypes interact with glycan structures on the cell surface for primary attachment. HSPG are utilized for AAV2, AAV3b and AAV6; N- or O-linked sialic acid (SA) for AAV1, AAV4, AAV5 and AAV6; and N-linked galactose for AAV9 (Table 2).

Table 2: AAV receptors

Serotype	Glycan receptors	Additional receptors
AAV1	α2,3/ α2,6 N-linked SA[85,86]	AAVR[87],
AAV2	HSPG[84]	AAVR[87], FGFR1[88], HGFR[89], LamR[90], CD9[91], integrin[92,93]
AAV3b	HSPG[94]	AAVR[87], FGFR1[95], HGFR[96], LamR[90]
AAV4	α2,3 O-linked SA[97]	unknown
AAV5	α2,3 N-linked SA[97,98]	AAVR[87], PDGFR[99]
AAV6	α2,3/ α2,6 N-linked SA[85], HSPG[86]	AAVR[87], EGFR[100]
AAV7	unknown	unknown
AAV8	unknown	AAVR[87], LamR[90]
AAV9	N-linked galactose[101,102]	AAVR[87], LamR[90]
AAVrh10	unknown	AAVR[103]
AAVpo1	unknown	unknown
AAV12	unknown	unknown

Abbreviations: AAV = adeno-associated virus, AAVR = AAV receptor, CD9 = tetraspanin, EGFR = epidermal growth factor receptor, FGFR1 = fibroblast growth factor receptor 1, HGFR = hepatocyte growth factor receptor, HSPG = heparan sulfate proteoglycan, LamR = laminin receptor, PDGFR = platelet-derived growth factor receptor, SA = sialic acid.

After attaching to the cell, internalization and trafficking are believed to be mediated by secondary proteinaceous receptors such as, in the case of AAV2, fibroblast growth factor receptor 1 (FGFR1), hepatocyte growth factor receptor (HGFR), laminin receptor (LamR), CD9 tetraspanin and αVβ5/α5β1 integrin. However, knockout studies for FGRF1 and HGFR demonstrated in several cell lines an unconvincing effect of those receptors[87] questioning their crucial role in this multifactorial procedure. In 2016, one noteworthy publication of Pillay *et al.* caught the attention of the AAV field by reporting the discovery of an

essential AAV receptor, consequently called AAVR[87]. In this work, AAVR dependency was shown for AAV1, AAV2, AAV3b, AAV5, AAV6, AAV8 and AAV9. A follow-up study could prove AAVR-mediated internalization of even more AAVs, including AAVrh10, but revealed receptor independence in the case of AAV4 and a closely related AAVrh32.33[103], suggesting an alternative entry route for these viruses. Interestingly, a viral overlay assay performed in another study demonstrated that AAVR and a 150 kDa large glycoprotein, originally discovered over 20 years ago[104], are identical[105].

Especially relevant for gene therapy applications in humans is that the assembled viral AAV particle offers extensive contact areas for neutralizing antibody interactions. As mentioned above, the variable regions are crucial for receptor binding making an antibody-induced impairment at this position particularly disruptive. The antigen-binding fragment (Fab) of the antibodies were shown to cover the protrusions surrounding the three-fold axis in AAV1, AAV2 and AAV6[106] and bind to specific surface epitopes on the capsid[107]. In general, the neutralization can occur prior to or post attachment to cellular receptors, in both cases preventing successful transduction. A major problem for the use of AAV in gene therapy is the high anti-AAV antibody prevalence in humans of 67%, 72%, 40%, 46%, 38% and 47% for AAV1, AAV2, AAV5, AAV6, AAV8 and AAV9, respectively[108]. Hence, the highest antibody abundance in our society is observed for AAV2, as further validated by analyzing 888 human serum samples from donors around the world[109]. Although the serotypes differ in their respective variable regions, cross-reactivity has been documented e.g. between AAV2 and AAV3[110], and even very weakly between AAV1 and AAV5[111] that share a low sequence homology (59%). The complex antibody-capsid interplay is not only biologically interesting but has major implications in clinical trials where appropriate solutions, such as generation of immune-evasive capsids, have to be found and applied.

1.1.3 INFECTION CYCLE

The life cycle of AAV is a multi-step process including virus binding to its receptor, internalization, endosomal trafficking, import to the nucleus, genome replication and gene expression. The different serotypes possess variable interaction partners on the cell surface (Table 2) but the initial membrane attachment is typically facilitated by glycan receptor binding. Subsequent invagination of the cellular membrane forms a vesicle around the receptor-bound AAV, a process whose mechanistic details remain elusive. Dependency

on dynamin- and clathrin-mediated endocytosis has been described for this process[112,113]. However, drug-induced inhibition of clathrin-coated vesicles showed an independency of AAV2 of this particular route, and alternative pathways such as the GPI-anchored-protein-enriched endosomal compartment as well as the clathrin-independent carriers were suggested instead[114]. After internalization, the AAV particle has to traffic to the nucleus, which was demonstrated to be a rate-limiting hurdle for the outcome of the infection[115]. The majority of internalized virions accumulate in the perinuclear region and only a fraction ends up in the nucleus after passing through early, late and recycling endosomes including the crossing of the Golgi complex and the endoplasmic reticulum[113,116,117]. The confinement in the endosomes is eventually circumvented by a pH-induced conformational change in the capsid structure of the AAV particle, leading to the translocation of the internal VP1/VP2 N-terminal region to the capsid surface[79,80]. This region comprises the PLA2 domain that plays a major role in endosome escape and therefore the release of the virus into the cytoplasm[78,118]. Next to the PLA2 domain, the externalized parts of VP1/VP2 additionally contain essential NLS mediating a translocation into the nucleus in a still poorly characterized process[81,119,120]. A recent study showed that AAV2 is transported through the nuclear pore complex (NPC) indicating yet another physiological barrier that the virus has to overcome[121] but at the same time implying intact particle transport across the NPC. Thus, it is assumed that genome uncoating occurs in the nucleus, albeit details of this process remain to be elucidated.

Once uncoated, the faith of the single-stranded AAV genome is dependent on the presence or absence of a helper virus. In the absence of a helper virus, the genome preferentially integrates into the AAVS1 locus on chromosome 19 to establish latency[122,123]. The region in close proximity to the locus contains Rep binding sites (RBS) for the Rep78 and Rep68 proteins which can tether the AAV genome to the chromosome by simultaneously interacting with RBS motifs in the ITRs[124–126]. The mechanism of integration is proposed to be non-homologous end joining (NHEJ) due to weak homology between the AAV genome and the AAVS1 locus[127]. This is further supported by the detection of several Rep-interacting DNA repair proteins by pull-down assays[128]. In the presence of a helper virus, the integrated AAV genome is activated to initiate replication and transcription. In a first step, the single-stranded AAV genome is converted to a double-stranded DNA by utilizing a strand displacement mechanism[129]. Therefore, the partially self-complementary ITR forms a secondary structure with an exposed 3′ hydroxyl group serving as a replication primer. Next, the host replication machinery is facilitating unidirectional synthesis of the complementary strand until reaching the 5′ end of the genome.

Subsequent binding of Rep78 and Rep68 to the Rep binding element (RBE) and RBS within the 3′ ITR leads to a conformational change and cleavage at the terminal resolution site (trs)[130]. This process is induced by the endonuclease, helicase and ATPase enzymatic activities of the Rep proteins and allows the replication of the 3′ ITR. Separation of the newly-generated double-stranded genome yields two DNA strands with a free 3′ hydroxyl group for further iteration. In parallel, the AAV promoters p5 and p19 are activated by the helper virus leading to the expression of the Rep proteins Rep78, Rep68, Rep52 and Rep40 which are fostering replication, expression and, in absence of a helper virus, integration. Expression of the essential structural components VP1, VP2 and VP3 as well as the assembly-activating protein (AAP) is driven by the p40 promoter. With the exception of AAV4 and AAV5, capsid assembly of all studied AAV serotypes is dependent on AAP[60,61]. The detailed mechanism of AAP-assisted particle formation is still unclear; however, a role as a scaffold or chaperone has been suggested[61–63] The encapsidation of one single-stranded genome through the five-fold symmetry pore is facilitated by the binding of the large Rep proteins to the ITRs and the VPs[77,131]. The translocation is assisted by the helicase domains of the smaller Rep proteins Rep52 and Rep40[132]. Infectious particles are then mostly released from the cell upon helper virus-induced cell lysis.

1.1.4 RECOMBINANT AAVS

Arguably one of the biggest advantages of AAV is the easy manipulation of its genome by replacing *rep* and *cap* with foreign DNA, such as a promoter and transgene of choice. Despite the size restrictions of ~4.7 kb, AAV leaves sufficient room for delivering intact gene copies, transcriptional regulators or gene editing tools. The only requirement for the production of such a recombinant AAV (rAAV) are *cis*-acting ITRs flanking the synthetic cargo as well as the supply of *rep* and *cap* in *trans*[133]. This offers the possibility to freely select a genome-capsid combination that is best suited for the individual task. To mimic an adenovirus infection needed for rAAV particle generation, a plasmid containing important adenoviral genes, namely E2A, E4 and VA RNA genes, is mandatory[134]. The fact that viral genes are solely present during virus production dramatically enhances the safety profile of rAAV. In contrast to the integration capability of *rep*/*cap*-bearing AAVs, recombinant vector sequences could not be detected in the AAVS1 locus due to the missing Rep proteins[135]. Instead it was demonstrated that the genome persists preferentially episomally[136], guaranteeing a stable expression over years in mice[137], rats[138], monkeys[138,139] and humans[140]. Although random integration is not fully

abolished[141], no oncogenic effects could be detected in mice for a recombinant AAV2 in an extensive study by Li *et al.*[142].

Genome engineering led to the development of self-complementary AAV vectors (scAAV) with the aim to circumvent the rate-limiting step of the second strand generation[143,144]. A mutation in the trs within the ITR prevents Rep-induced nicking of the DNA, leading to a double-stranded genome that can directly serve as a template for transcription[145,146]. Initial tests *in vitro* revealed dramatic effects of up to 140-fold increased transduction efficiency[147]. Additionally, superior transduction of muscle and liver tissue was shown *in vivo*[145]. The downside of these scAAVs is that the already limited packaging capacity is cut in half to ~2.2 kb, restricting the design of expression cassettes. Solutions to enhance the extent of available genetic information include splitting the *cis*-acting sequence elements in half, to later reunite the two fragments by either homologous recombination or RNA splicing[148–151]. In conclusion, the favorable characteristics of single-stranded and self-complementary rAAV led to the initiation of dozens of clinical trials for recessive monogenic disorders over the past decades, further illustrating the potential impact of basic AAV biology research for future applications[152].

1.2 CAPSID ENGINEERING

Although AAV exhibits many advantageous characteristics for successful use in gene therapy, concerns persist about insufficient tissue specificity and clearance by the host immune system. Due to the simple nature of this virus, the exposed capsid epitopes are directly interacting with cellular receptors and antibodies, promoting the search for beneficial variations in these regions. To address these needs, several approaches exist. For instance, the Wilson group is mining for natural AAV isolates in different species and thus assembling a comprehenisve collection of novel capsids. However, it has been shown that most AAV serotypes preferentially transduce the liver[153], leaving room for improvement by utilizing technologies for the design of synthetic capsids, commonly referred to as capsid engineering.

One possibility is the introduction of mutations into the *cap* gene in a random fashion by error-prone PCR. It has previously been demonstrated that already a single amino acid change can restore defective AAV isolates[154], providing the rationale to screen libraries consisting of AAV mutants. Perabo *et al.* and, in the following year, Maheshri *et al.* made use of an AAV2-based mutant collection and reported evidence for an improved immune

evasion[155,156]. Since crystal structures exist for most of the commonly used AAV serotypes, the approach can be fine-tuned by limiting the random mutagenesis to regions which are important for receptor or antibody binding[157,158]. Furthermore, detailed knowledge about particle structures allows rational mutagenesis to, for instance, mediate immune evasion or improve capsid stability by masking of proteasome-associated tyrosine residues[159–164]. As an example of how beneficial a single point mutation can be, the change of phenylalanine to leucine in AAV6 resulted in a capsid termed AAV6.2 that showed enhanced transduction of murine lung tissue and human airway epithelium[165].

Two studies published in 2015 by the groups of Vandenberghe and Schaffer took a highly innovative approach to capsid engineering by aiming to discover ancestral AAVs through computational analysis. Both groups phylogenetically compared naturally occurring AAV isolates to predict common amino acids of putative ancestors. For positions where no clear prediction could be made, a library comprising the potential residues was generated and subsequently screened in cell lines. The most promising candidates demonstrated increased expression in muscle tissue for AAVC7[166] and enhanced transduction of liver, muscle and retina for AAVAnc80L65[167]. The latter was studied more extensively in follow-up publications illustrating its great potency in the inner ear. AAVAnc80L65 was able to transduce all inner hair cells and the majority of outer hair cells in an adult murine cochlea[168]. Additionally, the ancestral vector showed a superior GFP expression in comparison to AAV1, AAV2, AAV6, AAV8 and AAV9 in organotypic cochlea explants[169] and could rescue mice with Usher syndrome type 1c[170].

1.2.1 DNA FAMILY SHUFFLING

In 1994, DNA family shuffling was introduced for the first time[171] and eventually adapted for the AAV field by Grimm *et al.* in 2008[172] as well as later in the same year by the groups of Samulski[173] and Schaffer[174]. The technique facilitates the directed evolution of novel synthetic AAV capsids in a high-throughput manner by exploiting the high homology of over 50% between the naturally occurring AAV serotypes (Table 1). In a first step, parental capsid genes undergo DNase-mediated fragmentation and subsequent primer-free PCR amplification. The ~300 bp-large pieces of the *cap* genes are capable of priming themselves in the elongation reaction, ultimately leading to the restoration of a chimeric full-length capsid sequence. Due to the shuffling of several parental sequences, the recombination possibilities are virtually

unlimited and and vastly exceed the typical library diversities of up to 10^7 variants[172]. Cloning of the chimeric sequences into an ITR- and *rep*-bearing plasmid allows production of the viral library that can then be utilized to screen for chimeras with enhanced efficiency or specificity *in vivo* or *in vitro*. By systematically rescuing AAV genomes by PCR from the cells or organs of interest, chimeras with increased capability to selectively transduce these targets are favored. Iterative rounds further boost chances to enrich promising variants.

The enormous potential of this approach was demonstrated by the isolation of AAVDJ in 2008, a chimera based on AAV2, AAV8 and AAV9, and by its high efficiency in the liver and additional cell types[172]. Subsequent to this work, several laboratories expanded the knowledge about the variant by testing its application in various tissues as well as by eventually solving its crystal structure[175–178]. Sparked by the success of AAVDJ, numerous groups adopted the technique to select novel variants in various tissues[179–188]. A noteworthy example is AAVM41, a chimera isolated after only two selection rounds from murine skeletal muscle that exhibits pronounced liver-detargeting as well as superior muscle efficiency compared to AAV6[189]. In 2016, Choudhury *et al.* identified a new capsid that efficiently transduces the central nervous system, AAVB1, and that is also more efficient than AAV9 in muscle, pancreas and lung[190]. The Kay group recently published two studies using a xenograft mouse model with implanted human hepatocytes for selection of clinically more relevant AAV chimeras. The rationale for this approach is the poor transduction of human hepatocytes by AAV8, despite its high potency in the murine liver. AAVLK03 was presented in the first publication as a promising variant for selective targeting of human cells and concurrent detargeting from murine hepatocytes[191]. The same vector was then outperformed in the more recent study by their lead candidate AAVNP59, which was 3-fold more efficient in human hepatocytes as compared to AAVLK03[192].

1.2.2 PEPTIDE DISPLAY

Another approach to engineer novel AAV capsids is peptide display, whereby small, mostly 7-9 amino acid-long peptides are integrated into exposed regions of the VP proteins by modification of the *cap* gene. Unlike DNA family shuffling, this technology is not restricted to the domains provided by the naturally occurring serotypes but allows introducing motifs that are entirely new in the context of AAV. This was demonstrated for the first time in 1999 by inserting an integrin-targeting peptide into different putative

loop structures of the AAV2 capsid proteins. One mutant managed to successfully infect AAV2-resistant cell lines, proving the feasibility to retarget the vector[193]. Many follow-up studies were carried out afterwards that similarly attempted to use peptides previously isolated by phage display to increase AAV transduction efficiency in various cell types or tissues (Table 3).

Table 3: AAV variants generated by rational peptide display

Target	Serotype	Insertion[a]	Peptide	Source
integrin	AAV2	587	AGTFALRGDNPQG	193
CD13	AAV2	588	NGRAHA	194
HUVEC	AAV2	587	SIGYPLP	195
integrin	AAV2	588	TGCDCRGDCFC	196
SMC	AAV2	587	EYHHYNK	197
HUVEC	AAV2	587	SMTPFPTSNEANLGGGS	198
Brain	AAV2	587	QPEHSST	199
Lung	AAV2	587	VNTANST	199
MT1-MMP	AAV2	587	CNHRYMQMC	200
Muscle	AAV2	587	TGASSLNIAGLS	201
Astrocytes	AAV9	588	GRGDLGLSA	202

[a]Insertion describes the amino acid position after which the peptide was inserted.
Abbreviations: AAV = adeno-associated virus, CD13 = alanyl aminopeptidase, HUVEC = human umbilical vein endothelial cell, MT1-MMP = membrane type 1 metalloprotease, SMC = vascular smooth muscle cell.

The altered behavior of these viruses is thought to be explained by the disruption of the HSPG motif[203,204]. A peptide insertion into position 587 or 588 of the AAV2 capsid protein is separating the essential arginines 585 and 588, which typically results in HSPG binding-deficient variants. Hence, detargeted vectors are subsequently able to utilize alternative pathways for cellular entry. In most cases, rationally designed variants possess limited chances to excel in specificity or efficiency, since the pre-selected peptides face different steric constraints when incorporated, for the first time, into AAV particles. Similarly, this was even shown for peptides selected in the context of AAV2 when displayed on AAV8 and AAV9[205]. To allow a peptide selection directly in the context of AAV, Perabo *et al.* and Müller *et al.* constructed random peptide display libraries to screen for promising peptide motifs by directed evolution[206,207]. Following the same principle as for DNA family shuffling (1.2.1), iterative selection rounds in the tissue or cells of interest favor candidates with peptide-induced transduction benefits. To monitor the enrichment of certain amino acid configurations, next-generation sequencing of the unselected and final library can nowadays be performed[208]. Since 2003,

numerous groups have adopted this technique and isolated novel, mainly AAV2-based vectors (Table 4).

Table 4: AAV variants generated by random peptide display

Target	Serotype	Insertion[a]	Peptide	Source
Mec1	AAV2	587	AAAGENQARSAA	206
M-07e	AAV2	587	AAARGDAVGVAA	206
HCAEC	AAV2	588	GNDVRAVSA	207
HCAEC	AAV2	588	GNSSRDLGA	207
Calu6	AAV2	588	GVTAGRAPA	209
PC3	AAV2	588	GDLSNLTRA	209
HSaVEC	AAV2	588	GNDVRSANA	209
HSaVEC	AAV2	588	GNDVRAVSA	209
Kasumi-1	AAV2	588	GNQVGSWSA	210
K562	AAV2	588	GEARVRPPA	211
CD34+ PBPC	AAV2	588	GNRTWEQQA	212
Lung	AAVDJ	588	GMVNNFEWA	172
Lung	AAVDJ	588	GNSSRDLGA	172
PymT	AAV2	588	GESGLSQSA	213
PymT	AAV2	588	GDLGSARAA	213
Lung	AAV2	588	GPRSTSDPA	213
PymT	AAV2	588	GRGDLGLSA	213
Heart	AAV2	588	GVNSTRLPA	214
HCAEC	AAV9	589	GSLRSPPSA	215
HCAEC	AAV9	589	GRGDLRVSA	215
Retina	AAV2	587	LALGETTRPA	186
Keratinocytes	AAV2	587	AAAPRGDLAPAA	216
Retina	AAV8	586	unknown	217
Lung	AAV2	588	GESGHGYFA	208
Brain	AAV2	588	GNRGTEWDA	218
Brain	AAV9	588	TLAVPFK	219
Brain	AAV9	588	YTLSQGW	219
Brain	AAV9	588	QAVRTSL	220

[a]Insertion describes the amino acid position after which the peptide was inserted.
Abbreviations: AAV = adeno-associated virus, Calu6 = mouse lung carcinoma cell line, CD34+ PBPC = primary human CD34-positive peripheral blood progenitor cells, HCAEC = human coronary artery endothelial cells, K562 = human myelogenous leukemia cell line, Kasumi-1 = human acute myeloid leukemia cell line, M-07e = human acute megakaryoblastic leukemia cell line, Mec1 = human B-cell chronic lymphocytic leukemia cell line, PC3 = human prostate carcinoma cell line, PymT = polyoma middle T antigen-induced breast cancer cells.

Of note, Grimm *et al.* made use of their newly-discovered AAVDJ as backbone for peptide insertions instead of the less efficient (*in vivo*) AAV2 in order to target the lung[172]. Varadi and colleagues likewise replaced AAV2 by the highly potent AAV9 and succeeded at improving its efficiency in endothelial cells by 40-fold when using a GSLRSPPSA or GRGDLRVSA peptide[215]. In 2016, two publications proved that AAV2 still remains a vital serotype for directed capsid evolution by random peptide display. The most promising peptide in the respective screenings, that has been isolated after five selection rounds, dramatically changed the tropism of the parental virus to the lung[208] and the brain[218], indicating a bigger influence of the peptide itself compared to the serotype. The findings reported in another 2016 study caught particular attention of many in the AAV field. Deverman *et al.* established a novel random peptide screening pipeline by using a Cre recombinase-transgenic mice strain specifically driving transgene expression of loxP site-comprising AAV genomes in astrocytes. The isolated lead candidate was able to robustly transduce the brain of C57BL/6J mice with superior efficiency as compared to the benchmark AAV9[219]. One year later, a slightly modified version of the peptide proved to further enhance the efficiency in the brain[220].

In conclusion, in peptide display, only little changes are made to the *cap* gene in contrast to the broader alterions caused by DNA family shuffling. Nevertheless, the observation that even these subtle modifications can result in significant retargeting also make this technology highly interesting for the development of tailored AAV vectors.

1.3 BARCODED AAVS

Synthetic AAV vectors isolated from capsid selection strategies such as mutagenesis, DNA family shuffling or peptide display exemplify the enormous potency of these techniques to enhance efficiency and specificity. However, even after several selection rounds in the target tissue or cell, hundreds or thousands of interesting candidates often remain and, complicate the final decision for a single variant. Illumina or PacBio[221] next generation sequencing nowadays facilitate this choice by offering the possibility to monitor the enrichment of certain favorable patterns. Based on this information, few lead candidates are usually selected and validated by individually testing them in the target of interest. Ideally, such validation experiments include essential benchmarks from the literature as controls concurrently increasing the required time, costs and amounts of animals.

A solution for these issues was presented by Adachi *et al.* in 2014 who introduced a barcode-based parallel screening system for novel AAV variants[222]. To this end, DNA barcodes were integrated into the AAV genome after the *rep* and *cap* genes enabling a tracking of the cognate capsids *in vivo* by detecting the capsid-assigned barcodes in the tissues. A screening of such barcoded AAV libraries dramatically cuts down animal numbers and downstream processing of massive sample amounts, while permitting a concurrent head-to-head comparison of all candidates in the same organism. By utilizing this approach, the group could identify amino acids in the AAV capsid that are important for receptor binding, tropism and neutralization. Later that year, Marsic and colleagues published a highly similar strategy that deviated in the construct design. Rather than incorporating the barcode into a wild type genome, it was placed into a recombinant AAV comprising a ubiquitously-expressing CBA promoter driving a luciferase and mApple reporter gene[223]. Next to the sequencing-based tracking of the barcode and thus capsid, this allows a simultaneous detection of the bioluminescence and fluorescence as functional readouts. The power of this adapted technique was exemplified in a separate paper in 2015 by the same group[224]. Moreover, the Björklund group presented a high-throughput approach where a collection of random barcode sequences can be linked to a plasmid library consisting of *cis*-regulatory elements (CRE). Ultimately, this generates libraries with several million unique barcodes placed in the 3'UTR of the gene cassette, therefore also permitting tracking of the viral transcripts. By sequencing the initial library, a link between the barcode sequence (unknown until this point) and the CRE can be established, facilitating subsequent identification in the tissues[225].

1.4 AIM OF THE THESIS

The aim of this work was to establish a pipeline for the parallel *in vivo* screening of novel pre-selected capsid variants in a high-throughput manner, by exploiting barcode-based tracking of the individual candidates in mice. Based on the knowledge provided by the literature (chapter 1.3) the barcode was placed into the 3'UTR of a reporter cassette enabling concomitant detection of the capsid on the cDNA and DNA level. Furthermore, the goal was to set up a comprehensive normalization strategy for the next generation sequencing data, to quantitatively and simultaneously characterize capsid behavior in terms of specificity and efficiency. The collection of variants to be screened comprised over 70 novel peptide-modified derivatives of natural AAV isolates that had already been studied extensively in our group *in vitro*. Exposing these capsids to the complex physiological environment of living

animals would ideally identify tissue-tropic or highly active vectors for potential use in clinical applications. Importantly, to improve the stringency of this screening and the results, the most popular benchmarks from the literature were included, such as AAVDJ, AAVLK03, AAVAnc80L65 and many others. This promised the possibilities to not only validate the pipeline by reproducing published results, but to potentially also identify and characterize superior candidates from our own pool.

2 MATERIALS

2.1 LABORATORY EQUIPMENT

Table 5: Laboratory equipment

Name	Vendor
4K15C	Merck KGaA
Accu-jet® pro	BRAND GmbH & Co. KG
Allegra X-12	Beckman Coulter
Aqualine AL 12	LAUDA
Aqualine AL 5	LAUDA
Avanti J-26 XP	Beckman Coulter
Axio Imager.A2	Carl Zeiss AG
Axio Scan.Z1	Carl Zeiss AG
Axiocam 503 color	Carl Zeiss AG
Basic Meter PB-11	Sartorius AG
CanoScan LiDE 70	Canon Inc.
Captair bio	erlab
CE Module	Bio-Rad Laboratories, Inc.
Centrifuge 5415R	Eppendorf AG
Centrifuge 5417R	Eppendorf AG
Centrifuge 5424R	Eppendorf AG
Centrifuge 5430R	Eppendorf AG
Centrifuge 5810R	Eppendorf AG
CKX41SF	Olympus Corporation
Countess	Invitrogen AG
Cryostar™ NX70	Thermo Fisher Scientific
Cytation 5 imaging reader	BioTek Instruments, Inc.
Cytomics FC 500 MPL	Beckman Coulter
E1-ClipTip 12.5, 200, 300, 1250	Thermo Fisher Scientific
E835	Consort
E-H2	Febikon Labortechnik GmbH
epMotion® 96	Eppendorf AG
FlexCycler	Analytik Jena AG
Forma -86 °C ULT Freezer	Thermo Fisher Scientific
Fragment Analyzer™	Advanced Analytical Technologies, Inc.
Function Line	Thermo Fisher Scientific
Galaxy MiniStar	VWR International

Gel Doc XR	Bio-Rad Laboratories, Inc
GenePulser Xcell™	Bio-Rad Laboratories, Inc.
HB-202	Biozym Scientific GmbH
HERAcell 150	Thermo Fisher Scientific
HERAsafe KS12	Thermo Fisher Scientific
HXP 120V	Carl Zeiss AG
KB 650-2NM	KERN & SOHN GmbH
MagMAX™ Express 96	Thermo Fisher Scientific
Mastercycler	Eppendorf AG
Mastercycler gradient	Eppendorf AG
Mastercycler pro S	Eppendorf AG
MF 22	Scotsman
Microlab STAR	Hamilton Robotics GmbH
Microwave oven	Sharp Electronics
Mixing Block MB-102	BIOER Technology
MPS C1000	Labnet International, Inc.
MSH basic yellow line	IKA-WERKE GmbH &Co. KG
Multitron	INFORS-HT
NANODROP 2000	Thermo Fisher Scientific
NextSeq™ 500	Illumina, Inc.
Optima L-90K Ultracentrifuge	Beckman Coulter
Owl EasyCast B1	Thermo Fisher Scientific
P2, P10, P20, P200, P1000	Gilson, Inc.
P93D	Mitsubishi Electric Corporation
PC Module	Bio-Rad Laboratories, Inc.
PCR Plate Spinner	VWR International
peqSTAR 96 Universal	VWR International
Precellys 24-Dual	Bertin Instruments
Premium Freezer -20 °C	Liebherr-International AG
Premium Fridge 4 °C	Liebherr-International AG
QuantStudio™ 6 Flex Real-Time PCR System	Thermo Fisher Scientific
Refractometer Model RMI	Exacta + Optech
RG-6000	Corbett Research
Rotor 70.1TI	Beckman Coulter
Rotor 70TI	Beckman Coulter
Shaker DOS-10L	neoLab Migge GmbH
Sonorex Super RK31	BANDELIN electronic GmbH & Co. KG
Sprout Mini Centrifuge	Heathrow Scientific
StepOnePlus	Applied Biosystems
Synergy™ HT	BioTek Instruments, Inc.

Tube Sealer 342428	Beckman Coulter
TW12 water bath	Julabo GmbH
Ultraviolet Sterilizing PCR Workstation	VWR International
U-RFL-T	Olympus Corporation
UST-30M-8E	Biostep GmbH
UVT-S-AR	Grant Instruments
Vac-Man®	Promega GmbH
Veriti 96 Well Thermal Cycler	Thermo Fisher Scientific
Vortex-Genie 2	Scientific Industries, Inc.

2.2 LABORATORY MATERIAL

Table 6: Laboratory material

Name	Vendor	Catalog#
0.2ml 8-Strip PCR Tube	STARLAB International GmbH	I1402-2900
0.2ml 8-Strip PCR Tube	STARLAB International GmbH	A1402-3700
1.5 ml tube	STARLAB International GmbH	E1415-2230
500ml Centrifuge Tube	Corning, Inc.	431123
5PRIME Phase Lock Gel	Quantabio	2302830
96 Well Cell Culture Plate	Greiner Bio One	655180
Amicon Ultra-15	Merck KGaA	UFC910008
Biosphere® Fil. Tip 10 µl	Sarstedt AG & Co. KG	70.1130.210
Biosphere® Fil. Tip 100 µl	Sarstedt AG & Co. KG	70.760.212
Biosphere® Fil. Tip 1000 µl	Sarstedt AG & Co. KG	70.762.211
Biosphere® Fil. Tip 20 µl	Sarstedt AG & Co. KG	70.760.213
Biosphere® Fil. Tip 200 µl	Sarstedt AG & Co. KG	70.760.211
Cell Culture Flask 550 ml	Greiner Bio One	660175
Cell lifter	Corning, Inc.	CLS3008
CELLSTAR® 15 ml	Greiner Bio One	188271
Centricon Plus-70	Merck KGaA	UFC710008
CK28 2 ml	Bertin Instruments	P000911-LYSK0-A
CK28 7 ml	Bertin Instruments	P000935-LYSK0-A
Combitips advanced® 0.5 ml	Eppendorf AG	0030089421
Combitips advanced® 1.0 ml	Eppendorf AG	0030089430
Combitips advanced® 10 ml	Eppendorf AG	00300089464

Combitips advanced® 2.5 ml	Eppendorf AG	0030089448
Combitips advanced® 5.0 ml	Eppendorf AG	0030089456
Costar Stripette 25 ml	Corning, Inc.	CLS4489
Costar Stripette 50 ml	Corning, Inc.	CLS4490
Countess™ cell counting chamber slides	Thermo Fisher Scientific	C10283
Disposable Scalpel	FEATHER Safety Razor Co., Ltd.	02.001.30.010
Easystrainer 70 µm	Greiner Bio One International GmbH	542070
Electroporation cuvettes	Biozym Biotech Trading GmbH	748050
Falcon™ 50 ml	Corning, Inc.	352070
Inoculation Loop	Greiner Bio One	731170
Luer-Lok™ 3ml Syringe	BD Biosciences	309658
Luer-Lok™ 5ml Syringe	BD Biosciences	309649
Micro tube 0.5ml	Sarstedt AG & Co. KG	72.699
Micro tube 1.5ml	Sarstedt AG & Co. KG	72.690.001
Micro-Fine™ + Demi	BD Biosciences	324826
Microlance 3™	BD Biosciences	301500
Nunclon™ Delta Surface	Thermo Fisher Scientific	168381
Optiseal Polypropylene Centrifuge Tubes	Beckman Coulter	361625
Pasteur pipette	BRAND GmbH & Co. KG	747720
Petri Dish	Greiner Bio One	633180
Pierce Protein Concentrator	Thermo Fisher Scientific	88537
Pipette 10 ml	Greiner Bio One	607180
Pipette 5 ml	Sarstedt AG & Co. KG	86.1253.001
Pipette Tips 10-200 µl	Greiner Bio One	739290
Pipette Tips 200-1000 µl	Greiner Bio One	740290
QiaShredder	Qiagen N.V.	79654
Quali-Pipette tips 10 µl	Kisker Biotech GmbH & Co. KG	GC.TIPS.B
Quick-Seal Centrifuge Tubes	Beckman Coulter	Z51218SCA
Re-Seal™ Polyallomer Centrifuge Tubes	Seton Scientific Corp.	9041
SafeSeal micro tube 2ml	Sarstedt AG & Co. KG	72.695.500
SafeSeal tube 1.5ml	Sarstedt AG & Co. KG	72.706
Slyde-A-Lyzer™ G2 Dialyse Cassettes	Thermo Fisher Scientific	87736

Strip Tubes and Caps, 0.1 ml	Qiagen N.V.	981103
Superfrost Ultra Plus®	Thermo Fisher Scientific	J4800AMNZ
TissueTek® Cryomold®	Sakura Finetek Europe *B.V. KvK*	4557
VacConnectors	Qiagen N.V.	19407

2.3 CHEMICALS

Table 7: Chemicals

Name	Vendor	Catalog#
Acetic acid	Merck KGaA	33209
Adenosine triphosphate (ATP)	Merck KGaA	A2383
Agarose	Biozym Biotech Trading GmbH	840004
Ampicillin	Carl Roth GmbH + Co. KG	K029.2
Aqua ad injectabilia	B. Braun Melsungen AG	-
Bacto™ agar	BD Biosciences	214010
Bacto™ tryptone	BD Biosciences	211705
Bacto™ yeast extract	BD Biosciences	212750
Bovine serum albumin (BSA)	Carl Roth GmbH + Co. KG	8076
Calcium chloride ($CaCl_2$)	Carl Roth GmbH + Co. KG	HN04.3
Cesium chloride ($CsCl_2$)	Carl Roth GmbH + Co. KG	8627.2
DEPC-Treated Water	Thermo Fisher Scientific	AM9916
Dimethyl sulfoxide (DMSO)	Thermo Fisher Scientific	F515
Disodium phosphate (Na_2HPO_4)	AppliChem GmbH	A3567
Dithiothreitol (DTT)	Thermo Fisher Scientific	15508013
Ethanol (EtOH)	VWR International	20821-330
Ethidium bromide	Merck KGaA	E1510
Ethylenediaminetetraacetic acid 0.1 M (EDTA)	Honeywell International Inc.	34550
Glucose	Merck KGaA	1.08342
Glycerol	VWR International	24388.260
HEPES	AppliChem GmbH	A3268
Hydrochloric acid (HCl)	Merck KGaA	35328
Isopropanol	Different manufacturer	-
Magnesium chloride ($MgCl_2$)	AppliChem GmbH	A3618
Magnesium sulfate ($MgSO_4$)	Merck KGaA	1.05886
Monopotassium phosphate (KH_2PO_4)	AppliChem GmbH	A3620

Nuclease-Free Water	Qiagen N.V.	1039498
OptiPrep™	PROGEN Biotechnik GmbH	1114542
Phenol red	Merck KGaA	1.07241
Polyethylene glycol (PEG)	Merck KGaA	81260
Polyethylenimine (PEI)	Polysciences, Inc.	23966-2
Potassium chloride (KCl)	AppliChem GmbH	A3582
Sodium chloride (NaCl)	Merck KGaA	31434
Sodium hydroxide (NaOH)	Merck KGaA	35256
Sucrose	Carl Roth GmbH + Co. KG	4661
TRIS	Carl Roth GmbH + Co. KG	4855.2
TWEEN®20	Merck KGaA	P9416
β-Mercaptoethanol (β-ME)	Merck KGaA	M3148

2.4 BUFFERS AND SOLUTIONS

Table 8: Commercial buffers and solutions

Name	Vendor	Catalog#
1 Kb Plus DNA Ladder	Thermo Fisher Scientific	10787018
Agencourt AMPure XP	Beckman Coulter	A63882
Agencourt RNAClean XP	Beckman Coulter	A63987
Buffer RLT	Qiagen N.V	79216
CD11b MicroBeads	Miltenyi Biotec	130-049-601
CD11c MicroBeads UltraPure	Miltenyi Biotec	130-108-338
CD19 MicroBeads	Miltenyi Biotec	130-052-201
CD3ε MicroBead Kit	Miltenyi Biotec	130-094-973
Chloroform:Isoamyl alcohol	Merck KGaA	25666
CutSmart® Buffer	New England Biolabs	B7204S
DPBS	Thermo Fisher Scientific	14190
DMEM	Thermo Fisher Scientific	61965
dNTP Mix	Thermo Fisher Scientific	R0193
Fetal bovine serum	Merck KGaA	F7524
GAPDH Primer/Probe Mix (60X)	Thermo Fisher Scientific	Mm00186825_cn
Gel loading dye Purple (6X)	New England Biolabs	B7024S
Illumina Resuspension Buffer	Illumina, Inc.	15026770
NEBuffer 2	New England Biolabs	B7002S
Penicillin-Streptomycin	Thermo Fisher Scientific	15140-122

Phenol-chloroform-isoamyl alcohol mixture	Merck KGaA	77617
Phusion HF buffer (5X)	Thermo Fisher Scientific	F518L
POLR2A Primer/Probe Mix (20X)	Thermo Fisher Scientific	Mm00839502_m1
ProLong™ Gold antifade reagent with DAPI	Thermo Fisher Scientific	P36935
Red Blood Cell Lysis Solution	Miltenyi Biotec	130-094-183
RNA*later*	Qiagen N.V	76106
RPMI	Thermo Fisher Scientific	61870
T4 DNA Ligase Reaction buffer	New England Biolabs	B0202S
Tango Buffer (10X)	Thermo Fisher Scientific	BY5
TE Buffer	Thermo Fisher Scientific	12090015
TissueTek® O.C.T Compound	Sakura Finetek Europe *B.V. KvK*	4583
Trypan Blue stain 0.4%	Thermo Fisher Scientific	T10282
Trypsin-EDTA (0.25%)	Thermo Fisher Scientific	25200-056
4% PFA Solution in PBS	Booster Biological Technology	AR1068

Table 9: Self-made buffers and solutions

Name	Composition	
15% Iodixanol	25% (v/v)	OptiPrep™
	75% (v/v)	PBS-MK-NaCl
25% Iodixanol	41.56% (v/v)	OptiPrep™
	58.19% (v/v)	PBS-MK
	0.25% (v/v)	Phenol red stock solution
40% Iodixanol	66.67% (v/v)	OptiPrep™
	33.33% (v/v)	PBS-MK
60% Iodixanol	99.75% (v/v)	OptiPrep™
	0.25% (v/v)	Phenol red stock solution
Benzonase Buffer	150 mM	NaCl
	50 mM	TRIS-HCl (pH 8.5)
	2 mM	$MgCl_2$
Lysogeny broth (LB) agar	1.5% (w/v)	Bacto™ agar
	1% (w/v)	NaCl
	1% (w/v)	Bacto™ tryptone

	0.5% (w/v)	Bacto™ yeast extract
Lysogeny broth (LB) media	1% (w/v)	NaCl
	1% (w/v)	Bacto™ tryptone
	0.5% (w/v)	Bacto™ yeast extract
MACS buffer (pH 7.2)	solvent	PBS (1X)
	0.5% (w/v)	BSA
	2 mM	EDTA
Na-HEPES resuspension buffer	150 mM	NaCl
	50 mM	HEPES
	25 mM	EDTA
PBS (1X)	137 mM	NaCl
	10 mM	Na_2HPO_4
	3 mM	KCl
	2 mM	KH_2PO_4
PBS-MK	solvent	PBS (1X)
	2.5 mM	KCl
	1 mM	$MgCl_2$
PBS-MK-NaCl	solvent	PBS-MK
	1 M	NaCl
PEG-NaCl solution	40% (w/v)	PEG
	1.915 M	NaCl
Phenol red stock solution	0.5%	Phenol red
SOB media	2% (w/v)	Bacto™ tryptone
	0.5% (w/v)	Bacto™ yeast extract
	0.05% (w/v)	NaCl
	10 mM	$MgSO_4$
	10 mM	$MgCl_2$
	2.5 mM	KCl
SOC media	solvent	SOB media
	20 mM	Glucose
Sucrose solution (30%)	solvent	DPBS
	30% (w/v)	Sucrose
TAE Buffer	5.71% (v/v)	Acetic acid
	2 M	TRIS
	50 mM	EDTA
Topping solution	solvent	Na-HEPES resuspension buffer
	3.27 M	CsCl (0.55 g/ml)

2.5 ENZYMES

Table 10: Enzymes

Name	Vendor	Catalog#
Antarctic Phosphatase	New England Biolabs	M0289S
Benzonase	Merck KGaA	1.01695.0001
BsaI-HF	New England Biolabs	R3535S
ClaI	New England Biolabs	R0197S
Esp3I	Thermo Fisher Scientific	ER0451
HindIII-HF	New England Biolabs	R3104S
NotI-HF	New England Biolabs	R3189S
NsiI	New England Biolabs	R0127S
OneTaq® Quick-Load MM	New England Biolabs	M0486S
Phusion Polymerase HF	Thermo Fisher Scientific	F530L
Phusion Polymerase HS	Thermo Fisher Scientific	F549L
PstI-HF	New England Biolabs	R3140S
QuantiFast PCR Master Mix	Qiagen N.V.	1044234
SfiI	New England Biolabs	R0123S
SpeI	New England Biolabs	R0133S
T4 DNA Ligase	New England Biolabs	M202L
XmaI	New England Biolabs	R0180S
XmnI	New England Biolabs	R0194S

2.6 KITS

Table 11: Kits

Name	Vendor	Catalog#
Allprep DNA/RNA 96 Kit	Qiagen N.V	80311
DNA Clean & Concentrator™	Zymo Research	D4013
High-Capacity cDNA Reverse Transcription Kit	Thermo Fisher Scientific	4368813
MinElute PCR Purification Kit	Qiagen N.V.	28006
NextSeq 500/550 High Output Kit v2 (75 cycles)	Illumina, Inc.	FC-404-2005
NucleoBond® Xtra Maxi	Macherey-Nagel GmbH & Co. KG	740414.100
NucleoSpin Plasmid Miniprep Kit	Macherey-Nagel GmbH & Co. KG	740588.250

Ovation Low Complexity Sequencing System	NuGEN Technologies, Inc.	9092-256
Pure Yield Plasmid Midiprep System	Promega GmbH	A2495
PureLink HiPure Plasmid Gigaprep Kit	Thermo Fisher Scientific	K210009
Qiaprep Spin Miniprep Kit	Qiagen N.V.	27106
QIAquick Gel Extraction Kit	Qiagen N.V	28706
QIAquick PCR Purification Kit	Qiagen N.V	28104
Quant-iT™ PicoGreen™ dsDNA Assay Kit	Thermo Fisher Scientific	P7589
RNase-free DNase Set	Qiagen N.V	79254
SensiMix™ II Probe Kit	Bioline	Bio-83020
Standard Sensitivity NGS Fragment Analysis Kit	Advanced Analytical Technologies, Inc.	DNF-473
TOPO™ TA Cloning™ Kit	Thermo Fisher Scientific	450641

2.7 LABORATORY ANIMALS

The inbred strain C57BL/6J (Janvier Labs) was used for *in vivo* experiments. Mice were kept and handled in accordance with the animal proposal G-126/14 and G-89/16.

2.8 BACTERIAL STRAINS

Table 12: Bacterial strains

Name	Vendor	Catalog#
5-alpha Competent *E.coli*	New England Biolabs	C2987H
MAX Efficiency™ DH5α™ Competent Cells	Thermo Fisher Scientific	18258012
MegaX DH10B™ $T1^R$ Electrocomp™ Cells	Thermo Fisher Scientific	C640003
One Shot® *ccd*B Survival™ 2 $T1^R$ Competent Cells	Thermo Fisher Scientific	A10460

2.9 DNA

2.9.1 PEPTIDE OLIGONUCLEOTIDES

Oligonucleotides listed in Table 13 were used for oligonucleotide annealing and subsequent peptide insertion cloning (3.2.2). Lowercase letters indicate overhangs.

Table 13: Peptide oligonucleotides

Name	Sequence (5′ to 3′)
A1_fw	AGGCATGCCATTAGGAGCGGCAGGCGCCCagg
A1_rev	GGGCGCCTGCCGCTCCTAATGGCATGCCTctc
A2_fw	AGGCAACTACTCCAGAGGAGTGGACGCCCagg
A2_rev	GGGCGTCCACTCCTCTGGAGTAGTTGCCTctc
A6_fw	AGGCAACGAGGCGCGGGTCCGGGAGGCCCagg
A6_rev	GGGCCTCCCGGACCCGCGCCTCGTTGCCTctc
BR1_fw	AGGCAATAGGGGGACGGAGTGGGACGCCCagg
BR1_rev	GGGCGTCCCACTCCGTCCCCCTATTGCCTctc
L1_fw	AGGCGAGTCAGGACATGGATATTTTGCCCagg
L1_rev	GGGCAAAATATCCATGTCCTGACTCGCCTctc
P2_fw	TGGCTGCGATTGCCGCGGCGATTGCTTTTGCGCCCagg
P2_rev	GGGCGCAAAAGCAATCGCCGCGGCAATCGCAGCCActc
P4_fw	TGGCAACGATGTGCGCAGCGCGAACGCCCagg
P4_rev	GGGCGTTCGCGCTGCGCACATCGTTGCCActc
P5_fw	TGGCAACGATGTGCGCGCGGTGAGCGCCCagg
P5_rev	GGGCGCTCACCGCGCGCACATCGTTGCCActc

2.9.2 OLIGONUCLEOTIDES

Table 14 shows the oligonucleotides which were used for overlap extension PCR, sequencing, golden gate cloning and regular PCR. Uppercase letters indicate binding regions, lowercase letters visualize overhangs. Restriction digest sites are marked by an underscore.

Table 14: Oligonucleotides

Name	Sequence (5′ to 3′)
#1310_AAV cap 4 correct pos.1630 fw	CTTTGCGGGGCCTAAACAGAACGGCAAC
#1311_AAV cap 4 correct pos.1630 rev	GTTGCCGTTCTGTTTAGGCCCCGCAAAG
#1318_Rep2 rev primer pos. 1827	TCATCCAAATCCACATTGAC
#1319_Rep2 rev primer pos. 944	CGTGGCCCATCCCAGAAAG
#1424_M13 Rev	CAGGAAACAGCTATGAC
#178_DJrev (real)	GTCGCAAAACACTCACGTGACCTC
#36_Pos680For13689	GAAATTGGCATTGCGATTCC
#37_Pos682For45	GATTGGCATTGCGATTCCAC
#412_CMV rv	ccgttaattaaGGCTGGATCGGTCCCGGTGTCTTC
#651_13_eGFPrev	TCCTCCTTGAAGTCGATGC
#652_14_eGFPfw	ATCTTCTTCAAGGACGACG
#653_15-CMVrevCTR	TTGATGTACTGCCAAGTGG
#678_pos. 1990 Amp Rev	GCCTCACTGATTAAGCATTGG
#679_pos. 1205 rep Rev	GCCTATGGAAAAACGCCAGCAAC
#680_E4 Rev pos. 35156 Ad 2	CCTGTTGTAAGACAGGCTTC
#682_E2A-1 Rev pos. 22320	GGCTGCGGAAGTAGGGCGAG
#683_E2A-2 pos. 26966	CAAAGCAGGGGCCAAGAAC
#684_E2A-3 For	CCAACTCCATGCTTAACAGTC
#685_E2A-end For	GAAGATTTGAGGAAGTTGTGG
#686_Rep upstr.	GAGTTTGATTAAGGTACGGTG
#688_E4-3 Rev	CAGTTTGCCTTTTGGAAGCC
#703_E2A-4 For	CACTTAAGCTCGCCTTCGATCTCAG
#704_E2A-5 For	GACAGCCTAACCGCCCCCTTTG
#705_E2A-6 For	CAGTTGGCGATGAGCAGCTG
#706_E2A-7 For	GTGGACGTCGGCTTACCTTC
#707_E4-4 For	GATGATCCATGGTTACGAGTCC
#708_bla For	GGTCTGACGCTCAGTGGAACG
#792_pEGFP_C2_FP	GATCACATGGTCCTGCTG

#822_LSeqFor	GATCTGGTCAATGTGGATTTG
#828_M13Rev	GGAAACAGCTATGACCATG
#833_Rep2for	AGACGCGGAAGCTTCGATCAA
#835_CMV-F	CGCAAATGGGCGGTAGGCGTG
AAV2-MTP_fw	acaggagcttcctccctcaacatcgccggattaagtAGACAAGCAG CTACCGCAGATG
AAV2-MTP_rev	acttaatccggcgatgttgagggaggaagctcctgtGTTGCCTCTCT GGAGGTTGG
AAV9_K1_fw	gccaagcaggcagtttgcgatccccgccatccgcccaggcggccACCG GCTGGGTTCAAAACC
AAV9_K1_rev	ggccgcctgggcggatggcggggatcgcaaactgcctgcttggcCACT CTGGTGGTTTGTGGCC
AAV9_K3_fw	gccaagcaggccgaggtgacctcagggtatccgcccaggcggccACCG GCTGGGTTCAAAACC
AAV9_K3_rev	ggccgcctgggcggataccctgaggtcacctcggcctgcttggcCACTC TGGTGGTTTGTGGCC
AAV9_LD_fw	GAATTTGCTTGGGCTGCAGCTTCTTCTTGG
AAV9_LD_rev	CCAAGAAGAAGCTGCAGCCCAAGCAAATTC
AAV9_PHP. eB _rev	cttaaaaggcaccgccaaagtcccatcACTCTGGTGGTTTGTGG CCAC
AAV9_PHP. S _rev	caaagacgtcctaaccgcctgTTGGGCACTCTGGTGGTTTGT G
AAV9_PHP.A_fw	tatactttgtcgcagggttggGCACAGGCGCAGACCGG
AAV9_PHP.A_rev	ccaaccctgcgacaaagtataTTGGGCACTCTGGTGGTTTGT G
AAV9_PHP.eB_fw	gatgggactttggcggtgccttttaagGCACAGGCGCAGACCG G
AAV9_PHP.S_fw	caggcggttaggacgtctttgGCACAGGCGCAGACCGG
Barcode #2	TGACGTCTCTGCTCNNNNNNNNNNNNNNNNCAG GCGAGACGTGACACTGC
Barcode #2_rv	GCAGTGTCACGTCTCGCCTG
EGFP_fw	cgtat<u>gcggccgc</u>ACCGGTCGCCACCATGG
EGFP_rev	agctgc<u>atcgat</u>TTACTTGTACAGCTCGTCCATGCCG
NGS_Fw4	ATCACTCTCGGCATGGACGAGC
NGS_Rev3	GGCTGGCAACTAGAAGGCACA
qPCR_EGFP_fw	GAGCGCACCATCTTCTTCAAG
qPCR_EGFP_rev	TGTCGCCCTCGAACTTCAC
qPCR_EYFP_fw	GAGCGCACCATCTTCTTCAAG
qPCR_EYFP_rev	TGTCGCCCTCGAACTTCAC
Rep2_front_rev	GGGAGCAAGTAATTGGGGATG
WHc1_NISrepair_fw	GGGACCGTGGCAGTCAATTTCCAGGGC

WHc1_NISrepair_rev	GCCCTGGAAATTGACTGCCACGGTCCC
WHc12_repair_fw	CAAGTACAACCACGCCGACGCCGAGTTCC
WHc12_repair_rev	GGAACTCGGCGTCGGCGTGGTTGTACTTG

2.9.3 PROBES

Table 15: Probes

Name	Sequence (5′ to 3′)
EYFP_Probe	FAM-ACGACGGCAACTACA-NFQ
EGFP_Probe	FAM-ACGACGGCAACTACA-BHQ1

2.9.4 CAPSID HELPER

Table 16: Capsid helper

Plasmid#	Name	Source
#0193	WH-Rep2-CapDJ	Eike Kienle
#0827	WH-Rep2-Cap9_P1	Eike Kienle
#0829	WH-Rep2-Cap9_P3	Eike Kienle
#1539	WH-Rep2-CapLK03	Marc Kay
#1610	WH-Rep2-Cap2NIS	Eike Kienle
#1611	WH-Rep2-Cap3NIS	Eike Kienle
#1612	WH-Rep2-Cap4mutNIS	Eike Kienle
#1613	WH-Rep2-Cap5NIS	Eike Kienle
#1614	WH-Rep2-Cap6NIS	Eike Kienle
#1615	WH-Rep2-Cappo1NIS	Eike Kienle
#1729	WH-Rep2-Cap1wt	Eike Kienle
#1730	WH-Rep2-Cap1_P2	This thesis
#1731	WH-Rep2-Cap1_P4	This thesis
#1732	WH-Rep2-Cap1_P5	This thesis
#1733	WH-Rep2-Cap1_A1	This thesis
#1734	WH-Rep2-Cap1_A2	This thesis
#1735	WH-Rep2-Cap1_A6	This thesis
#1736	WH-Rep2-Cap2wt	Eike Kienle
#1737	WH-Rep2-Cap2_P2	Eike Kienle
#1738	WH-Rep2-Cap2_P4	Eike Kienle
#1739	WH-Rep2-Cap2_P5	This thesis
#1740	WH-Rep2-Cap2_A1	This thesis
#1741	WH-Rep2-Cap2_A2	This thesis

#1742	WH-Rep2-Cap2_A6	This thesis
#1743	WH-Rep2-Cap3bwt	Eike Kienle
#1744	WH-Rep2-Cap3b_P2	Eike Kienle
#1745	WH-Rep2-Cap3b_P4	Eike Kienle
#1746	WH-Rep2-Cap3b_P5	This thesis
#1747	WH-Rep2-Cap3b_A1	This thesis
#1748	WH-Rep2-Cap3b_A2	This thesis
#1749	WH-Rep2-Cap3b_A6	This thesis
#1750	WH-Rep2-Cap4wt	Stefanie Große
#1751	WH-Rep2-Cap4_P2	This thesis
#1752	WH-Rep2-Cap4_P4	This thesis
#1753	WH-Rep2-Cap4_P5	This thesis
#1754	WH-Rep2-Cap4_A1	This thesis
#1755	WH-Rep2-Cap4_A2	This thesis
#1756	WH-Rep2-Cap4_A6	This thesis
#1757	WH-Rep2-Cap5wt	Eike Kienle
#1758	WH-Rep2-Cap5_P2	Eike Kienle
#1759	WH-Rep2-Cap5_P4	Eike Kienle
#1760	WH-Rep2-Cap5_P5	This thesis
#1761	WH-Rep2-Cap5_A1	This thesis
#1762	WH-Rep2-Cap5_A2	This thesis
#1763	WH-Rep2-Cap5_A6	This thesis
#1764	WH-Rep2-Cap6wt	Eike Kienle
#1765	WH-Rep2-Cap6_P2	Eike Kienle
#1766	WH-Rep2-Cap6_P4	Eike Kienle
#1767	WH-Rep2-Cap6_P5	This thesis
#1768	WH-Rep2-Cap6_A1	This thesis
#1769	WH-Rep2-Cap6_A2	This thesis
#1770	WH-Rep2-Cap6_A6	This thesis
#1771	WH-Rep2-Cap7wt	Eike Kienle
#1772	WH-Rep2-Cap7_P2	Eike Kienle
#1773	WH-Rep2-Cap7_P4	Eike Kienle
#1774	WH-Rep2-Cap7_P5	Eike Kienle
#1775	WH-Rep2-Cap7_A1	Eike Kienle
#1776	WH-Rep2-Cap7_A2	Eike Kienle
#1777	WH-Rep2-Cap7_A6	Eike Kienle
#1778	WH-Rep2-Cap8wt	Eike Kienle
#1779	WH-Rep2-Cap8_P2	Eike Kienle
#1780	WH-Rep2-Cap8_P4	Eike Kienle
#1781	WH-Rep2-Cap8_P5	Eike Kienle
#1782	WH-Rep2-Cap8_A1	Eike Kienle

#1783	WH-Rep2-Cap8_A2	Eike Kienle
#1784	WH-Rep2-Cap8_A6	Eike Kienle
#1785	WH-Rep2-Cap9wt	Eike Kienle
#1786	WH-Rep2-Cap9_P2	Eike Kienle
#1787	WH-Rep2-Cap9_P4	Eike Kienle
#1788	WH-Rep2-Cap9_P5	Eike Kienle
#1789	WH-Rep2-Cap9_A1	Eike Kienle
#1790	WH-Rep2-Cap9_A2	Eike Kienle
#1791	WH-Rep2-Cap9_A6	Eike Kienle
#1792	WH-Rep2-Caprh10wt	Eike Kienle
#1793	WH-Rep2-Caprh10_P2	Eike Kienle
#1794	WH-Rep2-Caprh10_P4	Eike Kienle
#1795	WH-Rep2-Caprh10_P5	Eike Kienle
#1796	WH-Rep2-Caprh10_A1	Eike Kienle
#1797	WH-Rep2-Caprh10_A2	Eike Kienle
#1798	WH-Rep2-Caprh10_A6	Eike Kienle
#1799	WH-Rep2-Cappo1wt	Eike Kienle
#1800	WH-Rep2-Cappo1_P2	Eike Kienle
#1801	WH-Rep2-Cappo1_P4	Eike Kienle
#1802	WH-Rep2-Cappo1_P5	This thesis
#1803	WH-Rep2-Cappo1_A1	This thesis
#1804	WH-Rep2-Cappo1_A2	This thesis
#1805	WH-Rep2-Cappo1_A6	This thesis
#1806	WH-Rep2-Cap12wt	Stefanie Große
#1807	WH-Rep2-Cap12_P2	This thesis
#1808	WH-Rep2-Cap12_P4	This thesis
#1809	WH-Rep2-Cap12_P5	This thesis
#1810	WH-Rep2-Cap12_A1	This thesis
#1811	WH-Rep2-Cap12_A2	This thesis
#1812	WH-Rep2-Cap12_A6	This thesis
#1813	WH-Rep2-Cap4mutwt	Eike Kienle
#1814	WH-Rep2-Cap4mut_P2	Eike Kienle
#1815	WH-Rep2-Cap4mut_P4	Eike Kienle
#1816	WH-Rep2-Cap4mut_P5	This thesis
#1817	WH-Rep2-Cap4mut_A1	This thesis
#1818	WH-Rep2-Cap4mut_A2	This thesis
#1819	WH-Rep2-Cap4mut_A6	This thesis
#1820	WH-Rep2-Cap1NIS	This thesis
#1821	WH-Rep2-Cap4NIS	This thesis
#1822	WH-Rep2-Cap12NIS	This thesis
#1925	WH-Rep2-Cap9LD	This thesis

#1926	WH-Rep2-Cap9K449R_PHP.eB	This thesis
#1927	WH-Rep2-Cap9K449R_PHP.S	This thesis
#1928	WH-Rep2-Cap2_MTP	This thesis
#1929	pGG-Rep2-B1	Addgene
#1930	pAnc80L65	Addgene
#1933	pAAV-Rep2-cap2_7m8	Boehringer Ingelheim
#1934	pAAV-Rep2-cap2_BR1	Boehringer Ingelheim
#1935	pAAV-Rep2-cap2_L1	Boehringer Ingelheim
#1936	pAAV-Rep2-Cap6ShH10	Boehringer Ingelheim
#1937	pAAV_AAV6.2	Boehringer Ingelheim
#1938	pAAV-Rep2-Cap9BI	Boehringer Ingelheim
#1939	pAAV-Rep2-cap2HBKO	Boehringer Ingelheim
#1940	pAAV-Rep2-Cap9K449R_PHP.B	Boehringer Ingelheim
#2004	WH-Rep2-CapAH3-5	Anne-Kathrin Herrmann
#2047	WH-Rep2-Cap9_K1	This thesis
#2048	WH-Rep2-Cap9_K3	This thesis
#2049	WH-Rep2-Cap9K449R_PHP.A	This thesis
#2050	WH-Rep2-Cap9LD_P1	This thesis
#2051	WH-Rep2-CapM41	This thesis
#2052	WH-Rep2-Cap9_BR1	This thesis
#2053	WH-Rep2-Cap4_L1	This thesis

2.9.5 BARCODED REPORTER PLASMIDS

All plasmids listed in Table 17 are based on Plasmid#552 and were cloned during this thesis as described in 3.2.3.

Table 17: Barcoded reporter plasmids

Plasmid#	Name	Barcode sequence
#2056	pJW1-CMV-EYFP-BC#A2	AGACTCGTTGTATAT
#2057	pJW2-CMV-EYFP-BC#A3	TAGAGATTTAAACCG
#2058	pJW3-CMV-EYFP-BC#A4	CGTGACAGCGGATGG
#2059	pJW4-CMV-EYFP-BC#A5	TGGGCGGTCAGGGTC
#2060	pJW5-CMV-EYFP-BC#A6	TTGCCGTCCTTCGAG
#2061	pJW6-CMV-EYFP-BC#A8	TTCAGCGGACGGGCC
#2062	pJW7-CMV-EYFP-BC#A9	GTCAGTCCGCTCTTT
#2063	pJW8-CMV-EYFP-BC#A11	TTAAGATCCTGGTCG
#2064	pJW9-CMV-EYFP-BC#A13	TCAACATGGGCAACG
#2065	pJW10-CMV-EYFP-BC#A14	CTTGATCGACGCCCA

#2066	pJW11-CMV-EYFP-BC#A15	TACGCTATTCAATCT
#2067	pJW12-CMV-EYFP-BC#A18	GTGCTTCTGGCGGAT
#2068	pJW13-CMV-EYFP-BC#A21	CGGCTGTCGGTCGCC
#2069	pJW14-CMV-EYFP-BC#A22	ATCGTACGTTACTGA
#2070	pJW15-CMV-EYFP-BC#A23	GATTCGAAAGCATAG
#2071	pJW16-CMV-EYFP-BC#A24	CGTATCGGGTCCGGA
#2072	pJW17-CMV-EYFP-BC#A25	TGGTTGGGTTTGTGG
#2073	pJW18-CMV-EYFP-BC#A26	TCGTTGTAACGGTAC
#2074	pJW19-CMV-EYFP-BC#A29	TAACGTTGGGTTGCC
#2075	pJW20-CMV-EYFP-BC#A30	GACCACTAGAAGGGC
#2076	pJW21-CMV-EYFP-BC#A32	CTGCATGGCGGAGTT
#2077	pJW22-CMV-EYFP-BC#A33	TCAACGATTGTCTGG
#2078	pJW23-CMV-EYFP-BC#A34	TGGTAGGTTCGAAAT
#2079	pJW24-CMV-EYFP-BC#A35	ACGTCGCACCGTTTG
#2080	pJW25-CMV-EYFP-BC#A37	CAGGCTTAACGCGGG
#2081	pJW26-CMV-EYFP-BC#A38	ACCATAGCGCCACGA
#2082	pJW27-CMV-EYFP-BC#A39	GTCCCGACTAGGACT
#2083	pJW28-CMV-EYFP-BC#A40	GTCTTGATTGCTTCG
#2084	pJW29-CMV-EYFP-BC#A41	ATTTGGCACAGGATG
#2085	pJW30-CMV-EYFP-BC#A42	GGCCACCGTGTGTGA
#2086	pJW31-CMV-EYFP-BC#A43	ATGAGCAGCGAATGA
#2087	pJW32-CMV-EYFP-BC#A44	ATGTTTAACGGCATA
#2088	pJW33-CMV-EYFP-BC#A45	TTGGACTCACAGATG
#2089	pJW34-CMV-EYFP-BC#A47	AAGGTGACCTAGTGT
#2090	pJW35-CMV-EYFP-BC#A48	CCCTCATGAGGTCCG
#2091	pJW36-CMV-EYFP-BC#A49	ATGACAATGTGCAGG
#2092	pJW37-CMV-EYFP-BC#A50	GCGAGGTCGTTAGTT
#2093	pJW38-CMV-EYFP-BC#A51	TAAGACTGTTCCGGG
#2094	pJW39-CMV-EYFP-BC#A52	GTTTGTAATCTCTAC
#2095	pJW40-CMV-EYFP-BC#A53	GTTAACGCGGCCATT
#2096	pJW41-CMV-EYFP-BC#A55	AGCGGCGTTTATCGT
#2097	pJW42-CMV-EYFP-BC#A56	TTGGTATGTGTCAAT
#2098	pJW43-CMV-EYFP-BC#A58	GTCGACTTCATGGCA
#2099	pJW44-CMV-EYFP-BC#A61	GAGCGTAATTGTGAG
#2100	pJW45-CMV-EYFP-BC#A62	CGTTAACCCGAAAGC
#2101	pJW46-CMV-EYFP-BC#A63	GTGACATGCAGGTAG
#2102	pJW47-CMV-EYFP-BC#A65	ACGATCGTACGTCTT
#2103	pJW48-CMV-EYFP-BC#A67	GTTCAGGTCAGGTCT
#2104	pJW49-CMV-EYFP-BC#A68	TAAGGAGGGCTGTAG
#2105	pJW50-CMV-EYFP-BC#A69	TATCAAGCTAACGTT
#2106	pJW51-CMV-EYFP-BC#A70	GCTCTGGATGTAGTA

#2107	pJW52-CMV-EYFP-BC#A71	TAGATGTGGCGGACA
#2108	pJW53-CMV-EYFP-BC#A74	GTCAACATCGTTACA
#2109	pJW54-CMV-EYFP-BC#A75	GGGCCCTAGCGCGTG
#2110	pJW55-CMV-EYFP-BC#A76	GATAGGCTGGTCCAA
#2111	pJW56-CMV-EYFP-BC#A77	TATTTGTGTCGTTCC
#2112	pJW57-CMV-EYFP-BC#A79	AGTTAGGGCGCTGCG
#2113	pJW58-CMV-EYFP-BC#A80	GCGGAACATAGGCGG
#2114	pJW59-CMV-EYFP-BC#A81	GCCCTTCAGTCAGCT
#2115	pJW60-CMV-EYFP-BC#A82	CGGTCGCGTGACGTG
#2116	pJW61-CMV-EYFP-BC#A83	GCCGGAGTCCCGGTA
#2117	pJW62-CMV-EYFP-BC#A84	CGAGTCGTATGTGGC
#2118	pJW63-CMV-EYFP-BC#A85	AGTAATTGGTCTTGG
#2119	pJW64-CMV-EYFP-BC#A86	GGTCTTTGCTCGGTG
#2120	pJW65-CMV-EYFP-BC#A87	GACTTGGTTGTGACG
#2121	pJW66-CMV-EYFP-BC#A90	TTGTTGTATGAGCAG
#2122	pJW67-CMV-EYFP-BC#A91	TCCACGGAGGCTGCG
#2123	pJW68-CMV-EYFP-BC#A94	CTACCTATTTACTCT
#2124	pJW69-CMV-EYFP-BC#A97	ACCGGGCGTTGAGGC
#2125	pJW70-CMV-EYFP-BC#A99	ACTGTGATGGGTTAG
#2126	pJW71-CMV-EYFP-BC#A100	TGGTTTACAAATTAT
#2127	pJW72-CMV-EYFP-BC#A101	TGTCCGGAAAGGACA
#2128	pJW73-CMV-EYFP-BC#A102	GTTGTGCCCTGAGTG
#2129	pJW74-CMV-EYFP-BC#A104	ACCGTATCTCTCCGG
#2130	pJW75-CMV-EYFP-BC#A107	TTGGAACGTGGGCTT
#2131	pJW76-CMV-EYFP-BC#A109	AGATTCAAAGCTGCG
#2132	pJW77-CMV-EYFP-BC#A110	TGTTGGAAGGTATCA
#2133	pJW78-CMV-EYFP-BC#A111	GTAGCTGAGGTTGGT
#2134	pJW79-CMV-EYFP-BC#A114	AGCCTAATCTTTGAC
#2135	pJW80-CMV-EYFP-BC#A115	AAGCACTAAAGAACA
#2136	pJW81-CMV-EYFP-BC#A116	GGTATGGCCTGCCGC
#2137	pJW82-CMV-EYFP-BC#A117	TGTTTAGGTGAGCCT
#2138	pJW83-CMV-EYFP-BC#A118	TGTGGTGTGACTCAG
#2139	pJW84-CMV-EYFP-BC#A119	TCGGGTTGGTCTTTG
#2140	pJW85-CMV-EYFP-BC#A120	ACATTGTGGTCATAG
#2141	pJW86-CMV-EYFP-BC#A121	AGACTTGGCGTTATG
#2142	pJW87-CMV-EYFP-BC#A122	ACGTGTCGTAGTAAG
#2143	pJW88-CMV-EYFP-BC#A124	TATATTGAGGCGTGT
#2144	pJW89-CMV-EYFP-BC#A126	TGAGAGTCATCCAAG
#2145	pJW90-CMV-EYFP-BC#A127	CCTAATCTCAGGCGG
#2146	pJW91-CMV-EYFP-BC#A129	CGTGACCCAGGAAGT
#2147	pJW92-CMV-EYFP-BC#A132	TCGTTAGTAGCGATC

#2148	pJW93-CMV-EYFP-BC#A135	GAGGTCCAGAGGAAG
#2149	pJW94-CMV-EYFP-BC#A138	ATGATCAGCGATATC
#2150	pJW95-CMV-EYFP-BC#A139	GGTGCCGGACAGCTC
#2151	pJW96-CMV-EYFP-BC#A141	TATAACTTAGCTGAT
#2152	pJW97-CMV-EYFP-BC#A142	CTTCTTCAGGCAACC
#2153	pJW98-CMV-EYFP-BC#A144	CCACTAGGATCCGGA
#2154	pJW99-CMV-EYFP-BC#A145	CAAGGCTTTCTGATC
#2155	pJW100-CMV-EYFP-BC#A146	ATCTCGAAGCGCGTA
#2156	pJW101-CMV-EYFP-BC#A147	GCAATTATCATAGTC
#2157	pJW102-CMV-EYFP-BC#A149	GACCTGCGCCTTACA
#2158	pJW103-CMV-EYFP-BC#A150	CGTCCGTCTAATGAA
#2159	pJW104-CMV-EYFP-BC#A151	GGTTGACAGTGGGCT
#2160	pJW105-CMV-EYFP-BC#A152	AGTTTAGGACAGGCA
#2161	pJW106-CMV-EYFP-BC#A155	TTCATCGGCCGCTAA
#2162	pJW107-CMV-EYFP-BC#A157	TACGTATCGCGTGAT
#2163	pJW108-CMV-EYFP-BC#A158	CTAGGCAGGACACCG
#2164	pJW109-CMV-EYFP-BC#A160	TTGGCAGAGGATCAC
#2165	pJW110-CMV-EYFP-BC#A161	TCGGCTCTGTTCTAG
#2166	pJW111-CMV-EYFP-BC#A162	TTTAGGCGCGGCTTG
#2167	pJW112-CMV-EYFP-BC#A163	CGTCCTGTAAGGAGT
#2168	pJW113-CMV-EYFP-BC#A164	TAGAGTATGAGTGGT
#2169	pJW114-CMV-EYFP-BC#A166	GAGCGGGCAGACGAT
#2170	pJW115-CMV-EYFP-BC#A169	GTGCGCAGGTTAGTG
#2171	pJW116-CMV-EYFP-BC#A171	CTCGCGGCCTGAGGG
#2172	pJW117-CMV-EYFP-BC#A172	CTAGATAAATGCGGT
#2173	pJW118-CMV-EYFP-BC#A173	ACCTGAGTTTGGTGG
#2174	pJW119-CMV-EYFP-BC#A175	CCGTCGAAGAAGGGA
#2175	pJW120-CMV-EYFP-BC#A179	GGCAGCGGACACGTG
#2176	pJW121-CMV-EYFP-BC#A180	ATCCTCTCCGCTACC
#2177	pJW122-CMV-EYFP-BC#A181	TAGCACCATTTACGG
#2178	pJW123-CMV-EYFP-BC#A184	CATGCCATGTGTATC
#2179	pJW124-CMV-EYFP-BC#A187	ACCAACCGGTGTGGG
#2180	pJW125-CMV-EYFP-BC#A189	GGTACAGGACGCAGG
#2181	pJW126-CMV-EYFP-BC#A190	GACCACTTATCGCCA
#2182	pJW127-CMV-EYFP-BC#A195	TCGGCGTGGCGGTCG
#2183	pJW128-CMV-EYFP-BC#A197	GACTTTGACATGTCA
#2184	pJW129-CMV-EYFP-BC#A198	TACATTTAACTGAAG
#2185	pJW130-CMV-EYFP-BC#A199	GGTCAGGACCATTGG
#2186	pJW131-CMV-EYFP-BC#A201	TGGGTTTCGGCATCA
#2187	pJW132-CMV-EYFP-BC#A202	TTACCTTCTAAGGGC
#2188	pJW133-CMV-EYFP-BC#A203	TGGTCGGCGAGTTTG

#2189	pJW134-CMV-EYFP-BC#A205	GGTTGGTTAGGCTGT
#2190	pJW135-CMV-EYFP-BC#A207	ACCGGCAATCCTAGC
#2191	pJW136-CMV-EYFP-BC#A208	GTGTGTTACCTAACA
#2192	pJW137-CMV-EYFP-BC#A209	TCATCTAGCATCGGG
#2193	pJW138-CMV-EYFP-BC#A210	GCCACAGGCATCGTG
#2194	pJW139-CMV-EYFP-BC#A211	CTTATGTGAAGAGAT
#2195	pJW140-CMV-EYFP-BC#A212	TAGTTTATCGCAGGG
#2196	pJW141-CMV-EYFP-BC#A213	GTACCTATCCGTTGT
#2197	pJW142-CMV-EYFP-BC#A214	TTCCGTGTGTTGTCT
#2198	pJW143-CMV-EYFP-BC#A215	CCCGTATGTCGGGTA
#2199	pJW144-CMV-EYFP-BC#A216	GAATCCATGACTTTG
#2200	pJW145-CMV-EYFP-BC#A217	GTTCGTTGCGGGATC
#2201	pJW146-CMV-EYFP-BC#A220	GTGCTTGTCATGCCG
#2202	pJW147-CMV-EYFP-BC#A221	AGTTCACGACTGCGA
#2203	pJW148-CMV-EYFP-BC#A222	GGACTCAGGCCTGGT
#2204	pJW149-CMV-EYFP-BC#A223	TTTGGTTGGAGTCTT
#2205	pJW150-CMV-EYFP-BC#A225	TTACGATTTATGCGC
#2206	pJW151-CMV-EYFP-BC#A226	CAATCCGGCGCGGGT
#2207	pJW152-CMV-EYFP-BC#A228	GTGTAGGTTATCATC
#2208	pJW153-CMV-EYFP-BC#A229	TCGCACGCTGATGTG
#2209	pJW154-CMV-EYFP-BC#A230	AGTTTCACATGACGG
#2210	pJW155-CMV-EYFP-BC#A232	GTTTACGGATCTCGG
#2211	pJW156-CMV-EYFP-BC#A233	TATATAGTCGGTTTG
#2212	pJW157-CMV-EYFP-BC#A236	ATGTCGAACCCAATC
#2213	pJW158-CMV-EYFP-BC#A237	TCTGTATGGGCCAGC
#2214	pJW159-CMV-EYFP-BC#A240	TGATCTGACCGTGTG

2.9.6 REPORTER PLASMIDS

Table 18: Reporter plasmids

Plasmid#	Name	Source
#552	pscAAV-CMV-EYFP-BGHpolyA	Eike Kienle
#2054	pscAAV-CMV-EGFP-BGHpolyA	This thesis
#2055	pscAAV-CMV-EYFP-ccdB-BGHpolyA	Florian Schmidt

2.10 SOFTWARE

Table 19: Software

Name	Vendor
ApE (A Plasmid Editor)	http://biologylabs.utah.edu/jorgensen/wayned/ape/
Gen5 2.09	BioTek Instruments, Inc.
GraphPad Prism 7	GraphPad Software
MendeleyDesktop	Mendeley Ltd.
NanoDrop 2000 v1.5	Thermo Fisher Scientific
NextSeq Control Software	Illumina, Inc.
Office 2007	Microsoft Corporation
PROSize Data Analysis Software	Advanced Analytical Technologies, Inc.
Python 2.7	Python Software Foundation
Quantity One 4.6.9	Bio-Rad Laboratories, Inc.
QuantStudio™ Software V1.3	Thermo Fisher Scientific
Rotor-Gene Q Series Software	Qiagen N.V

3 METHODS

3.1 GENERAL CLONING TECHNIQUES

3.1.1 POLYMERASE CHAIN REACTION (PCR)

In order to amplify DNA fragments for subsequent cloning steps, a PCR was performed with 10-100 ng template, 10 µl Phusion HF buffer, 1 µl dNTPs (10 mM), 1.5 µl forward primer (10 µM), 1.5 µl reverse primer (10 µM), 1.5 µl DMSO, 0.5 µl Phusion Polymerase HS and filled up to 50 µl with nuclease-free H_2O. After mixing, PCR was run under cycling conditions listed in Table 20. Annealing temperature (step 3) and extension time (step 4) varied depending on the utilized primer combination and amplicon length respectively.

Table 20: PCR cycling conditions

Step	Temperature [°C]	Time	Repetitions
1	98	30 sec	
2	98	10 sec	35 cycles
3	56-70	20 sec	
4	72	15 sec/kb	
5	72	5 min	
6	4	hold	

3.1.2 GEL ELECTROPHORESIS

For a size-depended separation of a DNA sample, a 1% agarose gel electrophoresis was carried out by mixing a solution of 1% agarose (w/v) in TAE Buffer with 0.5 µg/ml of ethidium bromide. After solidification of the gel, Gel loading dye Purple (6X) was added to the DNA sample and mixture was loaded into gel pockets. Electrophoresis was performed at 120 V for ~20-30 min. DNA bands were visualized with UV light and optionally excised for DNA purification.

3.1.3 RESTRICTION DIGEST

To enzymatically cleave DNA fragments, restriction endonucleases were used (Table 10) by incubating 3 µg of plasmid DNA or 50 µl of purified PCR product with 1 µl of the respective restriction enzyme in its corresponding buffer for 3-4 h at 37 °C. Outcome was checked by gel electrophoresis as described in 3.1.2.

3.1.4 DNA PURIFICATION

Excised DNA bands from 3.1.2 were purified with QIAquick Gel Extraction Kit according to manufacturer's instructions and eluted in 50 µl of nuclease-free H_2O.

PCR amplicon purifications without gel separation were performed with MinElute PCR Purification Kit, DNA Clean & Concentrator™ or QIAquick PCR Purification Kit following manufacturer's instructions.

3.1.5 LIGATION

For the ligation of a PCR-amplified insert into a restriction enzyme-digested plasmid backbone, 5 molar parts of the insert were combined with 1 molar part of the backbone (60 ng), 1 µl T4 DNA Ligase Reaction buffer, 0.5 µl T4 DNA Ligase and filled up to 10 µl with nuclease-free H_2O. The mix was incubated for 30 min at RT and afterwards directly used for bacteria transformation.

3.1.6 TRANSFORMATION

3 µl of the ligation from 3.1.5 was mixed with 50 µl of MAX Efficiency™ DH5α™ Competent Cells, One Shot® *ccd*B Survival™ 2 T1^R Competent Cells or 5-alpha Competent *E.coli* and incubated for 30 min on ice. Subsequent heat-shock for 45 sec at 42 °C allowed plasmid uptake. Bacterial cells were placed on ice for 2 min before plating on LB agar dishes and incubating overnight at 37 °C.

3.1.7 ELECTROPORATION

30 µl of MegaX DH10B™ T1R Electrocomp™ Cells were mixed with 1.5 µl of ligation mixture from 3.2.3 or 3.1.5 and transferred to pre-cooled cuvettes. Electroporation was performed with GenePulser Xcell™ with the settings 1800 V, 25 µF and 200 Ω. Immediately after completion, 1 ml of pre-warmed SOC medium was added to the cuvette for recovery. The solution was transferred to a tube and incubated at 700 rpm for 1 h at 37 °C. 100 µl of undiluted, 1:10 and 1:100 dilution was streaked on LB agar dishes with appropriate antibiotic resistance and plates were incubated overnight at 37 °C.

3.1.8 PLASMID DNA PREPARATION

For the preparation of plasmid DNA, 3-2000 ml of LB media, depending on the respective kit, with the adequate antibiotic resistance (50 µg/ml) was inoculated with a single colony or glycerol stock and incubated at 180 rpm overnight at 37 °C. Extraction of the plasmid was performed with commercial kits Qiaprep Spin Miniprep Kit, NucleoSpin Plasmid Miniprep Kit, Pure Yield Plasmid Midiprep, NucleoBond® Xtra Maxi or PureLink HiPure Plasmid Gigaprep Kit according to manufacturer's instructions.

3.2 SPECIFIC CLONING PROCEDURES

This chapter describes the cloning of the capsid helper plasmids (Table 16) and the barcoded reporter constructs (Table 17).

3.2.1 OVERLAP EXTENSION PCR

Overlap extension PCR was used to create capsid helper plasmids for the published AAV variants AAV2_MTP, AAV9_K1, AAV9_K3, AAV9LD, AAV9K449R_PHP.eB, AAV9K449R_PHP.S and AAV9K449R_PHP.A. Additionally, mutations in the WH-Rep2-CapNIS plasmids of AAV1, AAV4 and AAV12 were corrected with the same technique. The PCR reaction was set up as shown in 3.1.1 using the corresponding forward primer in combination with primer #178 and the reverse primer with #833. Upon completion of the cycler program (Table 20), two capsid fragments were generated for each AAV variant as seen by gel electrophoresis (3.1.2). Both bands were extracted and

purified for the following second PCR which uses the self-priming fragments to restore the full-length capsid gene. For that 0.5 µl 5′ fragment, 0.5 µl 3′ fragment, 10 µl Phusion HF buffer, 1 µl dNTPs (10 mM), 1.5 µl DMSO, 0.5 µl Phusion Polymerase HS and 31.2 µl nuclease-free H_2O were mixed and run according to the cycling conditions listed in Table 21.

Table 21: Overlap extension part 1 cycling conditions

Step	Temperature [°C]	Time	Repetitions
1	98	30 sec	12 cycles
2	98	10 sec	
3	72	90 sec	
4	4	hold	

Afterwards 2.5 µl #178 primer as well as 2.5 µl #833 primer was added to the reaction and the capsid amplification was completed by starting the following cycler program:

Table 22: Overlap extension part 2 cycling conditions

Step	Temperature [°C]	Time	Repetitions
1	98	30 sec	
2	98	10 sec	25 cycles
3	60	15 sec	
4	72	90 sec	
5	72	10 min	
6	4	hold	

PCR product was separated on an agarose gel (3.1.2) and appropriate band was purified (3.1.4) before digesting the fragment and a WH-rep2 helper plasmid with HindIII-HF and SpeI (3.1.3). Full-length capsid gene was subsequently ligated into the gel-purified plasmid backbone (3.1.5) and transformed (3.1.6). Outcome of the DNA preparation (3.1.8) was used for either virus production or peptide insertion cloning.

3.2.2 PEPTIDE INSERTION

To clone the missing AAV peptide insertion variants where the new insertion site (NIS) is utilized, the AAV serotype 1, 2, 3, 4, 4mut, 5, 6, po1 and 12 capsid gene, in the corresponding WH-Rep2-CapNIS plasmids (Table 16), was digested with SfiI to enable oligonucleotide integration. Restriction digest and subsequent DNA purification was performed as described in chapter 3.1.3

and 3.1.4 respectively. DNA sequences of the peptides P2, P4, P5, A1, A2, A6 were ordered (Merck KGaA) and double-stranded fragments with sticky overhangs for the SfiI-digested capsid helper plasmids were produced. For that, 5 µl forward and reverse peptide oligonucleotide (Table 13) as well as 5 µl NEBuffer 2 and 35 µl nuclease-free H_2O were mixed. Cycler program listed in Table 23 was executed.

Table 23: Oligonucleotide annealing cycling conditions

Step	Temperature [°C]	Temperature decrease [°C/sec]	Time [min]
1	95		5
2	75		3
3	65	0.1	2
4	55	0.1	2
5	45	0.1	2
6	4		hold

The newly generated plasmids harboring the capsid genes with the inserted peptide-encoding DNA stretches are shown in Table 16.

3.2.3 BARCODED REPORTER PLASMIDS

For the generation of barcoded AAV reporter plasmids, an oligonucleotide bearing a 15 nt-long stretch of randomized nucleotides flanked by two Esp3I sites was ordered (Table 14, Barcode #2). A special request for a guaranteed 1:1:1:1 ratio of the N-wobble was sent to the manufacturer (Merck KGaA). The synthesis of the second strand was performed by mixing 0.5 µl Barcode #2 (100 µM), 0.5 µl Barcode #2_rv (100 µl), 10 µl Phusion HF buffer, 1 µl dNTPs (10 mM), 1.5 µl DMSO, 0.5 µl Phusion Polymerase HS and filled up to 50 µl with nuclease-free H_2O. Cycler program depicted in Table 20 was used with an extension time of 5 sec.

After a subsequent PCR clean-up 5 molar parts of double-stranded barcode oligonucleotide were mixed with 1 molar part of pscAAV-CMV-EYFP-ccdB-BGHpolyA, 1 µl ATP (10 mM), 1 µl DTT (10 mM), 1 µl Tango Buffer (10X), 1 µl T4 DNA Ligase, 0.75 µl Esp3I and 1.3 µl nuclease-free H_2O. Golden gate reaction was carried out with cycling conditions listed in Table 24.

Table 24: Golden gate cycling conditions

Step	Temperature [°C]	Time	Repetitions
1	37	5 min	20 cycles
2	16	5 min	
3	65	20 min	

Golden gate reaction mix was then directly used for electroporation (3.1.7). Individual colonies, each of them theoretically containing a unique barcode sequence, were picked and grown in 3 ml of LB media before extracting the plasmid DNA (3.1.8). All barcoded constructs were double-digested with PstI-HF and XmaI as described in 3.1.3 to check the ITR integrity. Positive clones were sent for sequencing with the primer #652 (Table 14). Barcodes with a length of more or less than 15 and with homopolymers >3 were excluded. Remaining barcodes were tested for their Hamming distance to each other utilizing a tailored Excel sheet. To guarantee a distinct identification during next generation sequencing, the cutoff for the Hamming distance was set to >4. 159 barcodes were generated matching all criteria and are depicted in Table 17.

3.3 VIRUS PRODUCTION

Subchapters of 3.3 comprise all necessary steps for the production of AAV vectors from cell seeding to determination of viral titers.

3.3.1 HEK293T SEEDING

Four days before transfection, 7.5×10^6 HEK293T cells per 175 cm^2 flask were seeded and grown in DMEM with 10% FBS, 1% P/S. After two days, cells of one flask were washed with 8 ml DPBS and subsequently harvested with 2 ml Trypsin-EDTA (0.25%). 8 ml of DMEM with 10% FBS, 1% P/S was used to stop the trypsinization. Cell count was determined and 4×10^6 cells per 15 cm dish were seeded resulting in 80% confluency after two days which provided optimal conditions for the transfection.

3.3.2 POLYETHYLENIMINE (PEI) TRANSFECTION

For the triple transfection a barcoded reporter (Table 17), capsid helper (Table 16) and adeno helper plasmid were combined in equimolar ratios

adding up to 25 µg total DNA. The final concentration of NaCl in the transfection mix was set to 300 mM and the N/P ratio to 30. PEI was added last to initiate the complex formation. After vigorous vortexing, the mix was incubated for 10 min at RT. 2 ml were evenly distributed on a 15 cm dish and the plates were kept at 37 °C, 5% CO_2 for three days before harvesting.

3.3.3 HEK293T HARVEST, LYSIS AND BENZONASE TREATMENT

Cells were detached with a cell scraper and the suspension was collected in a 500 ml conical tube. HEK293T cells were pelleted at 1000 rcf for 15 min and supernatant was discarded. 5 ml of Benzonase Buffer (15 ml for a large iodixanol gradient) were used to resuspend the pellet. Afterwards, the cells were lyzed by four freeze-thaw cycles to release viral particles. 75 U of Benzonase were added per 15 cm dish and the suspension was incubated for 1 h at 37 °C with occasional inverting in order to break down residual plasmid DNA, genomic DNA and RNA. Two subsequent centrifugation steps at 4000 rcf for 15 min at 4 °C were carried out to remove cellular debris from the virus-containing supernatant.

3.3.4 AAV PURIFICATION BY IODIXANOL GRADIENT

Iodixanol gradient was prepared by inserting a Pasteur pipette into an ultracentrifuge tube. Supernatant from 3.3.3 was transferred to the tube followed by 1.5 ml of 15%, 25%, 40% and 60% iodixanol solution in succession to build the gradient. For the large iodixanol gradient 7 ml, 5 ml, 4 ml and 4 ml were used for the phases, respectively. Afterwards, the Pasteur pipette was carefully removed and a 5 ml syringe with Benzonase Buffer was used to fill the tube to the top. Tubes were sealed with the Tube Sealer and balanced to each other (allowed deviation +/- 0.01 g). Ultracentrifugation was carried out in Rotor 70.1TI at 50000 rpm for 2 h at 4 °C (70TI at 63000 rpm for 2 h at 4 °C for large gradient). Upon completion, ultracentrifuge tube was punctured at the top with a 19 G needle to release vacuum and 3 mm below the 40%/60% barrier to extract ~1.2 ml of virus-containing fraction with a syringe (~2.5 ml for large gradients). 12 ml of DPBS was added to the purified virus solution and loaded on an Amicon Ultra-15 for dialyzing and concentrating by centrifuging at 3000 rcf for 1-5 min. Centrifugation steps were carried out multiple times and the solution was mixed in between until ~1.5 ml residual volume. Process was

repeated twice by filling the Amicon tube again with DPBS to further eliminate the iodixanol content. Final concentrate was aimed to have 0.5-0.8 ml and stored at -80 °C.

3.3.5 AAV PURIFICATION BY CESIUM CHLORIDE GRADIENT

The supernatant from 3.3.3 was combined with 1 M $CaCl_2$ to reach a final concentration of 25 mM $CaCl_2$ and incubated for 1 h on ice to precipitate the proteins. Subsequent centrifugation at 10000 rcf for 15 min at 4 °C was performed to pellet the proteins. Supernatant was taken, mixed with ¼ volumes of PEG-NaCl solution and incubated overnight on ice. Solution was centrifuged at 2500 rcf for 30 min at 4 °C and resulting supernatant was discarded. Pellet was resuspended with 10 ml Na-HEPES resuspension buffer, followed by centrifuging at 2500 rcf for 30 min at 4 °C. Supernatant was filled up to 24 ml with Na-HEPES resuspension buffer and 13.2 g of CsCl was added. Refractive index (RI) of virus solution was determined with refractometer and adjusted to 1.3710 by adding CsCl or Na-HEPES resuspension buffer. After transferring the solution to an OptiSeal ultracentrifuge tube and filling the tube with Topping solution, tubes were balanced to each other (allowed deviation +/- 0.01 g) and centrifuged at 45000 rpm for 21-23 h at 21 °C in a 70TI rotor. To harvest the virus, fractions were taken by puncturing the tube at the bottom with a 19 G needle. 3, 3, 0.5, 0.5, 0.5, 5, 0.5, 0.5, 0.5 and 3 ml fractions are collected dropwise and RI-values were measured. Fractions in the range of 1.3711-1.3766 were pooled, filled up to 9 ml with DPBS and transferred to a Slyde-A-Lyzer™ G2 Dialyse Cassette for dialysis against 700 ml cold DPBS. DPBS was replaced after 30 min without stirring. The next buffer exchanges were performed after 1 h, 2 h, overnight, 2 h and 2 h. Concentration of the 9 ml to ~1 ml was achieved by Amicon Ultra-15 centrifugation as described in 3.3.4. Purified virus was stored at -80 °C.

3.3.6 AAV TITRATION BY QPCR

For the quantification of viral titers, 10 µl of a purified virus sample from 3.3.4 or 3.3.5 was combined with 10 µl TE Buffer and 20 µl of 2 M NaOH. Solution was vortexed and incubated for 30 min at 56 °C to break up the viral capsids. Neutralization was performed by adding 38 µl of 1 M HCl before vortexing again and adding 922 µl of nuclease-free H_2O. The highest standard for the qPCR was set to $5x10^8$ copies of double-stranded DNA and serially diluted to $5x10^3$ copies. 5 µl of all 6 standards and 5 µl of the alkaline lysis were combined

with 17.5 µl SensiMix™ II Probe No-ROX (2x), 1.4 µl qPCR_EGFP_fw (10 µM), 1.4 µl qPCR_EGFP_rev (10 µM), 0.35 µl EGFP_Probe (10 µM) and 9.35 µl nuclease-free H_2O, respectively. Mix was vortexed and 10 µl were pipetted in triplicates into Strip Tubes and run with the cycler program listed in Table 25.

Table 25: AAV titration cycling conditions

Step	Temperature [°C]	Time	Repetitions
1	95	10 min	
2	95	10 sec	40 cycles
3	60	20 sec	

Output values of the cycler, x, were corrected for the two dilution steps and the 10 µl input volume to get to viral genomes per ml.

$$vg/ml = x \times 7 \times 100 \times 100$$

3.4 WORKFLOW FOR VARIANT VALIDATION

The following subchapters of 3.4 describe the full workflow for the *in vivo* validation of a barcoded AAV-library from the injection into mice to the analysis by next generation sequencing.

3.4.1 IN VIVO PROCEDURES

Seven-week-old mice ordered from Janvier Labs were i.v. injected with ~$1x10^{12}$ vg/mouse of the barcoded AAV library via the tail vein. After 1-2 weeks abdominal aorta, thoracic aorta, brain, biceps, blood cells, colon, diaphragm, duodenum, eye, brown fat, white fat, heart, inner ear, kidney, liver, lung, ovaries, pancreas, quadriceps femoris, spleen and stomach were harvested and tissue pieces were submerged in RNA*later* solution before storing at -20 °C.

3.4.2 MACS FOR IMMUNE CELLS

Isolation of CD3ε-, CD11b-, CD11c-, CD19-positive cells was performed by harvesting the mandibular, accessory mandibular, subiliac, proper axillary, accessory axillary and medial iliac lymph nodes as well as the spleen. Tissues were transferred to a 70 µm strainer and homogenized with a plunger. After

washing the strainer with MACS buffer, resulting cell suspension was centrifuged at 1000 rcf for 5 min. Supernatant was aspirated and pellet was resuspended in 10 ml RBC lysis solution before incubating 5 min at RT. Cells were centrifuged again at 1000 rcf for 5 min and resuspended in 1 ml MACS buffer yielding approximately $1x10^8$ cells/ml. Cell suspension was split into two 500 µl fractions. 100 µl CD11c and CD11b MicroBeads were added respectively and following steps were carried out according to manufacturer's instructions. Flow-through of both purifications was kept and used to isolate CD19- and CD3-positive cells respectively by following manufacturer's instructions. Purified cells were counted and subsequently pelleted before freezing in liquid nitrogen for storage at -80 °C.

3.4.3 TISSUE HOMOGENIZATION

Isolated tissues were removed from RNA*later* solution and weighed at RT. After transferring the tissue to a respective Precellys® tube, 350 µl of RLT, 1% β-ME was added for every 10 mg of tissue. Tubes were placed into Precellys® 24-Dual homogenizer and homogenized by using program 1 with 5500 rpm for 20 sec. Procedure was repeated for samples with insufficient homogenization. Lysates were stored at -80 °C (or at 4 °C for 1-2 h).

Pellets of the purified cells from 3.4.2 were resuspended with 300 µl RLT, 1% β-ME for every $1x10^6$ cells and incubated for 5 min at RT. Lysates were transferred to a QiaShredder tube and centrifuged at 13000 rcf for 2 min.

3.4.4 PHENOL-CHLOROFORM EXTRACTION

PLG-tubes were centrifuged at 16000 rcf for 30 sec to collect the gel at the bottom of the tube. Afterwards 400 µl Phenol:Chloroform:Isoamylalcohol was added. Tissue lysates were thawed and subsequently centrifuged at 4000 rpm for 4 min to pellet potential debris. 400 µl of tissue lysate was transferred to a prepared PLG-tube and shaken vigorously for 15 sec. After centrifugation at 16000 rcf for 5 min, 400 µl Chloroform:Isoamylalcohol was added and PLG-tubes were again shaken vigorously for 15 sec. Tubes were incubated for 3 min at RT before centrifuging at 16000 rcf for 5 min. 350 µl of the aqueous phase was transferred to a 96-deepwell plate and stored at -80 °C (or 4 °C for 1-2 h).

3.4.5 DNA/RNA EXTRACTION

For the isolation of DNA and RNA from the 350 µl aqueous phase from 3.4.4 and the 300 µl RLT-lysate of the immune cells from 3.4.3 the Allprep DNA/RNA 96 Kit was used. Steps 3-10 of the manufacturer's instructions were followed. Step 11 was performed with only 400 µl of RW1 and followed by a DNase on-column digest. DNase stock solution was prepared by adding 550 µl RNase-free H_2O to one vial of lyophilized DNase I. DNase I incubation mix was prepared by adding 70 µl RDD buffer to 10 µl DNase stock solution and gently mixing. 80 µl of DNase I incubation mix was directly added to the RNeasy column in each well. Plate was sealed with a new sheet of AirPore Tape and incubated for 15 min at RT. Step 11 was repeated with only 400 µl RW1. Steps 12/16, 13/17, 14/18 and 15/19 were performed in parallel. RNA was eluted twice with 50 µl RNase-free H_2O, DNA twice with 75 µl EB buffer.

3.4.6 DNASE TREATMENT

To guarantee a complete removal of remaining gDNA in RNA samples, 212 ng of the RNA isolated in 3.4.5 was digested with DNase I. DNase I stock solution was prepared as described in 3.4.5. DNase I incubation mix was prepared by adding 1 µl DNase I stock solution and 4 µl RDD buffer to the 212 ng RNA. Final volume was filled up to 40 µl with nuclease-free H_2O. RNA was incubated for 15-30 min at RT and DNase I was subsequently heat-inactivated for 10 min at 75 °C. DNase I-treated RNA was stored at -80 °C.

3.4.7 CDNA SYNTHESIS

DNase I-treated RNA from 3.4.6 was directly used for the High-Capacity cDNA Reverse Transcription Kit. Kit components were thawed on ice and incubation mix was prepared by adding 4 µl 10X RT buffer, 1.6 µl 25X dNTP Mix, 4 µl 10X RT Random Primers and 2 µl MultiScribe Reverse Transcriptase to 28.4 µl DNase I incubation mix from 3.4.6 containing 150 ng of RNA. PCR cycler was used to incubate the mix for 10 min at 25 °C, 120 min at 37 °C and 5 min at 85 °C. Synthesized cDNA was stored at -20 °C.

3.4.8 AMPLIFICATION OF BARCODE REGION

To amplify the barcode region of the viral transcripts (from 3.4.7) or genomes (from 3.4.5) a PCR was performed resulting in a 112 bp amplicon. For the reaction 10 µl 5X Phusion HF buffer, 1 µl dNTPs, 0.25 µl Fwd4 (100 µM), 0.25 µl Rev3 (100 µM), 0.5 µl Phusion Hot Start II Polymerase and 25 ng of cDNA or gDNA template were mixed and filled to 50 µl with DEPC-treated H_2O. The PCR cycler program shown in Table 26 was used.

Table 26: Barcode region PCR cycling conditions

Step	Temperature [°C]	Time	Repetitions
1	98	30 sec	40 cycles
2	98	10 sec	
3	72	20 sec	
4	72	5 min	
5	4	hold	

PCR reaction was subsequently cleaned up with the MagMAX Express-96 Magnetic Particle Processor by adding 100 µl of Agencourt AMPure XP beads to the 50 µl of PCR reaction. Sample was mixed thoroughly by pipetting up and down 10 times and incubated for 10 min at RT. Two MagMAX wash plates were prepared with 150 µl 80% EtOH and one MagMAX plate with 25 µl Illumina Resuspension Buffer. MaxMAX program "AMPure_Trueseq96stan" was started and instructions of the machine were followed. After completion of the run 25 µl eluate was transferred to a 96-well plate and stored at -20 °C.

PCR outcome and DNA concentration was analyzed by using a Fragment Analyzer with the Standard Sensitivity NGS Fragment Analysis Kit according to manufacturer's instructions.

3.4.9 LIBRARY PREPARATION

In order to allow sequencing on the NextSeq500 platform a library preparation was performed where the PCR amplicons from 3.4.8 are ligated to sequencing adaptors. The Ovation Library System for Low Complexity Samples Kit was followed according to manufacturer's instructions to process 20-30 ng of amplicon DNA per sample. Result was monitored by running the processed samples on a Fragment Analyzer with the Standard Sensitivity NGS Fragment Analysis Kit according to manufacturer's instructions.

3.4.10 DNA QUANTIFICATION WITH PICOGREEN

To determine the DNA concentration of the sequencing library samples from 3.4.9, the Quant-iT PicoGreen dsDNA Assay Kit was used. PicoGreen was thawed and 1X TE buffer was prepared with the 20X TE stock solution. 1 µl of PicoGreen was added to 200 µl 1X TE buffer for each sample/standard to be analyzed. 200 µl of the mixture was transferred to a black 96-well plate for each sample. Eight DNA standards were prepared with a serial dilution ranging from 100 ng/µl to 1.56 ng/µl. 2 µl of standard or sample were added to the 200 µl PicoGreen/TE buffer solution, mixed by vortexing and measured with a Cytation 5 imaging reader by using the "QuantiT_PicoGreen_dsDNA" program (filter settings 485/20,530/25).

3.4.11 PREPARATION OF LIBRARY POOLS AND STARTING NEXTSEQ

Based on the DNA concentrations obtained in 3.4.10, a 2 nM dilution was prepared for each sample with Illumina Resuspension Buffer, 0.1% Tween20. 10 µl of every 2 nM dilution with a unique reverse adaptor which is supposed to be multiplexed on the flow cell were mixed and stored at -20 °C until library denaturation.

For the denaturation of the library fragments 5.3-6.0 µl of the library pool were used and filled up to 10 µl with Illumina Resuspension Buffer, 0.1% Tween20. 10 µl of 0.2 M NaOH were added, vortexed and incubated for 5 min at RT to denature the DNA strands. For the neutralization 10 µl of 200 mM Tris-HCl, pH 7.0 were added and sample was vortexed. Denatured library pool dilution was filled to 1 ml with 970 µl of pre-chilled HT1 buffer, mixed and 117 µl was combined with 1183 µl of pre-chilled HT1 buffer. 2 µl of 20 pM PhiX control was spiked in. Finished library pool dilution was vortexed thoroughly, spun down and loaded into a NextSeq500 cartridge.

For starting the NextSeq500 machine, instructions on the screen were followed. Read 1 was set to 84 and Index 1 to 8. SampleSheet.csv which is needed for subsequent demultiplexing was placed in the automatically created run folder.

3.4.12 DETECTION OF VIRAL GENOMES BY QPCR

To determine the EYFP and GAPDH copy number in the extracted DNA from 3.4.5, a TaqMan qPCR was performed by using 15 µl QuantiFast PCR Master Mix, 0.5 µl 60X Primer-Probe Mix (EYFP or GAPDH) and 14.5 µl sample (75 ng) or standard. Mix was vortexed and 10 µl of each sample or standard mix were transferred to a 384-well plate in duplicates. Plate was sealed and centrifuged at 800 rcf for 5 min. qPCR was started with following cycler program:

Table 27: qPCR cycling conditions

Step	Temperature [°C]	Time	Repetitions
1	50	2 min	
2	95	10 min	
3	95	15 sec	40 cycles
4	60	1 min	

Determined copy number of GAPDH was divided by two to obtain the number of cells. EYFP copy number was divided by the amount of cells resulting in viral genomes per diploid genomes (cells). Those values were used for data normalization.

3.4.13 NGS DATA NORMALIZATION

The NGS data obtained from 3.4.11 were processed by using a modified Python 2.7 script[224] (modified by Josefine Sippel and Jonas Weinmann) which uses the demultiplexed reads from the sequencer and searches for the known 15 nt-long barcode sequences. The output file lists the unknown sequences as well as the variant-assigned barcodes with their corresponding read counts.

A second Python 2.7-based script (written by Sabrina Weis) utilizes the output files from the first script and performs a multi-step normalization procedure which corrects for the variations in the total read counts of each flow cell, unbalanced composition of the initial viral injection mixture and different efficiencies of the AAV library in the analyzed tissues. In the first step the script is normalizing the read counts R of all variants α in tissue β to the sum of all variants α in β to obtain the proportion $P_{\alpha\beta}$.

$$P_{\alpha\beta}=\frac{R_{\alpha\beta}}{\sum_{\alpha} R_{\alpha\beta}}$$

The second step normalizes $P_{\alpha\beta}$ to the proportion of each variant α in the initial library L_{α} which corrects for the uneven composition in library.

$$P^*_{\alpha\beta}=\frac{P_{\alpha\beta}}{L_{\alpha}}$$

In the third step $P^*_{\alpha\beta}$ is normalized to the qPCR-determined vg/dg (see 3.4.12), termed G_{β}, to allow a comparison of one variant α over all analyzed tissues β.

$$B_{\alpha\beta}=\frac{P_{\alpha\beta}}{L_{\alpha}}\times G_{\beta}$$

At this point $B_{\alpha\beta}$ values were used and depicted directly to generate heat maps visualizing the differences of all variants α in all tissues β. $B_{\alpha\beta}$ values can also be shown as proportion of the sum over α or β of $B_{\alpha\beta}$.

$$V_{\alpha\beta}=\frac{B_{\alpha\beta}}{\sum_{\alpha} B_{\alpha\beta}}$$

$$T_{\alpha\beta}=\frac{B_{\alpha\beta}}{\sum_{\beta} B_{\alpha\beta}}$$

$V_{\alpha\beta}$ values were taken to create bar plots which demonstrate the proportion of all variants α in one tissue β and therefore exemplify the efficiency of the individual vectors. Bar plots using $T_{\alpha\beta}$ values show the proportion of one variant α in all tissues β allowing an analysis of the tissue specificity.

3.5 HISTOLOGY

In order to validate promising candidates from the barcode-based *in vivo* screening, C57BL/6J mice were i.v. injected with $5x10^{11}$ vg/mouse and kept for 2 weeks before harvesting the biceps, diaphragm, heart, liver and quadriceps femoris. Injected viruses carried a CMV promoter -driven *egfp* with a BGH poly-A (Table 18). Tissues were fixed in 4% PFA for 15-22 h and subsequently transferred to 30% sucrose solution until the tissue sinks to the bottom of the tube (~6 h). Afterwards, organs were embedded in TissueTek® O.C.T Compound, frozen on dry ice and stored at -80 °C. 12 µm sections were cut and embedded in ProLong™ Gold antifade reagent containing DAPI. Sections were scanned with Axio Scan.Z1 detecting the DAPI and GFP signal.

4 RESULTS

4.1 ESTABLISHMENT OF BARCODE-BASED AAV CAPSID SCREENING

The following chapter encompasses an overview of the barcode-based capsid screening pipeline to describe the fundamental principle of this workflow. Data generated by using the pipeline and more detailed information about the individual experimental outlines are shown in the chapters 4.3, 4.4 and 4.5.

In order to enable a barcode-based capsid screening *in vivo*, randomized 15 nucleotide-long DNA sequences were cloned into the 3'UTR of a CMV promoter-driven *eyfp* gene by Golden Gate cloning (3.2.3). Resulting clones were tested for the presence of the essential ITRs by restriction digest (3.1.3) and the barcode region was subsequently sequenced. Barcodes with a length differing from 15 nucleotides or comprising homopolymers longer than 3 nucleotides were excluded. The Hamming distance of the remaining pool was assessed and sequences with variations to every other barcode in at least five positions were kept.

A total of 159 barcodes could be generated matching all criteria (Table 17) and were used for vector production. During the latter, one barcoded construct was transfected into HEK293T cells together with a plasmid bearing the *rep* gene of AAV2 and a *cap* gene of choice. Hence, a tight linkage of a barcode to its respective capsid was established (Figure 4A). Each variant was produced separately and eventually pooled to create a barcoded library. Afterwards, the viral library was dialyzed as well as concentrated (Figure 4B).

For parallel validation in mice, $1x10^{12}$ vector genomes (vg) per mouse were injected into six C57BL/6J mice via the tail vein. Mice were kept for 1-2 weeks before tissues and cells of interest were harvested (Figure 4C). Steps for the extraction of DNA and RNA, the subsequent PCR amplification of the barcode region and the mandatory clean-up of the PCR product were optimized (data not shown). After completing the library preparation (3.4.9), the samples were multiplexed and processed by next generation sequencing (NGS) to identify the proportion of every barcode in the analyzed tissues (Figure 4D). To this end, a multi-layer normalization strategy was applied that corrects for the total read count differences of each flow cell, the variations in particle abundance in

the initial viral injection mixture and unequal transduction efficiencies of the AAV library in the tissues (3.4.13).

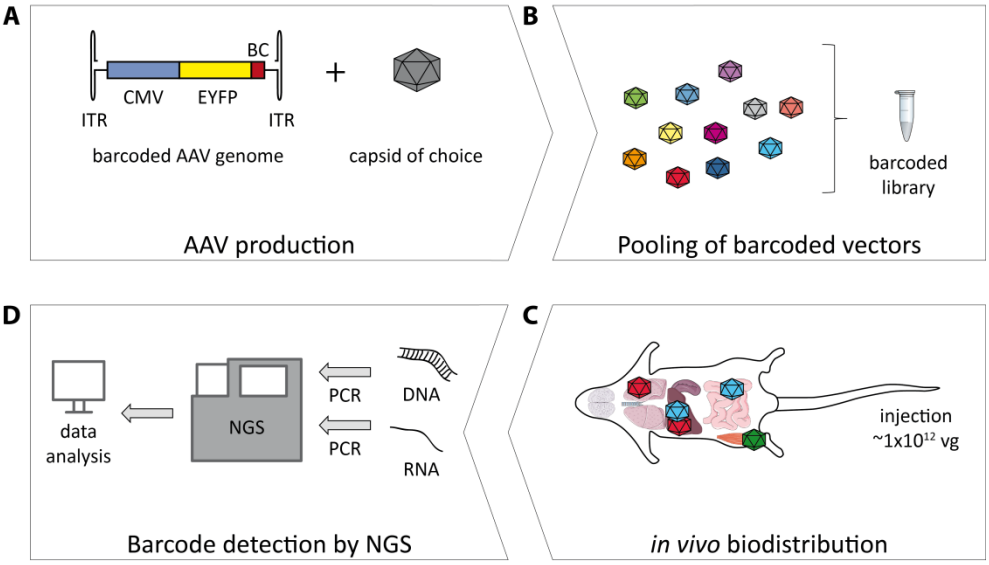

Figure 4: Workflow for a barcode-based AAV capsid validation
(A) Self-complementary AAV genome comprising a CMV promoter-driven *eyfp* transgene flanked by ITRs. The 3'UTR contains a 15 nucleotide-long barcode (BC) allowing capsid tracking on the DNA and cDNA level. During AAV production, barcoded genomes were paired with a *cap* gene of choice. **(B)** Each barcoded variant was produced separately and eventually pooled to generate a barcoded library. **(C)** C57BL/6J mice were i.v. injected with $1x10^{12}$ vg/mouse and kept for 1-2 weeks before harvesting tissues and cells. **(D)** DNA and RNA were extracted from all samples and the barcode region was amplified by PCR. NGS was performed to determine barcode read counts of all variants in the tissues.

To test the pipeline, 12 barcoded reporter constructs were transfected into HEK293T cells and barcode sequences were amplified from the cell lysate. All 12 DNA patterns could be detected by sequencing (data not shown). To further evaluate if the NGS sensitivity was high enough for a barcode identification from *in vivo* samples, two barcoded vectors, AAV2wt (wild type) and AAV8wt, were produced and mixed, and $7.22x10^{11}$ vg/mouse were injected into two female C57BL/6J mice. After two weeks, liver, heart and kidney were harvested and barcode abundance was measured in samples from DNA (data not shown) and cDNA (Figure 5).

Both barcodes could be found in the three tissues on the DNA and cDNA level. On the cDNA level, AAV8wt demonstrated superior efficiency in the liver, heart and, to a lesser extent, in the kidney compared to AAV2wt. Evidence is that in all cases but one, over 95% of the detected barcode sequences belonged to AAV8wt which is in line with expectations from the literature[153].

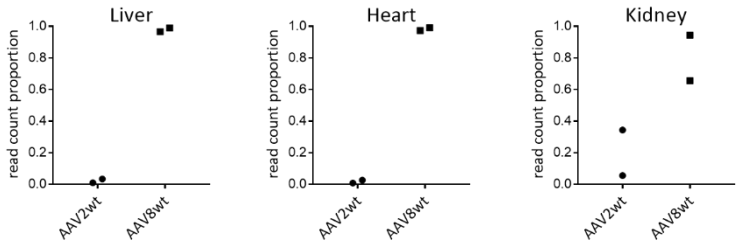

Figure 5: Barcode abundance in a pilot run with AAV2wt and AAV8wt
Two C57BL/6J mice were i.v. injected with 7.22×10^{11} vg/mouse of a mixture of AAV2wt and AAV8wt. Depicted is the proportion of the corresponding barcodes of the serotypes in the liver, heart and kidney cDNA samples. Dots represent individual mice.

4.2 AAV VARIANTS USED IN THE SCREENINGS

In this work, three independent library screenings were performed with varying vector compositions. To point out the differences between the individual screening rounds, this chapter provides an overview of all analyzed capsid variants in the three screenings and the respective library compositions (Table 28) as well as information about the origin of the capsids. Results obtained by applying these libraries *in vivo* are described in the chapters 4.3, 4.4 and 4.5.

The 1st generation library contained 91 capsids, among them 13 parental serotypes and 78 peptide-modified variants based on these natural AAVs. A highly similar panel was previously tested extensively *in vitro* in our laboratory by primarily Kathleen Börner and Eike Kienle (more information is found in the doctoral thesis of Eike Kienle). During this doctoral work, variants displaying the peptides P2, P4, P5, A1, A2 and A6 were cloned utilizing an alternative insertion site for the AAV serotypes 1, 2, 3, 4, 5, 6, po1 and 12 after amino acid position 588, 587, 588, 586, 577, 588, 569, 594, respectively. Structural modeling of the integration site, the variable region VIII, hinted towards a potential transduction improvement (unpublished data) by slightly shifting the peptide insertion site, aiming to better match the insertion position in the VRVIII loop of AAV7, AAV8, AAV9 and AAVrh10 (position 589, 590, 588 and 590, respectively). The amino acid after which the peptide is displayed varies marginally due to differences in the total VP protein size of the naturally occurring AAV isolates.

The first screening revealed dramatic variations in production efficiency of the analyzed variants (4.3). Therefore, poor producers were excluded for the production of the 2nd generation library but replaced with published benchmarks from the literature, namely, AAV2_7m8[186], AAV2_BR1[218],

AAV2_L1[208], AAV2HBKO[203], AAV6.2[165], AAV9_PHP.B[219], AAVDJ[172], AAVLK03[191] and AAVshH10[179].

Remains of the 2nd generation library were used to create the 3rd generation library by spiking in the benchmarks AAV2_L1mut1[208], AAV2_L1mut2[208], AAV2YF[161], AAV9_K1[215], AAV9_K3[215], AAV9_PHP.A[219], AAV9_PHP.eB[220], AAV9_PHP.S[220], AAV9LD[222], AAVAnc80L65[167], AAVB1[190] and AAVM41[189]. Additionally, the library was enriched with 30 chimeric variants selected in stellate cells (work of Anne-Kathrin Herrmann) and 34 chimeras isolated from muscle tissue selections (work of Jihad El Andari).

Table 28: Variants in the screenings

Variant	Source	1st library	2nd library	3rd library
AAV1wt	[43]	x	x	x
AAV1_A1	Our lab	x	x	x
AAV1_A2	Our lab	x	x	x
AAV1_A6	Our lab	x	x	x
AAV1_P2	Our lab	x		
AAV1_P4	Our lab	x	x	x
AAV1_P5	Our lab	x	x	x
AAV2wt	[43]	x	x	x
AAV2_7m8	[186]		x	x
AAV2_A1	Our lab	x	x	x
AAV2_A2	Our lab	x	x	x
AAV2_A6	Our lab	x	x	x
AAV2_BR1	[218]		x	x
AAV2_L1	[208]		x	x
AAV2_L1mut1	[208]			x
AAV2_L1mut2	[208]			x
AAV2_MTP	[201]			x
AAV2_P2	Our lab	x	x	x
AAV2_P4	Our lab	x	x	x
AAV2_P5	Our lab	x	x	x
AAV2HBKO	[203]		x	x
AAV2YF	[161]			x
AAV3bwt	[46]	x	x	x
AAV3b_A1	Our lab	x	x	x
AAV3b_A2	Our lab	x	x	x
AAV3b_A6	Our lab	x	x	x
AAV3b_P2	Our lab	x		
AAV3b_P4	Our lab	x	x	x

AAV3b_P5	Our lab	x	x	x
AAV4wt	[44]	x	x	x
AAV4_A1	Our lab	x	x	x
AAV4_A2	Our lab	x	x	x
AAV4_A6	Our lab	x	x	x
AAV4_L1	Our lab		x	x
AAV4_P2	Our lab	x	x	x
AAV4_P4	Our lab	x	x	x
AAV4_P5	Our lab	x	x	x
AAV4mutwt	Our lab	x		
AAV4mut_A1	Our lab	x		
AAV4mut_A2	Our lab	x		
AAV4mut_A6	Our lab	x		
AAV4mut_P2	Our lab	x		
AAV4mut_P4	Our lab	x		
AAV4mut_P5	Our lab	x		
AAV5wt	[45]	x	x	x
AAV5_A1	Our lab	x	x	x
AAV5_A2	Our lab	x	x	x
AAV5_A6	Our lab	x	x	x
AAV5_P2	Our lab	x		
AAV5_P4	Our lab	x	x	x
AAV5_P5	Our lab	x	x	x
AAV6wt	[46]	x	x	x
AAV6_A1	Our lab	x		
AAV6_A2	Our lab	x		
AAV6_A6	Our lab	x		
AAV6_P2	Our lab	x		
AAV6_P4	Our lab	x	x	x
AAV6_P5	Our lab	x		
AAV6.2	[165]		x	x
AAV7wt	[47]	x	x	x
AAV7_A1	Our lab	x	x	x
AAV7_A2	Our lab	x	x	x
AAV7_A6	Our lab	x	x	x
AAV7_P2	Our lab	x	x	x
AAV7_P4	Our lab	x	x	x
AAV7_P5	Our lab	x	x	x
AAV8wt	[47]	x	x	x
AAV8_A1	Our lab	x	x	x
AAV8_A2	Our lab	x	x	x

AAV8_A6	Our lab	x	x	x
AAV8_P2	Our lab	x	x	x
AAV8_P4	Our lab	x	x	x
AAV8_P5	Our lab	x	x	x
AAV9wt	[48]	x	x	x
AAV9_A1	Our lab	x	x	x
AAV9_A2	Our lab	x	x	x
AAV9_A6	Our lab	x	x	x
AAV9_BR1	Our lab		x	x
AAV9_K1	[215]			x
AAV9_K3	[215]			x
AAV9_P1	Our lab[202]		x	x
AAV9_P2	Our lab	x		
AAV9_P3	Our lab			x
AAV9_P4	Our lab	x	x	x
AAV9_P5	Our lab	x	x	x
AAV9K449R_PHP.A	[219]			x
AAV9K449R_PHP.B	[219]		x	x
AAV9K449R_PHP.eB	[220]			x
AAV9K449R_PHP.S	[220]			x
AAV9BI	Boehringer Ingelheim		x	x
AAV9LD	[222]			x
AAVrh10wt	[48]	x	x	x
AAVrh10_A1	Our lab	x	x	x
AAVrh10_A2	Our lab	x	x	x
AAVrh10_A6	Our lab	x	x	x
AAVrh10_P2	Our lab	x	x	x
AAVrh10_P4	Our lab	x	x	x
AAVrh10_P5	Our lab	x	x	x
AAVpo1wt	[50]	x	x	x
AAVpo1_A1	Our lab	x	x	x
AAVpo1_A2	Our lab	x	x	x
AAVpo1_A6	Our lab	x	x	x
AAVpo1_P2	Our lab	x		
AAVpo1_P4	Our lab	x	x	x
AAVpo1_P5	Our lab	x	x	x
AAV12wt	[49]	x	x	x
AAV12_A1	Our lab	x		
AAV12_A2	Our lab	x		
AAV12_A6	Our lab	x		

AAV12_P2	Our lab	x		
AAV12_P4	Our lab	x		
AAV12_P5	Our lab	x		
AAVAnc80L65	[167]			x
AAVB1	[190]			x
AAVDJ	[172]		x	x
AAVDJYF	Our lab			x
AAVLK03	[191]		x	x
AAVM41	[189]			x
AAVshH10	[179]		x	x
AAVAH chimeras	Our lab			x
AAVJEA chimeras	Our lab			x

4.3 1ST GENERATION LIBRARY SCREENING

The first generation library comprised a total of 91 different AAV variants (Table 28), each of them containing a barcode that served as unique identifier. To save time, two 15 cm dishes of HEK293T cells were used for the production of each vector. Afterwards, the cell lysates of the individual productions were pooled and subsequently purified by one cesium chloride gradient. The library was then processed on a NextSeq500 sequencer to determine the proportion of the variant-encoding barcodes in this mixture. This step is essential to verify the presence of each barcode and therefore the cognate capsid. Additionally, the respective proportion was used to normalize for potential variations in production efficiencies. Therefore, the theoretical mean proportion in case of an equimolar library was calculated and fold changes compared to this value are depicted in Figure 6. Values close to 1 or -1 reflect a production behavior according to the expectations.

However, drastic differences in barcode abundance could be observed for the variants in the library. Peptide insertion mutants frequently demonstrated a proportional decrease whereas vectors from the AAV5 or AAV9 family were generally over-represented. The most pronounced reduction in barcode abundance could be detected for peptide insertion variants of AAV6 and AAV12, with up to 978-fold deviation from the theoretical mean proportion (Figure 6). Also worth noting is that P2-modified capsids typically gave the lowest yields within the respective family. The overall lowest amount of read counts was found for AAV4mut_A2 whose titer was 3600-fold decreased versus the mean.

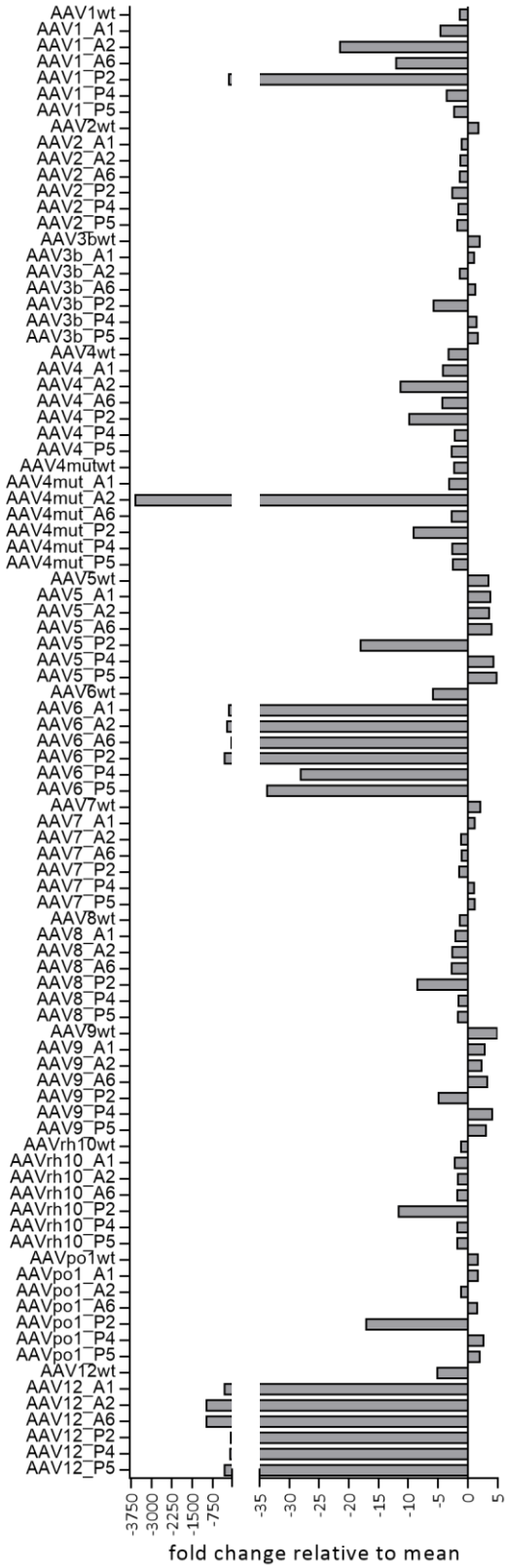

Figure 6: Composition of 1st generation library

Shown are fold changes to the theoretical mean proportion of each barcoded variant. A value close to 1 or -1 reflects the expected production behavior. Negative fold changes imply under-representation of the respective variant, positive values illustrate over-representation.

To study the library *in vivo*, $1x10^{12}$ vg/mouse were injected i.v. into three female and three male C57BL/6J mice. Mice were kept for two weeks before harvesting abdominal aorta, thoracic aorta, brain, colon, diaphragm, duodenum, eye, brown fat, white fat, heart, inner ear, kidney, liver, lung, pancreas, quadriceps femoris and spleen. DNA and RNA were extracted and the workflow described in 4.1 was followed. Sequencing data was normalized to the bias of the viral injection mixture (see above and Figure 6) and the resulting normalized proportions of each variant on the cDNA level in the analyzed tissues are shown as a bar plot (Figure 7).

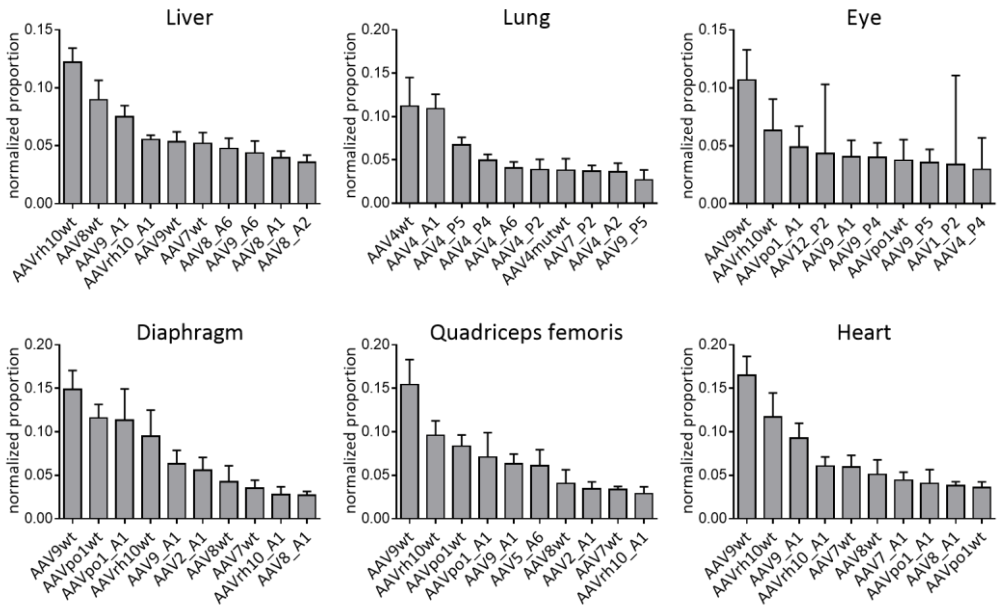

Figure 7: Transcriptional efficiency in various tissues
The depicted bar plots show the transcriptional efficiency as normalized proportion of the top 10 AAV variants from the 1st generation library in the liver, lung, eye, diaphragm, quadriceps femoris and heart. The cDNA values are the average from six C57BL/6J mice with SD.

In most of the studied organs, the wild type versions of AAV9, AAVrh10, AAV8 and AAV7 exhibited the highest efficiencies in this order, followed by peptide insertion variants based on these serotypes (Supplementary information, Figure 30). Different effects could be observed in the liver, lung, eye, diaphragm, quadriceps femoris and heart (Figure 7).

In the liver, AAVrh10wt displayed the highest efficiency even displacing the potent AAV8wt[226,227]. Peptide integrations into either AAVrh10wt or AAV8wt could not boost their effect. A strong enrichment of AAV4 and its related variants was detected in the lung, with AAV4wt being the top hit. AAV4mutwt that differs in only one amino acid (K544E) showed a 3-fold reduction as compared to its unmodified counterpart.

In the eye and the three muscle tissues diaphragm, quadriceps femoris and heart, a previously barely characterized AAV isolate, AAVpo1wt[50,228], appeared in the top 10 list of the most abundant barcodes. Furthermore, the porcine variant and its A1-modified version were found in the inner ear (Supplementary information, Figure 30). The two capsids demonstrated only weak efficiencies in other organs, indicating a preferential targeting of muscle. AAV9wt exhibits the highest normalized proportion in muscle tissues in line with its reputation as a gold standard for muscle transduction[229]. Of note, AAV9wt was not among the top 10 hits in the liver, potentially explaining the pronounced effects in several other tissues.

4.4 2ND GENERATION LIBRARY SCREENING

Because of the up to 3600-fold deviation from the theoretical mean proportion of the 1st generation library (Figure 6) and the resulting major implications for the normalization of the results, the production procedure was altered for the second library. For the first screening, two 15 cm dishes had been used to produce each variant, and the resulting particles had been pooled and concurrently purified without prior titration (4.3). However, as shown, this led to a heterogeneous vector abundance. Based on these findings and on experiences made by individually testing the production efficiencies of several wild type capsids (collected in a newly created internal AAV production database), the number of dishes required to achieve comparable yields was calculated for every variant and found to range from only one plate for highly potent producers, such as AAV5wt, to 120 plates for very poor candidates, such as AAV2_L1. Variants that would have required unfeasible amounts of plates, e.g. many P2-modified mutants or AAV6 and AAV12 with peptide insertions, were excluded. For each of the selected 82 vectors (Table 28), the aim was to reach 1.2×10^{11} vg after individual purification over an iodixanol gradient. Viral titers were determined by qPCR, and equimolar amounts were pooled and subsequently concentrated as well as dialyzed using Amicon Ultra-15 tubes. This 2nd generation library was sequenced to monitor the composition of the mixture and to generate seminal basal values for the normalization strategy (Figure 8).

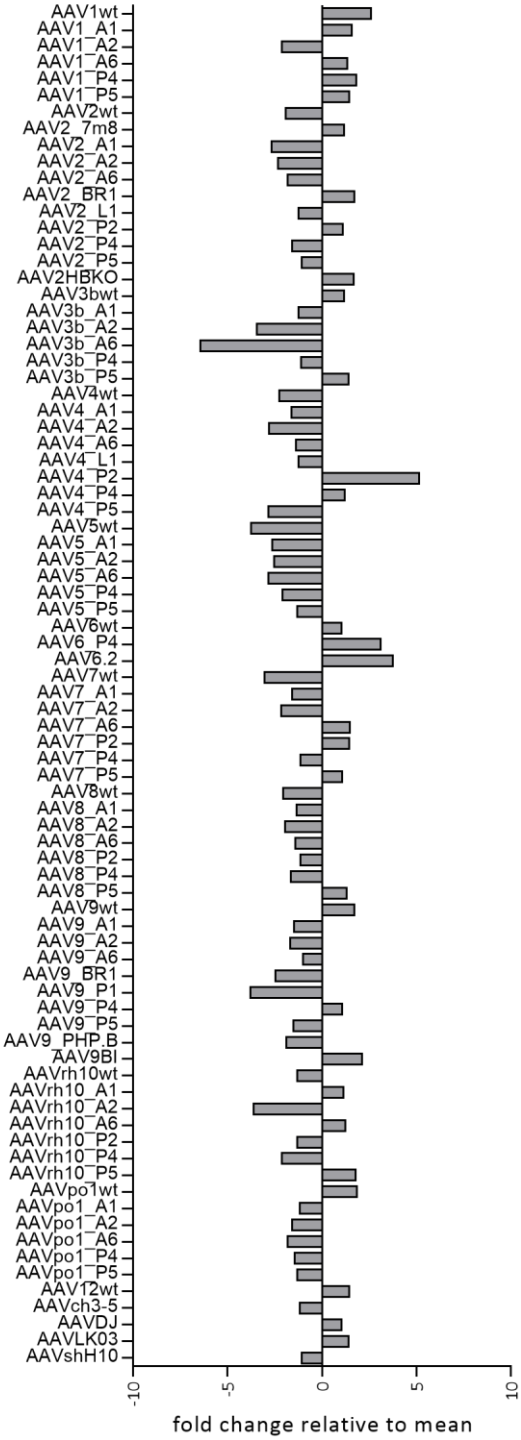

Figure 8: Composition of 2nd generation library
Shown are the fold changes to the theoretical mean proportion of each barcoded variant. A value close to 1 or -1 reflects the expected production behavior. Negative fold changes imply under-representation of the respective variant, positive values illustrate over-representation.

Unlike in the first library, substantial improvements could be observed for the second AAV pool, which showed only up to 6.4-fold under-representation of individual capsids (AAV3b_A6) or 5.1-fold over-representation (AAV4_P2).

Most others oscillated around the expected values of 1 or -1, illustrating a very homogenous capsid distribution and lowering the risk of normalization artefacts.

After completing the quality control step, $1x10^{12}$ vg/mouse were injected i.v. into six female C57BL/6J mice. After one week, mice were sacrificed and abdominal aorta, thoracic aorta, blood cells, brain, colon, diaphragm, duodenum, eye, brown fat tissue, white fat tissue, heart, inner ear, kidney, liver, lung, ovaries, pancreas, quadriceps femoris, spleen and stomach were extracted. DNA and RNA were isolated for subsequent deep sequencing. Furthermore, a qPCR was performed to determine the viral genomes per diploid genome (vg/dg) in each tissue. These values are depicted in Figure 9 and were additionally used to normalize the sequencing data. By implementing this step, for the first time, a comparison of one variant across all analyzed tissues was enabled (3.4.13), providing the opportunity to concurrently gather data on capsid efficiency and specificity.

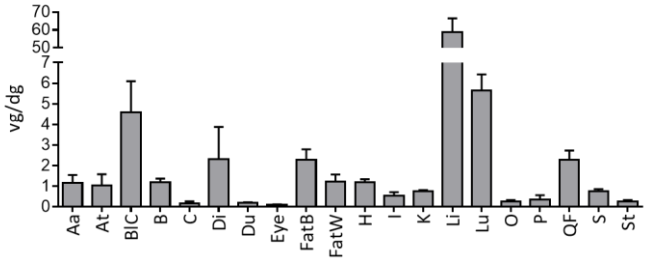

Figure 9: Viral DNA distribution of the 2nd generation library
The depicted bar plot shows the viral DNA distribution from the 2nd generation library after systemic injection into C57BL/6J mice across abdominal aorta (Aa), thoracic aorta (At), blood cells (BlC), brain (B), colon (C), diaphragm (Di), duodenum (Du), eye, brown fat (FatB), white fat (FatW), heart (H), inner ear (I), kidney (K), liver (Li), lung (Lu), ovaries (O), pancreas (P), quadriceps femoris (QF), spleen (S) and stomach (St). Detected viral genomes (EYFP probe) were normalized to GAPDH as a housekeeper. Depicted values represent the average of six mice with SD.

By determining the vg/dg values, the distribution of the 2nd generation library could be tracked across different tissues. As expected, the liver harbors the largest proportion of the viral particles with 59 vg/dg, followed by the lung (5.6 vg/dg) and the blood cells (4.6 vg/dg). In abdominal aorta, thoracic aorta, brain, diaphragm, brown fat tissue, white fat tissue, heart, inner ear, kidney, quadriceps femoris and spleen, roughly one viral genome was found in every cell (assuming a diploid genome per cell). Tissue types from the digestive tract, namely colon, duodenum and stomach, demonstrated very low values. In detail, only one viral genome could be detected in every fifth cell. The lowest transduction of only 0.1 vg/dg was found in the eye.

During processing of the NGS data, $B_{\alpha\beta}$ values were calculated by a custom-made Python script (3.4.13). These values allow the generation of a heat map displaying the full biodistribution of each variant in the library on the cDNA level (Figure 10). Thus, the transcript abundance of vectors can be compared either within the same organ or across all tissues. The color scheme was set to the highest value in the screening and a logarithmic scale was chosen to adjust for the over-representation in the liver.

As already indicated above, most of the screened AAV variants showed a pronounced liver tropism. Nevertheless, highly interesting differences between the serotypes could be observed. Derivatives of AAV7, AAV8, AAV9 and AAVrh10 generally demonstrated a broad transcriptional activity whereas members of the AAV2 and AAV3b family largely remained inactive in all tissues. AAV4wt and related peptide insertions predominantly showed up in the blood cells and the lung, but avoided the liver almost entirely. An as-of-yet unknown tropism could be found for AAVpo1wt and AAVpo1_A1. Both capsids were detargeted from the liver but were transcriptionally active in muscle, especially in the diaphragm and the quadriceps femoris.

Aside from the naturally occurring serotypes and their peptide-modified derivatives, some of the published benchmarks gave remarkable results. For instance, the peptide insertion variant AAV2_L1[208] displayed a significantly higher efficiency than its parental virus AAV2wt (Figure 10). Moreover, its activity was limited mainly to the lung and, to a lesser extent, to the brain and blood cells. Another AAV2 peptide-displaying mutant, AAV2_BR1[218], showed strong specificity for the lung and the brain. The latter was even more specifically targeted by the AAV9-based peptide insertion variant AAV9_PHP.B[219], which was restricted to the brain. Of note, the well-known chimeric capsid AAVDJ[172] was confirmed as a highly specific liver-targeting vector, as it barely showed any activity in off-targets. Surprisingly, the P1 peptide-presenting variant AAV9_P1 - previously identified as lead candidate in cultured human astrocytes[202] - could be detected mainly in the screened muscle tissues, namely, the diaphragm, quadriceps femoris and heart (Figure 10). Next to this noticeable improvement in muscle specificity as compared to AAV9wt, an increase in efficiency was observable.

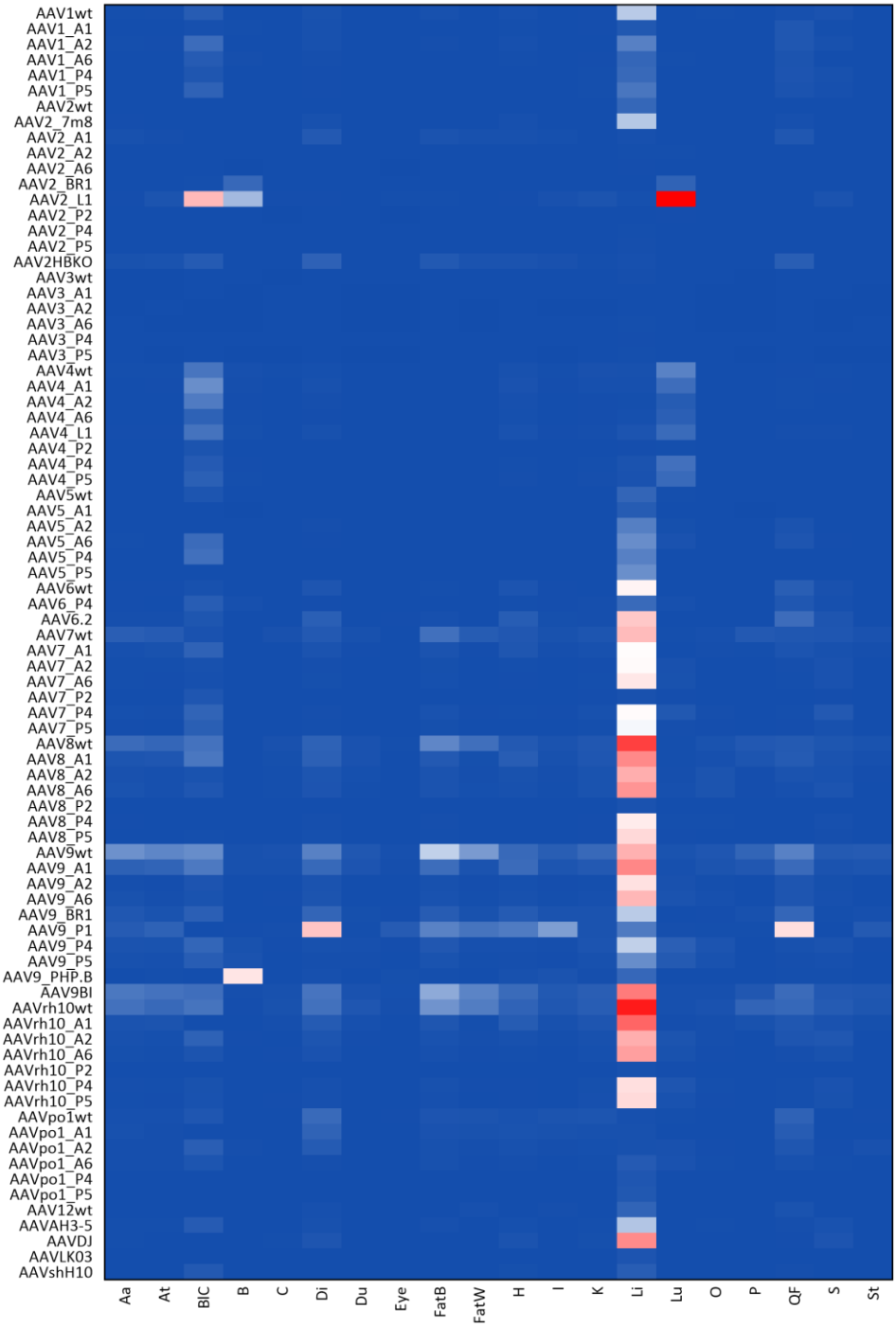

Figure 10: Biodistribution of all variants of the 2nd generation library
Calculated $B_{\alpha\beta}$ values are depicted as a heat map simultaneously illustrating the transcriptional efficiency and specificity of all variants in the 2nd generation library in the abdominal aorta (Aa), thoracic aorta (At), blood cells (BlC), brain (B), colon (C), diaphragm (Di), duodenum (Du), eye, brown fat (FatB), white fat (FatW), heart (H), inner ear (I), kidney (K), liver (Li), lung (Lu), ovaries (O), pancreas (P), quadriceps femoris (QF), spleen (S) and stomach (St). A logarithmic scale is used with blue representing the value 0, white 0.47 and red 4.74.

Although the heat map provides a comprehensive overview of the full dataset, subtle differences are difficult to spot. Therefore, to better illustrate such details, a bar plot depiction of important highlights was chosen. This either illustrates the efficiency of all variants within one organ, referred to as $V_{\alpha\beta}$ values (Figure 11 and Figure 12), or the specificity of one variant across all tissues, termed $T_{\alpha\beta}$ values (Figure 13, Figure 14 and Figure 15). A detailed description of how these values were generated is found in 3.4.13.

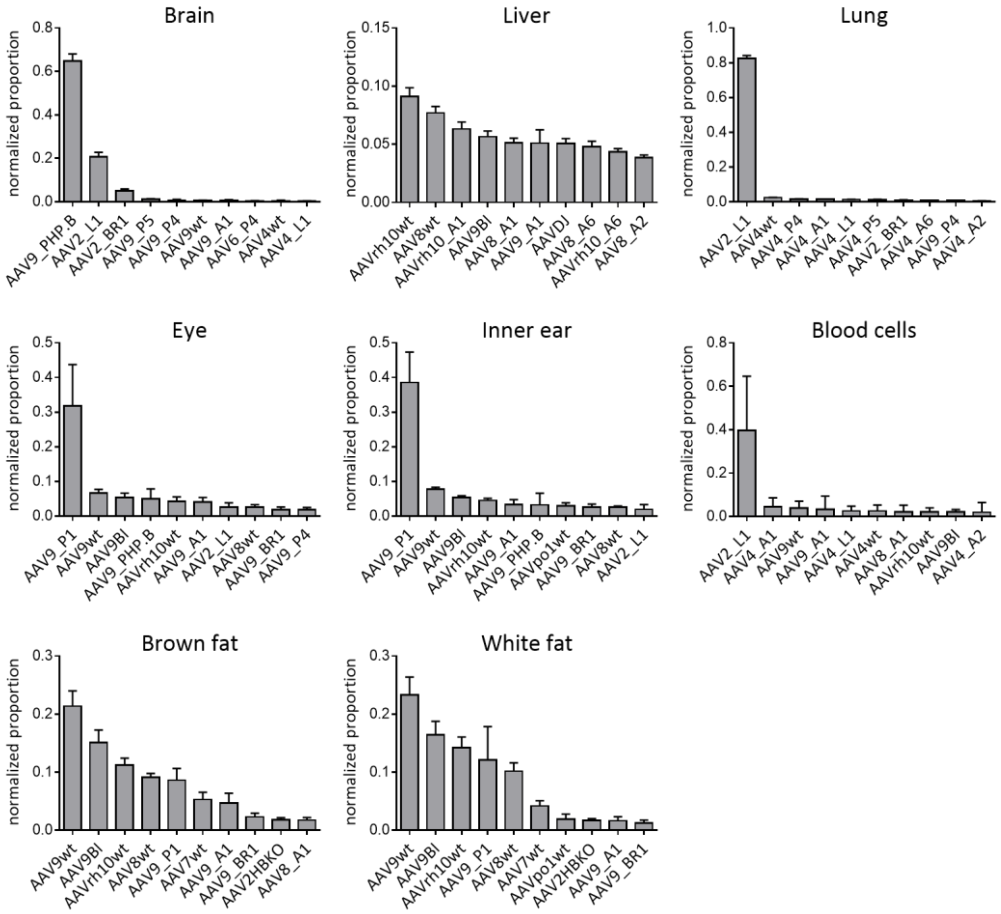

Figure 11: Transcriptional efficiency in various tissues
The depicted bar plots show the transcriptional efficiency as normalized proportion of the top 10 AAV variants from the 2nd generation library in the brain, liver, lung, eye, inner ear, blood cells, brown fat tissue and white fat tissue. The cDNA values are the average from six C57BL/6J mice with SD.

In Figure 11, $V_{\alpha\beta}$ values of various organs are shown to illustrate the efficiency of single AAV variants within the same tissue. As already visible in the heat map, AAV9_PHP.B showed evidence for robust brain activity as over 60% of all barcodes detected in this tissue belong to this capsid variant. Two AAV2wt-based variants, AAV2_L1 and AAV2_BR1, ranked second and third, respectively; however, the former was 3.2-fold and the latter 12.8-fold less efficient than AAV9_PHP.B. Strikingly, all three variants outperformed AAV9wt, which is widely used for passing the blood brain barrier and

robustly transducing the brain. The peptide insertion mutants AAV9_P5 and AAV9_P4 once more demonstrated to be the most efficient in the brain of all vectors in the 1st generation library, replicating the result of the first screening (Supplementary information, Figure 30).

Furthermore, AAVrh10wt and AAV8wt gave the highest normalized values in the liver (Figure 11), mirroring the effects observed for the 1st generation library in this tissue (Figure 7). Interestingly, AAVDJ, the chimera previously selected in hepatocytes, was 1.8-fold less abundant than the top hit AAVrh10wt.

The collection of AAV4 capsids again exhibited a strong lung affinity, as evidenced by the fact that they occupied 7 out of the top 10 spots. Nevertheless, AAV2_L1 clearly outcompeted the AAV4 variants by at least 34.3-fold, representing 82% of all capsids in the lung tissue. Worth noting is that capsids that appeared in the lung were also mostly present in blood cells.

A new addition in the second screening, AAV9_P1, was the most efficient capsid in the eye and in the inner ear with 31% and 38% of all hits, respectively, after systemic injection (Figure 11). This particular virus could also be found in the brown and white fat tissue, albeit it did not reach the top 3. Most impressively, AAV9_P1 was the lead candidate in the diaphragm, quadriceps femoris and the heart, overtaking the gold standard for muscle transduction, AAV9wt, by 10.6-fold, 7.2-fold and 1.5-fold, respectively (Figure 12). The promising vectors from the first screening, AAVpo1wt and AAVpo1_A1 further proved their muscle efficiency in diaphragm and quadriceps femoris but were clearly inferior to AAV9_P1.

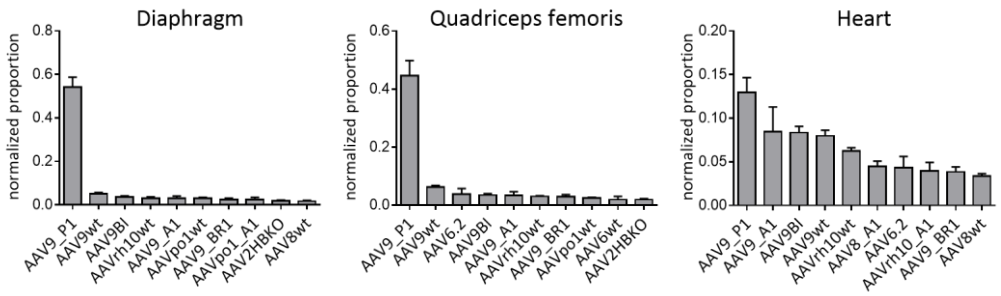

Figure 12: Transcriptional efficiency in muscle tissues
The depicted bar plots show the transcriptional efficiency as normalized proportion of the top 10 AAV variants from the 2nd generation library in the diaphragm, quadriceps femoris and heart. The cDNA values are the average from six C57BL/6J mice with SD.

Another possibility to interpret the screening data is to use the $B_{\alpha\beta}$ values in order to calculate the proportion of one variant in each tissue, termed $T_{\alpha\beta}$ value. Selected highlights of this analysis are shown in Figure 13, Figure 14 and

Figure 15. Importantly, these specificity values ($T_{\alpha\beta}$ values) cannot be directly compared to the efficiency values ($V_{\alpha\beta}$ values) shown above and are therefore herein described separately.

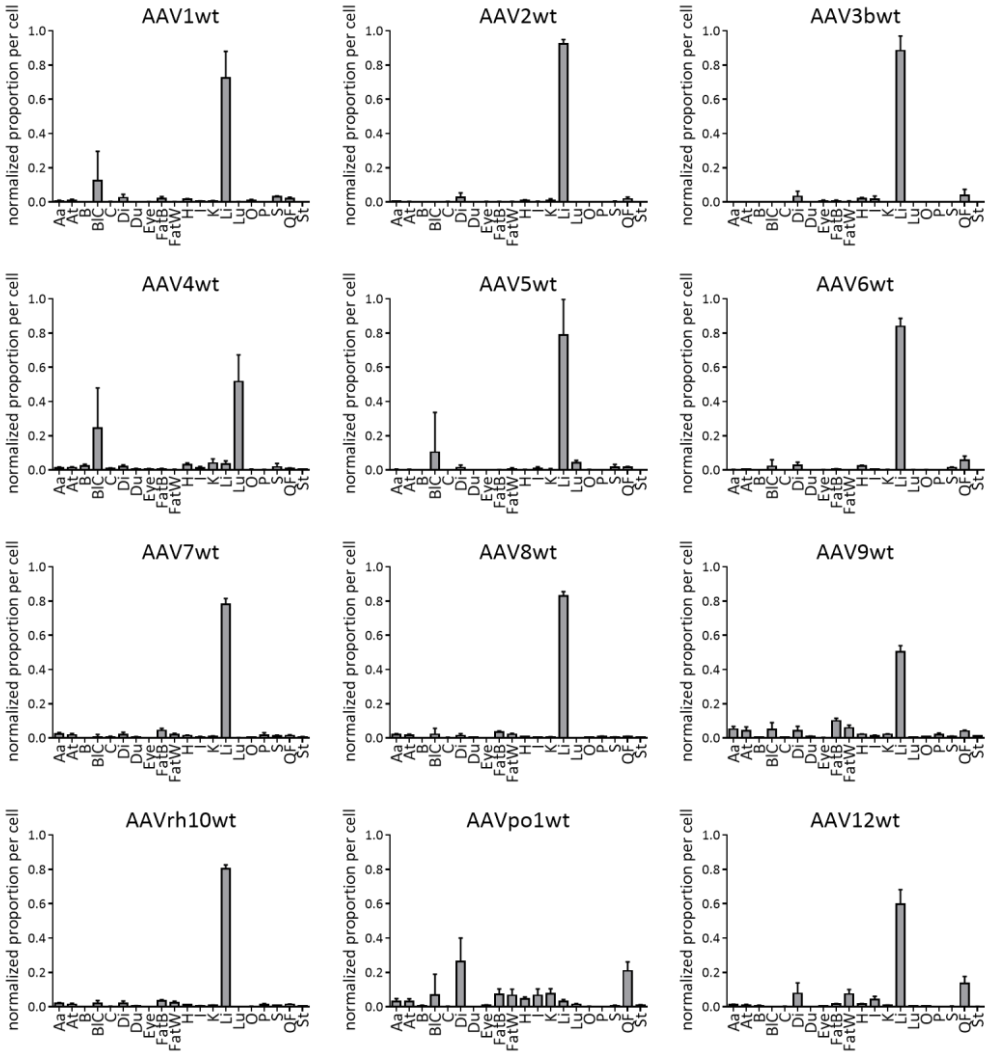

Figure 13: Transcriptional specificity of common AAV serotypes
The depicted bar plots show the transcriptional specificity as normalized proportion per cell of common serotypes from the 2nd generation library in abdominal aorta (Aa), thoracic aorta (At), brain (B), blood cells (BlC), colon (C), diaphragm (Di), eye, brown fat (FatB), white fat (FatW), heart (H), inner ear (I), kidney (K), liver (Li), lung (Lu), ovaries (O), pancreas (P), spleen (S), quadriceps femoris (QF) and stomach (St). Depicted is the average of cDNA values from six C57BL/6J mice with SD.

Analysis of the specificity of wild type AAVs revealed a pronounced bias towards the liver, which typically harbored over 80% of the respective virus. Exceptions were AAV4wt, AAV9wt and AAVpo1wt. Instead of targeting the liver, AAV4wt ended up predominantly in the lung (51%) and to a lesser extent in blood cells (24%). The first barcode screening had already implied a muscle-tropic behavior of AAVpo1wt (Figure 7). The improved normalization strategy could now verify these data by revealing a 52% proportion of

AAVpo1wt in the three muscle tissues diaphragm, quadriceps femoris and heart, with off-targeting mainly to the brown and white fat tissue, inner ear and kidney (Figure 13). Of note, AAV9wt exhibited the broadest activity of all 82 candidates in this screening and, based on the $V_{\alpha\beta}$ values, also the highest efficiency in the majority of the organs (Figure 11 and Figure 33). However, most of the capsid still ended up in the liver (50%) after tail vein injection (Figure 13).

Published synthetic AAV capsids included in this screening round offered the possibility to validate the robustness of the pipeline by attempting to reproduce data from the literature. For instance, in Figure 14, AAVDJ[172] showed a high specificity for the liver (97%) with negligible off-targeting to the diaphragm (1%) and spleen (0.7%), further improving on the already liver-tropic competitor AAV8wt (Figure 13). A more recently published chimera that was selected for human hepatocyte transduction, AAVLK03[191], demonstrated a 92% proportion in the murine liver (Figure 14) but was 200-fold less efficient than AAV8wt (data not shown). AAV6.2[165], deviating in only one amino acid from AAV6wt, behaved identical to its unmodified wild type parent concerning specificity.

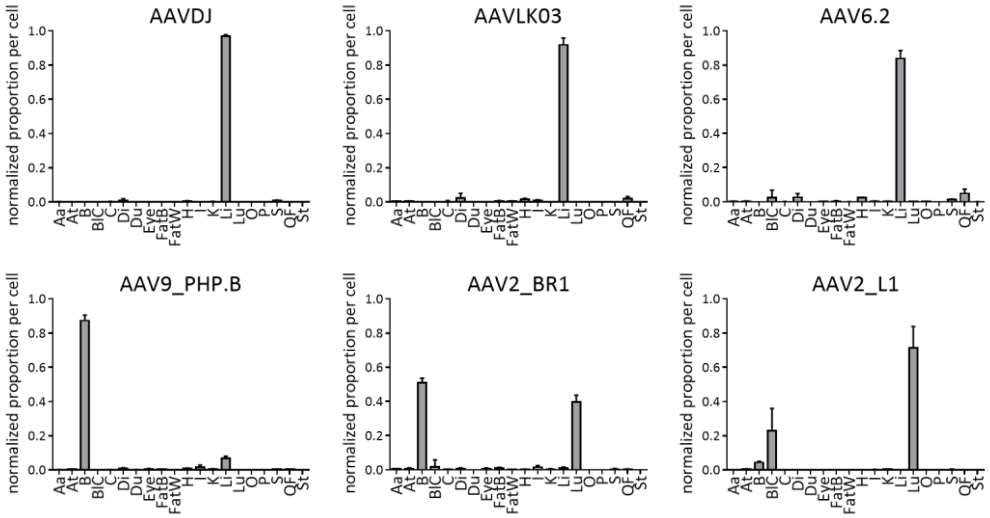

Figure 14: Transcriptional specificity of published AAV variants
The depicted bar plots show the transcriptional specificity as normalized proportion per cell of published AAV variants from the 2nd generation library in abdominal aorta (Aa), thoracic aorta (At), brain (B), blood cells (BlC), colon (C), diaphragm (Di), eye, brown fat (FatB), white fat (FatW), heart (H), inner ear (I), kidney (K), liver (Li), lung (Lu), ovaries (O), pancreas (P), spleen (S), quadriceps femoris (QF) and stomach (St). Depicted is the average of cDNA values from six C57BL/6J mice with SD.

Remarkably, AAV9_PHP.B[219] not only proved to be a highly efficient capsid but furthermore excels in targeting the brain tissue (87%), with minor transcriptional activity in the liver (6.7%). AAV2_BR1, a peptide-displaying

variant selected for the brain[218], could not match AAV9_PHP.B since 51% of the AAV2_BR1 transcripts were detected in the brain and 40% in the lung (Figure 14).

The highest specificity for the lung was observed for the selected AAV2_L1[208] capsid, where it made up 71% of all hits. Of the remaining 29%, 23% were found in blood cells as the major off-target and 4% of the capsid ended up in the brain. When comparing this synthetic capsid to the lung-tropic AAV4wt (Figure 13), superior efficiency (Figure 11) and specificity (Figure 14) were observed for the peptide insertion variant, exemplifying the power of directed evolution approaches. It was thus tempting to test whether additional improvements could be achieved with an AAV4-based capsid displaying the lung-tropic L1 peptide. The resulting rationally designed variant was called AAV4_L1, and the corresponding results on specificity can be seen in Figure 15.

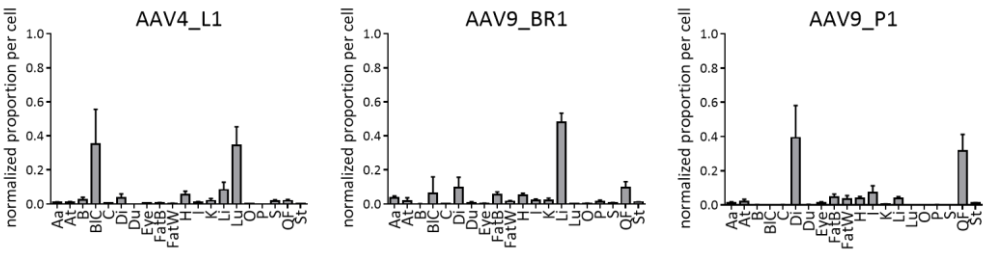

Figure 15: Transcriptional specificity of novel AAV variants
The depicted bar plots show the transcriptional specificity as normalized proportion per cell of novel AAV variants from the 2nd generation library in abdominal aorta (Aa), thoracic aorta (At), brain (B), blood cells (BlC), colon (C), diaphragm (Di), eye, brown fat (FatB), white fat (FatW), heart (H), inner ear (I), kidney (K), liver (Li), lung (Lu), ovaries (O), pancreas (P), spleen (S), quadriceps femoris (QF) and stomach (St). Depicted is the average of cDNA values from six C57BL/6J mice with SD.

Curiously, AAV4_L1 showed no improvement in vector specificity; on the contrary, lung-targeting even decreased compared to AAV4wt. Additionally, the efficiency was similar to that observed for the other AAV4 peptide insertion variants (Figure 11).

This rational approach was further applied by integrating the BR1 peptide into the most efficient wild type capsid for brain transduction, AAV9wt. However, the newly generated variant AAV9_BR1 was unable to selectively target the brain; instead, it was mainly active in the liver (Figure 15).

Most notably, AAV9_P1 showed a marked increase in muscle specificity with a cumulated proportion of 75% in the three muscle tissues diaphragm, quadriceps femoris and heart. Identified off-targets were brown and white fat tissue, inner ear and the liver (Figure 15).

4.5 3RD GENERATION LIBRARY SCREENING

For the third barcode-based variant screening, remains of the 2nd generation library were enriched with 64 chimeric capsids that had been independently generated, *in vivo* selected and pre-validated by two other members of the Grimm laboratory. Thirty of them were selected in stellate cells by Anne-Kathrin Herrmann and the other 34 in different muscle tissues by Jihad El Andari. The latter were especially important for this work since the newly discovered benefits of AAV9_P1 in muscles should be validated against variants isolated from state-of-the-art selection strategies, such as DNA family shuffling that was used by the other two group members. Furthermore, next to a variety of additional published benchmarks, the most promising muscle-tropic capsids from the literature were added, namely AAVM41[189], AAVB1[190] and AAV2_MTP[201]. Thus, opportunities for a fair comparison to well-established capsids were provided. On top, two more peptide insertion variants were added, AAV9_P3 and AAV9_K3[215], comprising a peptide motif that is highly similar to P1 and thereby potentially helping to elucidate the role of the peptide itself in determining capsid tropism. Finally, to study brain transduction, the successors of AAV9_PHP.B[219] were spiked in, referred to as AAV9_PHP.A[219], AAV9_PHP.eB[220] and AAV9_PHP.S[220] (Table 28).

All 75 additional variants were individually produced and purified, including the 64 extra capsids that were selected by the two colleagues (see above) and produced by them. After virus titration, equimolar amounts were pooled to create a preliminary library. This library was subsequently titrated together with the 2nd generation library. Based on the number of AAV variants in the respective libraries, molar shares for the final pooling were calculated.to end up with equimolar shares for each vector in the resulting 3rd generation library. Afterwards, the mixture was concentrated and dialyzed using an Amicon Ultra-15 tube.

As before, the library composition was assessed by NGS (Supplementary information, Figure 36). As compared to the 2nd generation library, the imbalance only marginally increased, as evidenced by a 7.4-fold deviation to the theoretical mean proportion for one of the newly introduced chimeras, AAVJEA3-H4. Importantly, small composition imbalances in this range can readily be corrected for during the multi-step normalization procedure.

For the *in vivo* screening, 1.57×10^{12} vg/mouse were injected i.v. into six female C57BL/6J mice. After one week, aorta, biceps, colon, diaphragm, duodenum, eye, brown fat tissue, white fat tissue, heart, inner ear, kidney,

liver, lung, ovaries, pancreas, quadriceps femoris and stomach were harvested. Additionally, lymph nodes and the spleen were extracted for subsequent isolation of CD3-, CD19-, CD11b- and CD11c-positive cells by MACS (in collaboration with Martin Busch). The brain was further dissected into the subventricular zone (SVZ) and the cortex (in collaboration with Sascha Dehler). From the SVZ, neural stem cells (NSC), neuroblasts, astrocytes and oligodendrocytes were extracted via FACS. Astrocytes and oligodendrocytes were collected from the cortex.

In parallel, four BALB/c mice were injected via the tail vein with 1.57×10^{12} vg/mouse of the same 3rd generation library. From these mice, hepatocytes, stellate cells, Kupffer cells and liver sinusoidal endothelial cells (LSECs) were isolated by MACS after perfusing the liver (in collaboration with the Dooley laboratory and Anne-Kathrin Herrmann). DNA and RNA were extracted, and qPCR-based determination of the viral genomes per diploid cell was performed for all samples except for the brain cells where only RNA could be collected. The distribution of the 3rd generation library in the comprehensive tissue collection from the C57BL/6J mice and the liver cells of the second mouse experiment can be seen in Figure 16.

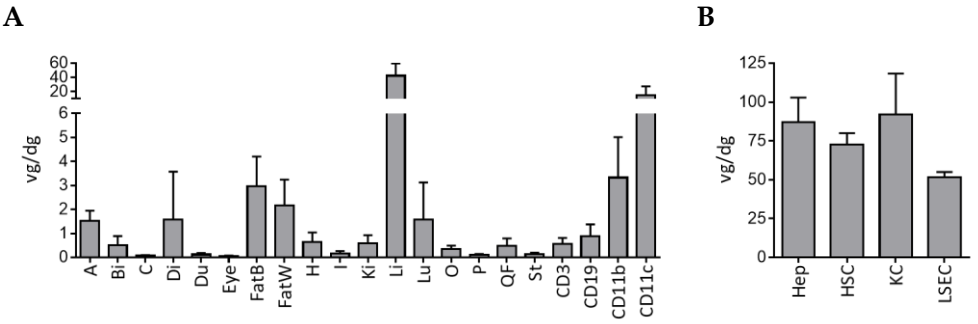

Figure 16: Viral DNA distribution of the 3rd generation library
(A) The depicted bar plot shows the viral DNA distribution from the 3rd generation library after systemic injection into six C57BL/6J mice across aorta (A), biceps (Bi), colon (C), diaphragm (Di), duodenum (Du), eye, brown fat (FatB), white fat (FatW), heart (H), inner ear (I), kidney (K), liver (Li), lung (Lu), ovaries (O), pancreas (P), quadriceps femoris (QF), stomach (St) and CD3-, CD19-, CD11b- as well as CD11c-positive cells. **(B)** Shows the distribution in the liver of four BALB/c mice across hepatocytes (Hep), stellate cells (HSC), Kupffer cells (KC) and liver sinusoidal endothelial cells (LSEC). Detected viral genomes (EYFP probe) were normalized to GAPDH as a housekeeper. Depicted values represent the average of the mice with SD.

As previously observed for the second variant screening (Figure 9), the majority of AAV particles ended up in the liver (42 vg/dg) followed by CD11c cells (14 vg/dg), CD11b cells (3.3 vg/dg), brown fat tissue (3 vg/dg) and white fat tissue (2.2 vg/dg). Aorta, biceps, diaphragm, heart, kidney, lung, ovaries, quadriceps femoris, CD3 and CD19 cells ranged between 0.35 and 1.6 vg/dg.

The digestive tract including colon, duodenum and stomach as well as the eye, inner ear and pancreas could only be weakly transduced (0.05-0.16 vg/dg). Concerning the transduction of liver cell types in the separate mouse study (Figure 16B), hepatocytes, stellate cells and Kupffer cells demonstrated similar levels, while LSECs were slightly lacking behind.

After processing the NGS data and analyzing the output files, C57BL/6J mouse numbers 3 and 4 were declared to be outliers due to unusually low AAV9_P1 abundance and therefore excluded from the analysis. The $V_{\alpha\beta}$ and $T_{\alpha\beta}$ values that are shown below hence depict the averages of mouse 1, 2, 5 and 6 with the corresponding SD. Figure 17 shows the efficiency of the top 10 AAV variants in the liver, lung, eye, inner ear, brown and white fat tissue.

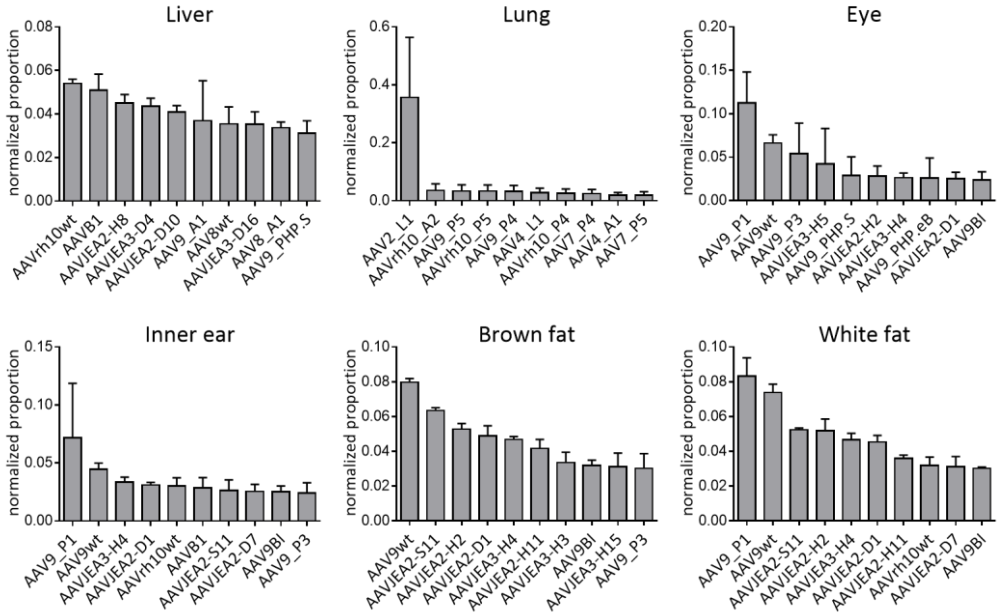

Figure 17: Transcriptional efficiency in various tissues
The depicted bar plots show the transcriptional efficiency as normalized proportion of the top 10 AAV variants from the 3rd generation library in the liver, lung, eye, inner ear, brown fat tissue and white fat tissue. The cDNA values are the average from four C57BL/6J mice with SD.

A first notable result was that AAVrh10wt outcompeted the other capsids in the library in the liver, reproducing the results from the first (Figure 7) and second (Figure 11) screening. AAV8wt was found in position 7, mostly separated from AAVrh10wt by capsids that were newly added in the third screening round (Table 28). One of those, AAVB1, was originally reported to excel in brain, muscle and pancreas[190] but turned out to be highly transcriptionally active in the liver. In the lung, the highly promising capsid AAV2_L1[208] was confirmed as lead candidate displacing the rationally designed peptide insertion variants.

From the 2nd generation library, AAV9_P1 has emerged as an efficient vector in the muscle tissues (Figure 12) and the off-targets eye, inner ear and fat tissue (Figure 11). Figure 17 partially exemplifies this phenomenon again by verifying AAV9_P1 as the most efficient vector in the eye and inner ear. Unlike what was observed in the second screen, the P1-displaying variant even marginally overtook AAV9wt in the white fat tissue.

More importantly, AAV9_P1 once more showed a superior efficiency compared to AAV9wt in the diaphragm, biceps, quadriceps femoris and heart where it outperformed its parent by 10.1-fold, 7.2-fold, 5.6-fold and 1.6-fold, respectively (Figure 18). Surprisingly, none of the published muscle benchmarks was able to reach the top 10 in any of the muscle tissues. Moreover, several of the newly generated, shuffled chimeras selected in these tissues were found in the top 10 albeit they remained below AAV9wt. Finally, an additional peptide insertion mutant from our laboratory, AAV9_P3, scored second to AAV9_P1 in the diaphragm, biceps and quadriceps femoris where it was 3- to 6-fold less efficient depending on the organ.

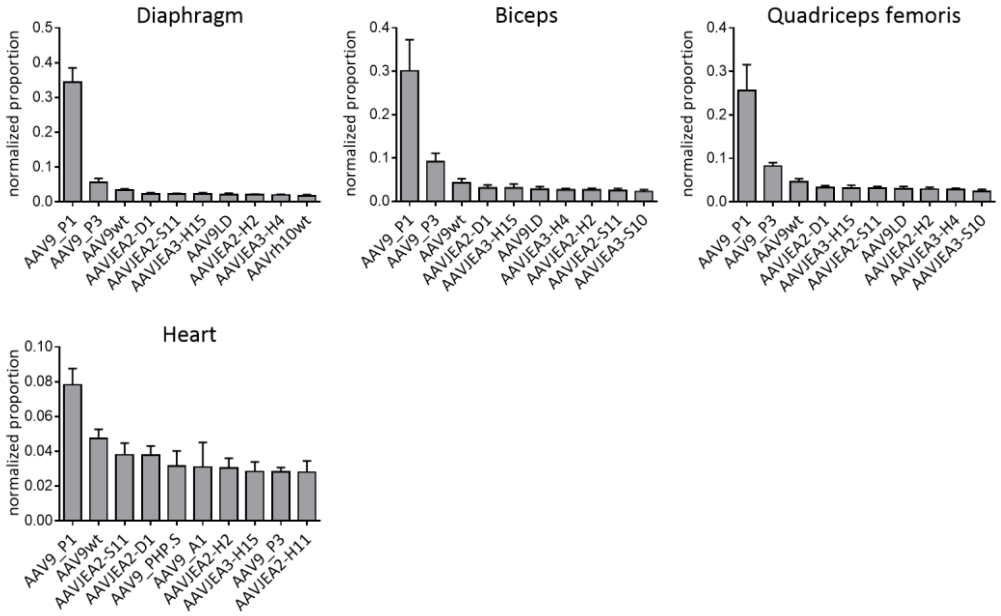

Figure 18: Transcriptional efficiency in muscle tissues
The depicted bar plots show the transcriptional efficiency as normalized proportion of the top 10 AAV variants from the 3rd generation library in the diaphragm, biceps, quadriceps femoris and heart. The cDNA values are the average from four C57BL/6J mice with SD.

Another promising feature of the 3rd generation library was the presence of three additional brain-targeting variants next to AAV2_BR1 and AAV9_PHP.B, offering the opportunity for a head-to-head comparison in the clinically highly relevant brain tissue. However, it has to be noted that the flow cytometry sorting of cells from this tissue was difficult, ultimately resulting in an

incomplete recovery of the samples and low cellular yields ranging from 330 to 941 cells depending on the fraction. Thus, the data must be interpreted with caution. Nonetheless, a trend towards AAV9_A2 could be observed (Supplementary information, Figure 39). Intriguingly, none of the capsids suggested by the literature appeared in the top 10.

In the second mouse experiment utilizing the third barcoded library, the liver of four BALB/c mice was dissected into hepatocytes, Kupffer cells, stellate cells and LSECs in collaboration with the Dooley laboratory in Mannheim and Anne-Kathrin Herrmann. Samples of extracted DNA and RNA were run through the established pipeline and normalized as previously described (3.4.13). As before, $V_{\alpha\beta}$ values were averaged across the mice and are depicted with the corresponding SD in Figure 19.

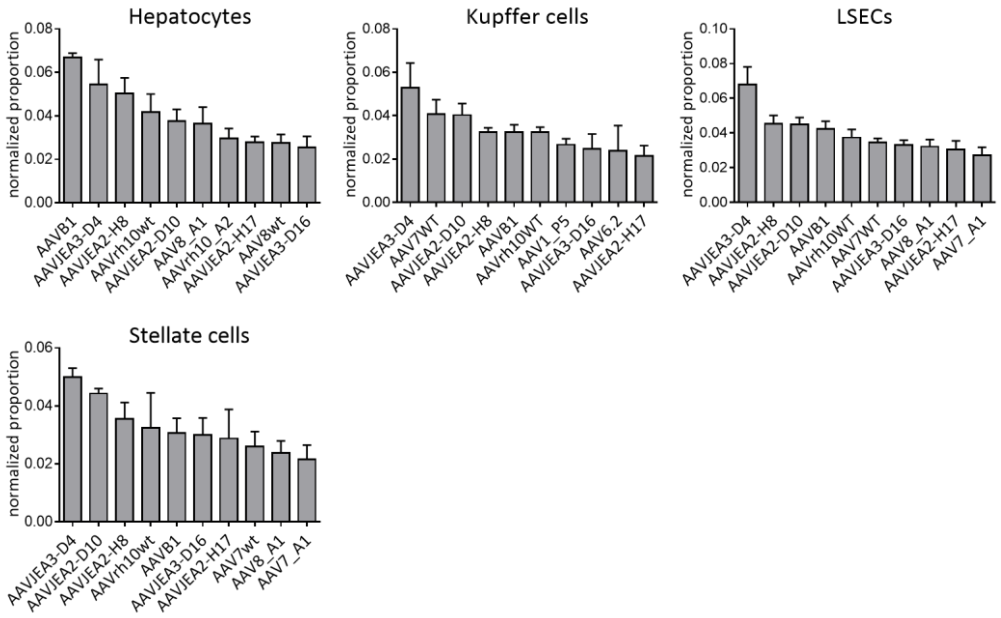

Figure 19: Transcriptional efficiency in liver cell types
The depicted bar plots show the transcriptional efficiency as normalized proportion of the top 10 AAV variants from the 3rd generation library in hepatocytes, Kupffer cells, liver sinusoidal endothelial cells (LSECs) and stellate cells. The cDNA values are the average from four BALB/c mice with SD.

The data from the whole liver (Figure 17) had already indicated a pronounced liver activity of AAVB1[190]. Analysis of the sub-cell types revealed the highest AAVB1 efficiency in hepatocytes where it outperformed all other variants. In addition, the capsid was detected in the top 10 of the remaining three cell types. In Kupffer cells, an interesting effect was noted, namely the appearance of variants that were generally less efficient, such as AAV1_P5 and AAV6.2. Curiously, none of the chimeras pre-selected for stellate cells was preferentially detected in these cells in this screen. More information on these

capsids and a more detailed description and discussion of these data is found in the doctoral thesis of Anne-Kathrin Herrmann.

The possibility to analyze the liver as a whole organ or divided into cell types massively enhances the understanding of the function of particular capsid variants. Accordingly, for capsids that were less frequently found in hepatocytes, their specificity was assessed by studying the corresponding $T_{\alpha\beta}$ values in all 21 tissues (Figure 20).

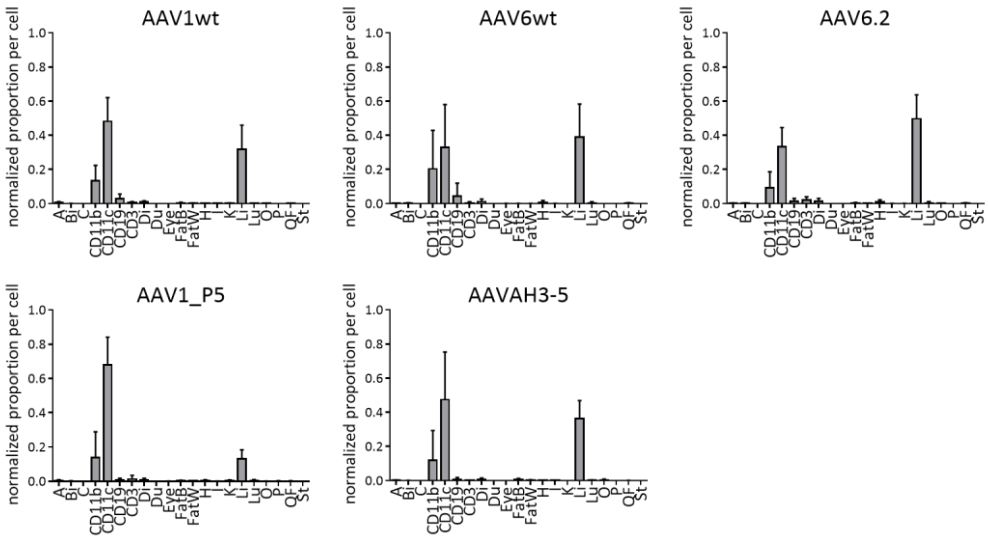

Figure 20: Transcriptional specificity of hepatocyte-detargeted variants
The depicted bar plots show the transcriptional specificity as normalized proportion per cell of hepatocyte-detargeted AAV variants from the 3rd generation library in aorta (A), biceps (Bi), colon (C), CD11b-, CD11c-, CD19-, CD3-positive cells, diaphragm (Di), duodenum (Du), eye, brown fat tissue (FatB), white fat tissue (FatW), heart (H), inner ear (I), kidney (K), liver (Li), lung (Lu), ovaries (O), pancreas (P), quadriceps femoris (QF) and stomach (St). Depicted is the average of cDNA values from four C57BL/6J mice with SD.

The selected candidates AAV1wt, AAV6wt, AAV6.2, AAV1_P5 and AAVAH3-5 demonstrated a highly similar tropism, by almost exclusively targeting the liver and CD11b- as well as CD11c-positive cells. Strikingly, three different capsid engineering approaches achieved the same result, namely DNA family shuffling with AAVAH3-5, peptide insertion with AAV1_P5 and introduction of single point mutations with AAV6.2. Even the very homologous, naturally occurring isolates AAV1wt and AAV6wt exhibited identical specificity patterns. Still, out of the five capsids, AAV1_P5 showed the most pronounced immune cell-targeting while its activity in the liver was limited to fewer than 20%.

Concerning the specificity within the liver tissue, i.e., information provided by the second study in BALB/c mice, a marked hepatocyte-detargeting could be observed for all five variants (Figure 21).

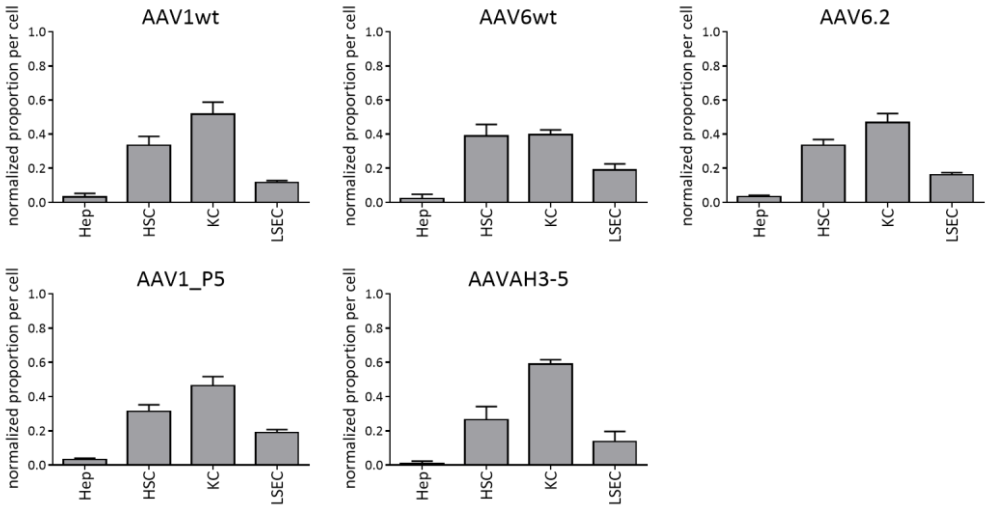

Figure 21: Transcriptional specificity of hepatocyte-detargeted variants
The depicted bar plots show the transcriptional specificity as normalized proportion per cell of hepatocyte-detargeted AAV variants from the 3rd generation library in hepatocytes (Hep), hepatic stellate cells (HSC), Kupffer cells (KC) and LSECs. Depicted is the average of cDNA values from four BALB/c mice with SD.

In more detail, the chimera AAVAH3-5 detargeted hepatocytes almost entirely (0.8%), followed by the other four variants with normalized proportions of roughly 3%. None of the selected vectors could discriminate between stellate cells, Kupffer cells or LSECs, including the shuffled chimera selected in stellate cells, AAVAH3-5. From the three mentioned cell types, the mentioned vectors could be predominately found in Kupffer cells with over 40% followed by stellate cells (~30%) and LSECs (~15%). Solely based on the specificity, no clear lead candidate emerged. However, AAV1_P5 was the most efficient in stellate cells where it marginally surpassed AAV6.2, AAV6wt, AAVAH3-5 and AAV1wt by 1.05-fold, 1.27-fold, 2.18-fold and 2.38-fold, respectively (corresponding $V_{\alpha\beta}$ values are not shown).

In the full organ biodistribution, a remarkable observation had been that AAV2_L1 demonstrated the highest specificity observed in all three screenings (Figure 22). 99% of the vector activity was measured in the lung, which is even higher than the 71% proportion found in the second screening round (Figure 14). This enhancement can most likely be explained by the fact that for the 3rd generation library, the two major off-targets, blood cells and brain, were not included in the analysis. This clearly exemplifies that determination of capsid specificity strongly depends on the investigated organs.

Vice versa, an example that implementing certain tissues can also decrease the tropism is shown for AAVDJ (Figure 22). The dataset of the 2nd generation library had revealed a very pronounced 97% specificity of this capsid for the liver (Figure 14). However, the most recent screening showed additional AAVDJ activity in CD11b- and CD11c-positive cells, lowering its value in the liver to 58%. Anne-Kathrin Herrmann made further modifications to the AAVDJ capsid by mutating three tyrosine residues to phenylalanines (AAVDJYF), hoping to evade proteasomal degradation of the particles. Nevertheless, this did not result in improved efficiency in the whole liver or enhanced specificity.

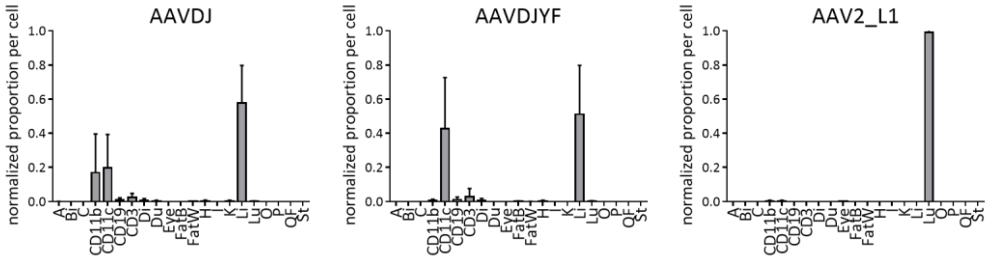

Figure 22: Transcriptional specificity of AAVDJ and AAV2_L1
The depicted bar plots show the transcriptional specificity as normalized proportion per cell of AAVDJ, AAVDJYF and AAV2_L1 from the 3rd generation library in aorta (A), biceps (Bi), colon (C), CD11b-, CD11c-, CD19-, CD3-positive cells, diaphragm (Di), duodenum (Du), eye, brown fat tissue (FatB), white fat tissue (FatW), heart (H), inner ear (I), kidney (K), liver (Li), lung (Lu), ovaries (O), pancreas (P), quadriceps femoris (QF) and stomach (St). Depicted is the average of cDNA values from four C57BL/6J mice with SD.

Within the liver, the chimera selected in hepatocytes, AAVDJ, was indeed found predominantly in hepatocytes (87%) followed by stellate cells (6.9%), Kupffer cells (4.3%) and LSECs (1.3%) (Figure 23). Also here, the mutations introduced in AAVDJYF did not alter capsid selectivity for the on-target. Worth noting are two peptide-modified variants, AAV5_P5 and AAV8_P5, that exhibited similar or even enhanced hepatocyte activity with 95% and 85%, respectively. Still, these vectors are slightly inferior to AAVDJ since their efficiency is 14-fold and 2.2-fold lower in whole liver, based on the data from the second screening. Finally, AAV8wt and AAVrh10wt, that were the most effective capsids in the liver on the cDNA level (Figure 11), showed a broad distribution in the four sub-cell types with minor preference to hepatocytes (Figure 23).

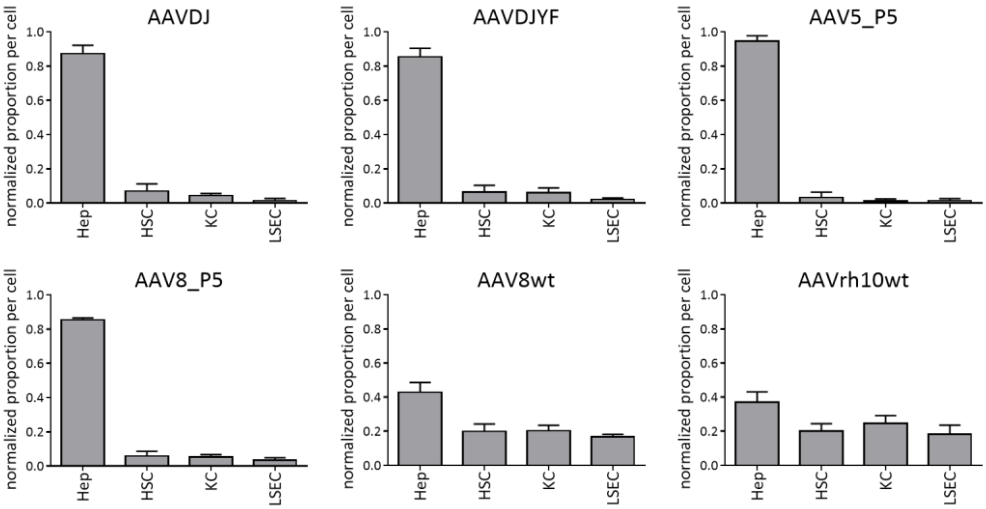

Figure 23: Transcriptional specificity for hepatocytes
The depicted bar plots show the transcriptional specificity as normalized proportion per cell of hepatocyte-targeting AAV variants from the 3rd generation library in hepatocytes (Hep), hepatic stellate cells (HSC), Kupffer cells (KC) and LSECs. Depicted is the average of cDNA values from four BALB/c mice with SD.

Next to liver, muscle is one of the preferred organs for gene therapy applications. The results from the 2nd generation library had already showed a massively improved targeting of muscle for AAV9_P1 (Figure 15). The 3rd generation library comprised important benchmarks from the literature, allowing for an extensive comparison with this lead candidate. The specificity values ($T_{\alpha\beta}$ values) of the mentioned benchmarks are depicted in Figure 24.

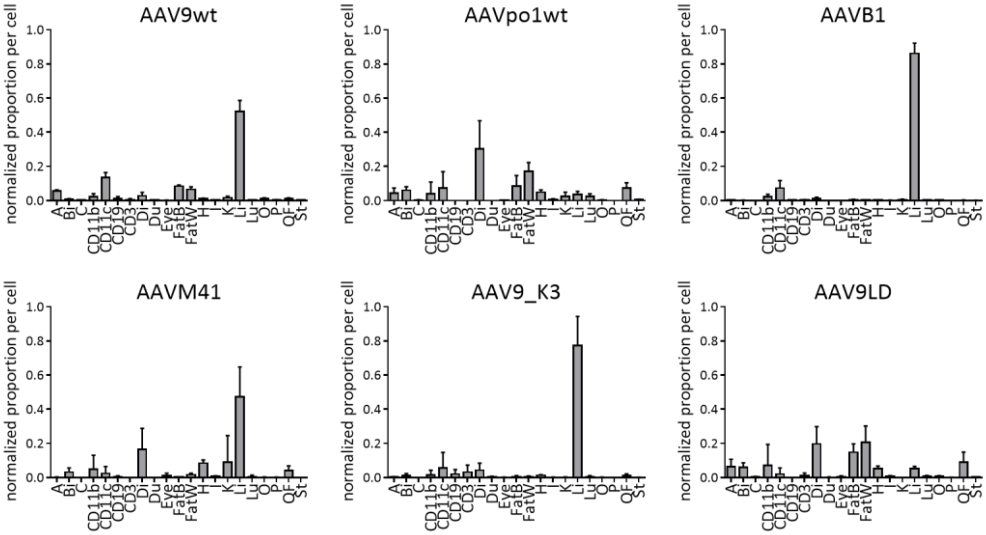

Figure 24: Transcriptional specificity of published benchmarks in muscles
The depicted bar plots show the transcriptional specificity as normalized proportion per cell of published muscle-tropic benchmarks from the 3rd generation library in aorta (A), biceps (Bi), colon (C), CD11b-, CD11c-, CD19-, CD3-positive cells, diaphragm (Di), duodenum (Du), eye, brown fat tissue (FatB), white fat tissue (FatW), heart (H), inner ear (I), kidney (K), liver (Li), lung (Lu), ovaries (O), pancreas (P), quadriceps femoris (QF) and stomach (St). Depicted is the average of cDNA values from four C57BL/6J mice with SD.

The widely used and broadly transducing AAV9wt once more particularly targeted the liver (52%), an effect already observed in the second screening (Figure 13). Previously observed, AAVpo1wt exhibited a tendency for transduction of diaphragm and quadriceps femoris (Figure 13). In this third screen, this tropism could be confirmed, with brown and white fat tissue appearing as major off-targets (Figure 24).

AAVB1, a chimera originally selected for the brain[190], was reported to transduce muscle tissues more robustly than AAV9wt, a notion that could not be reproduced here. In fact, AAVB1 was 3-, 3.5-, 7.6- and 9.2-fold less efficient in the heart, diaphragm, quadriceps femoris and biceps, respectively (corresponding $V_{\alpha\beta}$ values are not shown). Regarding specificity, the chimeric AAVB1 capsid was preferentially detected in the liver with 86% (Figure 24).

Another chimera from Yang and colleagues, AAVM41, was isolated after two selection rounds in muscle tissue and showed a trend towards a muscle tropism[189]. Here, 16.5% of the transcriptional activity could be found in the diaphragm, 8.3% in the heart, 3% in the biceps and 4% in the quadriceps femoris. However, the capsid was roughly 10-fold less efficient than AAV9wt.

The peptide-inserted mutant AAV9_K3 was selected in endothelial cells[215] but used in the third screening due to its peptide sequence that deviated from P1 in only two of the nine amino acids. Surprisingly, the variant could not target the muscles. Instead, 77% of the viral activity was measured in the liver.

Two point mutations, P504A and G505A, were introduced to the AAV9wt capsid proteins by Adachi *et al.*, yielding variant AAV9LD[222] that was reported to be liver-detargeted. Indeed, the strong exclusion of the liver could be reproduced in this study by detecting 140-fold less viral transcripts as compared to AAV9wt. Interestingly, this effect resulted in a preferred targeting of muscle and fat tissues as depicted in Figure 24. However, the mutations mildly decreased the efficiency by roughly 1.5-fold in diaphragm, biceps and quadriceps femoris as well as, more prominently, in the heart by 3-fold compared to the parental virus, AAV9wt.

So far, none of the benchmarks could reach similar levels of specificity or efficiency in comparison to AAV9_P1 in the second screening. New in the successive round were chimeric synthetic capsids selected and pre-validated in muscle tissues by Jihad El Andari, including an independent NGS screen. Thirty-four of these shuffled variants were included in the library and assessed for specificity, and the most promising are depicted in Figure 25. All of them showed a convincing detargeting from the liver while increasing the

proportion in muscle tissues, as hoped for. As previously observed for AAVM41, AAV9LD and AAV9_P1 (Figure 24 and Figure 15), off-targeting to the brown and white fat tissue was measured.

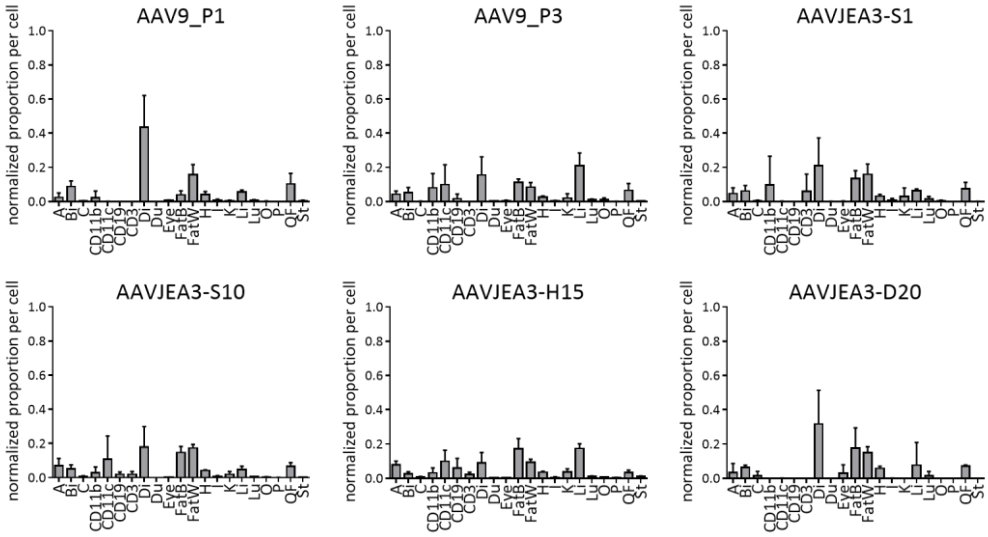

Figure 25: Transcriptional specificity of novel variants in muscle tissues
The depicted bar plots show the transcriptional specificity as normalized proportion per cell of novel muscle-tropic AAV variants from the 3rd generation library in aorta (A), biceps (Bi), colon (C), CD11b-, CD11c-, CD19-, CD3-positive cells, diaphragm (Di), duodenum (Du), eye, brown fat tissue (FatB), white fat tissue (FatW), heart (H), inner ear (I), kidney (K), liver (Li), lung (Lu), ovaries (O), pancreas (P), quadriceps femoris (QF) and stomach (St). Depicted is the average of cDNA values from four C57BL/6J mice with SD.

The lead candidate of the second screening, AAV9_P1, once more exhibited a strong muscle tropism (66%) outcompeting all other variants in the 3rd generation library (Figure 25). In addition, AAV9_P1 was roughly 10-fold more efficient in the on-targets than the best chimeric AAVJEA vectors. Of note, AAV9_P3 also behaved similar to AAV9_P1 in terms of specificity, but the cumulative value for all muscles was lower (30%).

4.6 VALIDATION OF AAV9_P1

As shown above, AAV9_P1 demonstrated compelling evidence for a high efficiency (Figure 12 and Figure 18) and specificity (Figure 15 and Figure 25) in murine muscle tissues. To independently verify this novel and exciting finding, further validation experiments had to be performed. One particularly important consideration was the potential occurrence of capsid interference in a library context, resulting from e.g. receptor competition or particle cross-interactions. Therefore, AAV9_P1 and the previous lead candidates from the first screening, AAVpo1wt and AAVpo1_A1, as well as AAV9wt as a control

were injected individually into three C57BL/6J mice at a dose of $1x10^{11}$ vg/mouse. The other supposedly muscle-tropic vectors from the literature, AAVB1[190], AAVM41[189], AAV9_K3[215], AAV9LD[222] and AAV2_MTP[201], were excluded from this study since none of them matched AAV9_P1 regarding efficiency (Figure 18) and muscle-targeting (Figure 24 and Figure 25). The two AAVpo1-based variants were included due to their roughly 50% proportion in the three muscle tissues (Figure 24). Intravenously injected mice were kept for one week before diaphragm, quadriceps femoris, heart and liver were harvested and analyzed by qPCR to detect the viral transcripts. The *eyfp* transgene signal was subsequently normalized to a POLR2A housekeeper. The values depicted in Figure 26 are *eyfp* relative quantities ($2^{-\Delta Ct}$) for the mentioned AAVs in the respective organs.

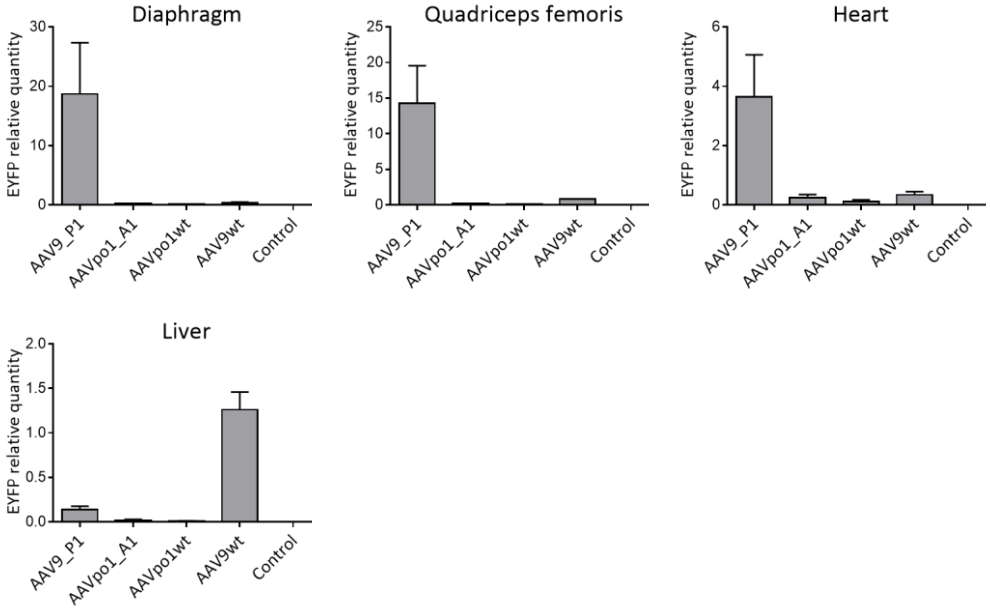

Figure 26: EYFP relative quantities of AAV9_P1
The depicted bar plots show EYFP relative quantities of AAV9_P1, AAVpo1_A1, AAVpo1wt, AAV9wt and an uninjected control mouse in the diaphragm, quadriceps femoris, heart and liver. Relative quantities ($2^{-\Delta Ct}$) were measured by detecting viral EYFP transcripts via qPCR as well as a POLR2A housekeeper. Depicted values are the average of three C57BL/6J mice with SD.

AAV9_P1 exhibited a dramatically improved transcriptional activity in the diaphragm, quadriceps femoris and heart, exceeding its parental capsid AAV9wt by 55-, 17- and 11-fold, respectively. Importantly, the P1-displaying variant was 9-fold less abundant in the major AAV off-target, the liver. Both AAVpo1-based variants showed less activity than AAV9wt, which is in line with the barcode screening data (Figure 12). Notably, AAVpo1wt and AAVpo1_A1 were especially detargeted from the liver, i.e., 126-fold and 63-fold, further improving on the already pronounced effect of AAV9_P1.

To study whether AAV9_P1 would exhibit additional beneficial effects on the protein level, the used AAV genome cassette was slightly modified by replacing the CMV promoter-driven *eyfp* gene with *egfp*. This guaranteed an optimal excitation at 488 nm for a fluorescence-based readout via histology. C57BL/6J mice were i.v. injected with $5x10^{11}$ vg/mouse and kept for two weeks before submerging organ pieces of the diaphragm, quadriceps femoris, biceps, heart and liver into a 4% paraformaldehyde solution for fixation. After an intermediate incubation in 30% sucrose, the samples were embedded and cryosections were generated. Both AAVpo1 variants were excluded from this experiment since their strong liver-detargeting did not outweigh the lower efficiency in the muscle tissues as compared to AAV9_P1. Next to AAV9_P1, AAV9wt and a PBS control, also a newly-cloned variant, AAV9LD_P1, was included in which the two mutations of AAV9LD[222] were introduced into the AAV9_P1 capsid. The expectation was that this rationally designed capsid may display enhanced liver-detargeting while maintaining the prominent activity in the muscle tissues.

During the dissection of the mice, a surprising effect observed for the AAV9_P1 group was that EGFP expression was visible to the naked eye. Images of a representative mouse in dorsal and ventral position are shown in Figure 27.

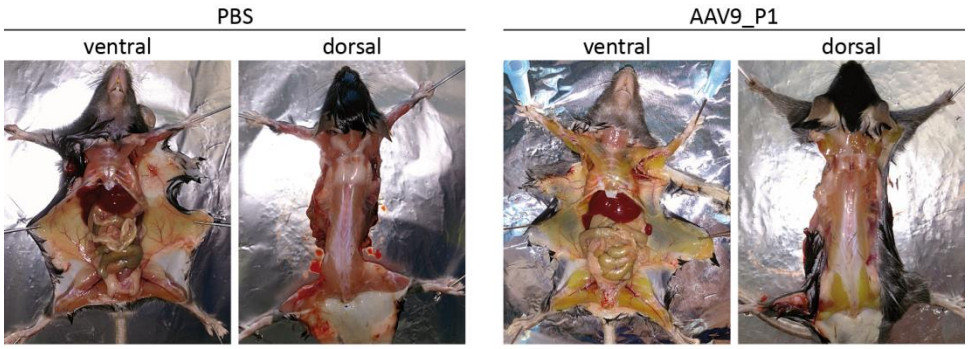

Figure 27: Dissection of an AAV9_P1-injected mouse
Dissection images show a representative C57BL/6J mouse in ventral and dorsal position from the PBS and AAV9_P1 group. Mice were injected i.v. with $5x10^{11}$ vg/mouse and kept for two weeks.

Although the pictures were taken under normal light conditions, a pronounced EGFP signal could be detected in the skeletal muscles of the AAV9_P1-injected mouse. Thus far, the superior activity of AAV9_P1 in the muscles was determined based on cDNA data of the quadriceps femoris and the biceps (Figure 26). However, these images led to the assumption that the capsid behaved equally efficient in the other skeletal muscles.

From the obtained cryosections, representative images were chosen for the liver diaphragm, heart, biceps and quadriceps femoris and assessed directly for EGFP-induced fluorescence signal (Figure 28 and Figure 29). The exposure was normalized to the highest signal in this experiment, i.e., the diaphragm of the AAV9_P1-injected mouse. The dataset for the PBS and AAV9wt group can be seen in Figure 28.

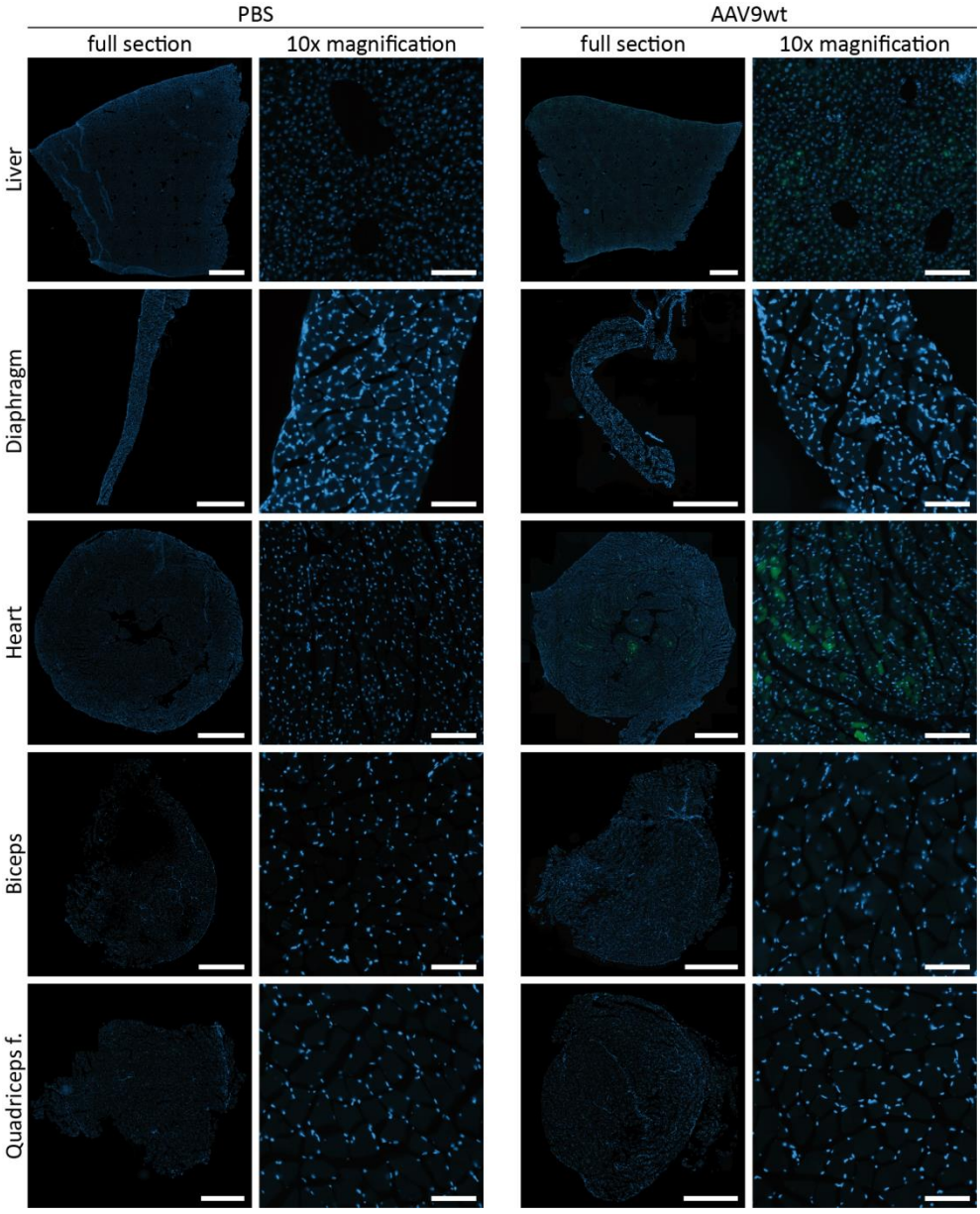

Figure 28: EGFP fluorescence of PBS and AAV9wt group
Images show 10 µm cryosections of the liver, diaphragm, heart, biceps and quadriceps femoris. Representative sections were chosen from C57BL/6J mice injected with $5x10^{11}$ vg/mouse of AAV9wt or PBS as a control. Direct EGFP fluorescence was detected (green) together with the DAPI signal (blue). Scale bar in the full section is 1 mm and 100 µm for the 10x magnification. Exposure was normalized to the diaphragm of the AAV9_P1 group (Figure 29).

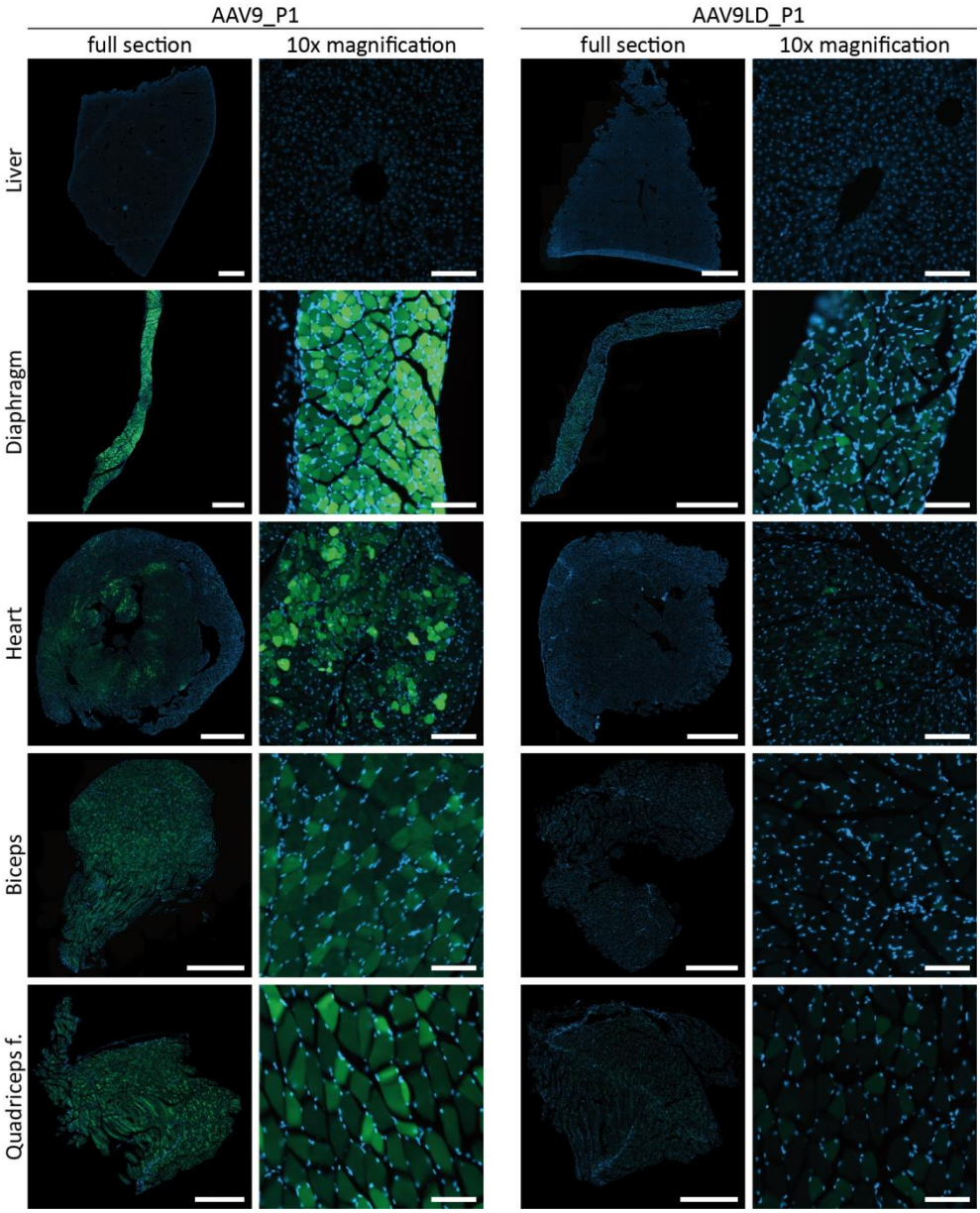

Figure 29: EGFP fluorescence of AAV9_P1 and AAV9LD_P1 group
Images show 10 µm cryosections of the liver, diaphragm, heart, biceps and quadriceps femoris. Representative sections were chosen from C57BL/6J mice injected with $5x10^{11}$ vg/mouse of AAV9_P1 or AAV9LD_P1. Direct EGFP fluorescence was detected (green) together with the DAPI signal (blue). Scale bar in the full section is 1 mm and 100 µm for the 10x magnification. Exposure was normalized to the diaphragm of the AAV9_P1 group.

As expected, no fluorescence was detected in organs of the PBS group. The sections of the AAV9wt-injected mice showed a faint signal in the heart and the liver. Strikingly, AAV9_P1 completely transduced the diaphragm and slightly less efficiently the biceps and quadriceps femoris (Figure 29). In the heart, an EGFP signal could be predominantly observed in the tissue layers surrounding the heart cavity, indicating that transduction in this organ is more heterogeneous than in the other muscles. Importantly, barely any fluorescence

was detected in the liver, further supporting the biodistribution illustrated in Figure 15 and Figure 25.

Concurrent with the results obtained for AAV9LD (Figure 24), the modified P1-displaying capsid, AAV9LD_P1, showed enhanced detargeting from the liver albeit this is difficult to spot with the used exposure settings (Figure 29). Unexpectedly, the rationally designed variant mediated lower EGFP expression in the four muscle tissues compared to AAV9_P1, especially in the heart. Importantly, AAV9LD_P1 still exceeded the benchmark AAV9wt.

Also surprising was that barely any fluorescence could be detected for AAV9wt in this work, at least with the exposure settings used in Figure 28 and Figure 29. To prove that AAV9wt was indeed above background level, the exposure was normalized to the EGFP signal in the liver of AAV9wt (Supplementary information, Figure 43). In the liver, AAV9wt demonstrated EGFP signals in the entire organ but preferentially surrounding the blood vessels. In the heart, a partial transduction was observed for AAV9wt supporting literature data that AAV9wt is highly efficient in this tissue[230–234]. It was already observed in Figure 13 and Figure 24 that most of the capsid ended up in the liver, which may contribute to the relatively weak fluorescence signals in the diaphragm, biceps and quadriceps femoris.

Not surprisingly, the images for AAV9_P1 with the modified settings (Supplementary information, Figure 44) were massively overexposed due to the high activity of the mutant in these tissues. Regardless, these alternative settings strongly supported the conclusion that AAV9_P1 was transducing every cell in the muscle tissues, except for the heart where the signal was weaker in the outer layers. In the liver, individual cells were hit, which is in contrast to AAV9LD_P1 where EGFP fluorescence was almost completely abolished.

5 DISCUSSION

5.1 BARCODE-BASED CAPSID SCREENING

High-throughput capsid engineering strategies such as DNA family shuffling, peptide display and random mutagenesis yield thousands of potentially promising variants with higher efficiency and specificity. However, selecting the best candidate from this pool remains challenging, raising the demand for strategies that facilitate and accelerate the process.

The aim of this work was to establish and apply a barcode-based AAV capsid screening pipeline enabling a simultaneous validation of multiple capsid variants *in vivo*.

5.1.1 ESTABLISHMENT, LIMITATIONS AND FURTHER OPTIMIZATION

Utilizing and building on the information provided in the literature[222,224,225], the barcode was integrated into the 3'UTR of a CMV promoter-driven *eyfp* (Figure 4). Thus, barcode abundance could be monitored on the DNA and RNA level. However, the latter poses further challenges for the normalization strategy.

First of all, a comparison of barcode-comprising viral transcripts across multiple organs or cells assumes equal promoter activity in all analyzed tissues. The CMV promoter is known for its ubiquitous expression, making it a suitable candidate for this screening[235]. Secondly, values for the activity of each virus in every tissue have to be determined which then allow calculating proportions for the individual organs, referred to as specificity. Next generation sequencing of the cDNA samples of each tissue reveals how much of all the detected barcode sequences belong to one capsid variant. However, this proportion alone cannot be used to predict the specificity of a given capsid across all tissues. For instance, a capsid could have a 20% share in the liver and a 20% share in the eye, but these numbers solely describe the efficiency within each organ compared to all other variants in the screening, while they do not allow a statement about the overall distribution of this variant in the body. This fundamental difference - efficiency within a single tissue versus specificity

across all studied tissues - is perhaps best illustrated by the bulk results depicted in Figure 9, which show that most of the library ends up in the liver. This implies that even if a capsid has a high efficiency in a non-liver tissue as compared to all other capsids in the same tissue, its actual main target in the whole body may still be the liver itself, which was indeed frequently in line with our observations. Hence, ideally the relative quantities of all viral transcripts have to be determined in every organ and then multiplied with the barcode proportion obtained from the deep sequencing, resulting in the relative quantities of one variant in the corresponding organ.

Unfortunately, this strategy would introduce a bias since the RNA expression levels of the commonly used housekeeping genes vary across the tissues[236,237]. Due to this reason, the proportional values were normalized to the total viral genomes per cell in the respective organs, by making the assumption that the genome delivered by a particular variant always produces the same amount of transcripts in all analyzed tissues. As previously mentioned, this only applies if the promoter activity is identical in all tissues, which, however, cannot be guaranteed even for a ubiquitous promoter. Thus, determining the specificity inevitably introduces a bias either because of the heterogeneous housekeeper expression or tissue-specific promoter activity. However, assessing the relative quantities of the viral transcripts would require an additional qPCR step for every analyzed tissue, and the resulting values would moreover have to be divided by the total viral genomes (G_β) of the same organ. This calculation normalizes for the potentially unequal promoter activity but cannot correct for the variable housekeeper expression. In conclusion, both approaches are appropriate, yet the strategy that multiplies the NGS proportions with solely the G_β-values was chosen for this work as it saves one extra qPCR step.

Importantly, the applied normalization strategy also corrects for the unbalanced composition of the initial library and for total read count differences between flow cells, ultimately leading to so-called $B_{\alpha\beta}$ values (3.4.13) that describe the overall biodistribution of every vector in the screening. These values can be depicted as proportion of one variant across all tissues ($T_{\alpha\beta}$ values) or as proportion of all variants within one tissue ($V_{\alpha\beta}$ values). The same formulas were used to process the DNA dataset but, unlike the transcripts, the viral genomes are independent of the CMV promoter activity thus omitting this particular bias in the analysis. In summary, the novel normalization procedure enables, for the first time, to simultaneously monitor specificity and efficiency thereby yielding essential information for the characterization of gene therapy vectors.

Most crucial is the design of the barcode itself. In the first attempt to produce barcoded AAV genomes, an oligonucleotide with 10 random bases was ordered and integrated into the 3'UTR via Golden Gate cloning. After bacteria transformation, theoretically every individual colony should comprise a distinct, 10 nucleotides-long barcode. Yet, after confirming that the clones possessed intact ITRs, Sanger sequencing revealed truncated barcodes, homopolymers or even identical DNA sequences. Due to those reasons, the approach had to be canceled and optimized. Consequently, the barcode was extended to 15 bases to increase the mathematically possible unique combinations from 4^{10} to 4^{15}. These second-generation barcodes exhibited a drastic improvement as evidenced by a reduced appearance of homopolymers or identical sequences. From the pool of extracted barcodes, sequences comprising homopolymers with more than 3 identical consecutive nucleotides were excluded since the NGS experiences difficulties when predicting the bases for such stretches. To prevent false assignment of barcodes due to sequencing errors, differences in at least five positions compared to all other barcodes in the library were required. Therefore, the Hamming distance was calculated, and sequences that failed to fulfill these criteria were excluded. In summary, 240 clones were picked, of which 11 lost their ITRs, 64 comprised homopolymers or truncated sequences and another six showed a Hamming distance below 5. Eventually, only 66% of the screened barcodes could be used illustrating how labor-intensive this process is.

An alternative but inevitably more expensive approach could be to order pre-defined barcodes as oligonucleotides. After self-annealing, the barcodes can be pooled and cloned into a backbone with complementary overhangs. To generate a library comprising 100 unique sequences, roughly 165 colonies have to be analyzed and only checked for ITR integrity as determined by probability theory. Pre-defined barcodes could be designed without any homopolymers and a sufficient Hamming distance. Most importantly, the length of the sequences could be cut down to eight nucleotides or even lower while still fulfilling the mentioned prerequisites, in turn providing more freedom when placing the primers for amplicon generation. In the current approach the amplicon is slightly too long, therefore only the reverse sequencing index can be read with a 75 cycle Illumina kit since the required amount of nucleotides to cover the capsid barcode had to be 84. Consequently, this prevents sequencing the forward index and ultimately limits the multiplexing to 32, due to 32 available reverse indexes in the Ovation Low Complexity kit that was used in this work. An optimized and therefore shorter barcode-comprising amplicon would offer the possibility to utilize the forward index, thus allowing to process substantially more samples on one flow cell.

Another crucial parameter during barcode-based capsid screening is the virus production and the resulting composition of the library. Variants for the 1st generation library were each produced using two HEK293T plates and eventually purified over one cesium chloride gradient, in order to save costs and time. Surprisingly, pronounced discrepancies could be detected regarding the production efficiency of the individual variants, creating a largely imbalanced library composition (Figure 6). Nevertheless, the normalization strategy corrects for such effects by using the variant proportions in the initial input library (L_α) and by normalizing the NGS-determined $P_{\alpha\beta}$ values to these ratios. Still, normalization artefacts were observed especially for capsids that were particularly under-represented, such as the peptide insertion mutants of AAV serotypes 1, 6 and 12. For instance, AAV12_P2 and AAV1_P2 were the fourth and ninth most efficient vector in the eye (Figure 7), but analysis of the raw data revealed that AAV12_P2 was only detected in four out of six mice, and AAV1_P2 in one out of six. Despite the minute amounts, the 77-fold and 132-fold under-representation triggered a high multiplication of the respective $P_{\alpha\beta}$ values during the data normalization, in turn leading to the observed artefacts. In fact, this phenomenon was also found in abdominal aorta, brain, brown and white fat tissue as well as kidney in the first screening for the cDNA data (Figure 30), and in abdominal aorta, thoracic aorta, brain colon, eye and white fat tissue for the gDNA data (Figure 31).

In an attempt to create a more homogenous library composition, virus production for the variants of the 2nd generation library was performed individually. To this end, the amount of needed plates was adapted and the respective lysates were run separately over iodixanol gradients. Pooling of equimolar amounts as determined by qPCR led to a balanced library (Figure 8) unlikely to produce normalization artefacts. However, significantly more time and consumables, i.e., 1148 plates and 114 individual iodixanol gradients, had to be invested to produce the 82 variant-comprising 2nd generation AAV library. Still, this is worth the effort as it substantially improves library quality and overall robustness of the pipeline, and as it is probably the only option for screenings of highly diverse capsids including different serotypes, peptide insertions and other mutants. Of note, the production scheme of the 1st generation library likely remains suitable for barcode screenings of lead candidates from directed evolution strategies, such as DNA family shuffling or peptide display, since these variants have inevitably also been selected for high production efficiency. Therefore, the viral particle yield per plate should be similar among these vectors, arguably favoring the less labor-intensive process.

Furthermore, the detection limit of the barcode-based variant screening should be considered. In fact, this limit is difficult to define due to the multifactorial dependency on the dose per variant, homogenous intravenous injections, incubation time before the harvest, transducability of the analyzed tissues, self-complementary or single-stranded AAV genome, sequencing depth, cDNA or DNA detection, the variant itself and the number of potential competitors in the library.

Regarding the dose, the aim was to inject $1x10^{10}$ vg per variant per mouse. However, the library imbalance of the first round resulted in a broad range of abundance between the individual variants. A good example for the detection limit was provided by AAV4mut_A2 which was the least abundant capsid in the library with $3x10^{6}$ vg/mouse, with a 3600-fold deviation from the mean. Serotype AAV4 and its peptide-modified variants as well as AAV4mut exhibited a strong lung tropism, taking eight spots in the top 10 list (Figure 7). The remaining peptide insertion mutants of AAV4mut could all be found in the top 25 except for AAV4mut_A2, for which no read counts were measured in the six mice. Most likely, this particular capsid would have demonstrated the same preferred lung-targeting if equimolar titers would have been used. The dose of $1x10^{10}$ vg per variant in the second screening was enough to detect read counts in all six mice for at least two thirds of the variants, even in poorly transduced tissues. This is sufficient to analyze the efficiency and specificity of promising candidates.

The incubation time before harvesting the organs was two weeks for the first screening and one week for the second and third. No noticeable difference could be detected when reducing the time. Further decreasing the incubation will eventually diminish the chance of the vectors to transduce the target tissues. Prolonging the time by several weeks probably enhances silencing of the CMV promoter[238].

Arguably the highest influence on the detection limit is exerted by the screened organs and variants themselves. As depicted in Figure 9, the vast majority of the library ended up in the liver. Organs such as the eye and the digestive tract were difficult to transduce and therefore required a more sensitive detection. Among the variants, dramatic variations in terms of efficiency were observed (Figure 10). For instance, vectors belonging to the AAV2, AAV3 and AAV5 family demonstrated a very weak *in vivo* activity, resulting in only a few barcode reads.

Equally important is the size of the library and the associated variant competition. These resulting interfering effects were illustrated in the third

screening, where the settings were kept identical to the second round while the library was enriched with further 75 capsids. For each of them, $1x10^{10}$ vg/mouse were injected intravenously. Although many results could be reproduced, the majority of weakly active AAVs hit the detection limit. For example, in the context of the 2nd generation library, AAV2wt exhibited a minor 0.14% share in the liver with a marginal standard deviation across the six mice. However, in the third screening, one out of four mice had no read counts even though AAV2wt preferentially targets the liver. As expected, the detection was even more challenging on the DNA level since only the barcodes on the viral genomes are measured, whereas promoter-amplified barcode-comprising transcripts are counted on the cDNA level. In general, a deeper sequencing of the samples could potentially rescue some inefficient variants but comes with increased costs.

In conclusion, the chosen experimental settings in this work were appropriate to identify highly efficient and specific AAV capsids in the mouse. For upcoming screenings, the parameters from the second *in vivo* characterization study should be copied and library diversities should be restricted to a maximum of 100 variants.

5.1.2 COMPARISON TO PUBLISHED DATA

After establishing the barcode-based capsid screening pipeline, the output values of important benchmarks can be compared to the literature to prove the robustness of the system. In this work, the DNA family shuffled variant AAVDJ demonstrated superior specificity for the liver (Figure 14) and efficiency scores slightly worse than AAV8wt (Figure 11). The efficiency was already studied in the original 2008 publication of Grimm *et al.* by showing hFIX expression levels similar to AAV8wt and AAV9wt *in vivo* up to a certain dose[172]. The specificity on the protein level was shown twice in 2016 by reports documenting highly liver-tropic detection of luciferase signals[164,178]. Of note, in all three publications and in this work, C57BL/6 mice were used supposedly explaining the comparable outcome. In the third barcode screening, AAVDJ had seemingly lost part of its liver specificity and in exchange exhibited activity in CD11b- and CD11c-positive cells (Figure 22). However, the applied normalization strategy calculates the proportion per cell, thus favoring minor cell populations like the mentioned immune cells. Since the liver is one of the largest organs, the AAVDJ proportion in this tissue greatly outweighs the off-targets when calculating the proportion per organ. So far unpublished is the selective hepatocyte-targeting of the chimera in an *in vivo* context (Figure 23),

which is congruent with the fact that AAVDJ was selected in human hepatocytes. Next to the shuffled vector, such a specific liver sub-cell type tropism could only be observed for AAV5_P5 and AAV8_P5.

Also very impressive are the results obtained for AAV2_L1. This peptide displaying variant showed a pronounced specificity for the lung and some off-targeting to blood cells and the brain (Figure 14). In the original work of Körbelin *et al.*[208], the blood cells were not analyzed as an off-target but several other tissues were assessed for luciferase expression. Strikingly, the lung/brain ratio of AAV2_L1 in the second barcode screening is identical to the one seen in Figure 4 of the original publication[208]. This comparison is especially important since both results were obtained with completely different techniques, further illustrating the robustness of the barcode-based system.

Another important benchmark that has recently attracted substantial attention in the AAV field is AAV9_PHP.B. This variant that has been selected for astrocyte-targeting demonstrated superior efficiency (Figure 11) and specificity (Figure 14) in the whole brain where it outcompeted the commonly used AAV9wt. This result is consistent with published histology data [219,239–241]. Of note, Hordeaux and colleagues showed that AAV9_PHP.B activity is limited to C57BL/6J mice[241] that were, by coincidence, also the mouse strain of choice for this work, therefore delivering evidence for the comparable outcome. The exceptional case of AAV9_PHP.B is discussed in more detail in chapter 5.5.

Next to the compelling confirmation of the benchmark results, the screenings additionally excelled in reproducibility between the screening rounds. For instance, AAVrh10wt proved to be the most efficient capsid in the liver in all three screens and even outperformed AAV8wt, albeit only marginally. In the third screening, AAV8wt was slightly separated from AAVrh10wt, mostly by capsids that were newly added in this round. Interestingly, the comparable efficiency of AAVrh10wt and AAV8wt has also been documented in the literature[242]. Nathwani *et al.* also showed highly similar vector genomes per cell for both wild types, which could be confirmed here in all three capsid screening rounds. Among the top 10 variants in the pancreas of the first and second screen, the first eight vectors were in identical positions when excluding the newcomers of the second round. In the same round, AAV9_P1 was 1.6-fold more effective than AAV9wt in the heart (Figure 12). The successive screening once more exhibited a 1.6-fold difference between the two vectors (Figure 18). These findings and several further, similarly consistent results observed in other tissues or for other variants convincingly prove the robustness and reproducibility of the barcode-based variant

screening pipeline that was established here. This conclusion is of high importance since only a stable system can be used to reliably identify the most potent variant from a pool of potential lead candidates.

5.2 CHALLENGES IN RATIONAL CAPSID DESIGN

The analyzed AAV variants in this work were generated with different capsid engineering techniques. Next to the directed evolution approaches like DNA family shuffling, peptide display and error-prone PCR that were used to create the benchmarks in our screens, over 70 mutants based on the naturally occurring serotypes were used that display elsewhere-selected peptides. A highly similar panel of viruses was previously tested in our laboratory (work of primarily Kathleen Börner and Eike Kienle) and proved to be vastly effective in cultured cells *in vitro* (manuscript in preparation). Interestingly, in this work, the peptide-modified vectors behaved very differently as compared to the respective wild type versions in mice, clearly exemplifying the very restricted transferability between *in vitro* and *in vivo* systems. The same effect was also observed for commonly used AAV serotypes by exhibiting diverging transduction profiles in cell lines[243] as compared to the *in vivo* situation in mice[153]. Plausible explanations are the additional barriers in a living organism including the more challenging accessibility of the target tissue and potential interactions with the host immune system. Moreover, it has to be noted that a direct comparison is difficult since the *in vitro* screenings were performed in one particular cell line whereas mostly whole organs were analyzed in the barcode-based screenings.

Further interesting and useful observations were made when attempting to rationally improve capsids by transferring peptides isolated through directed AAV evolution between two different capsids. In detail, the natural isolate AAV4wt and the previously selected peptide display mutant AAV2_L1[208] were detected preferentially in the lungs. The fact that AAV2wt itself predominantly targets the liver suggested that it is the L1 peptide that mediates the lung tropism and thus pointed towards the possibility to further improve AAV4's activity in the lung by integrating the L1 peptide into AAV4wt. Curiously, though, the resulting AAV4_L1 variant exhibited equal efficiencies as the other AAV4-based peptide insertion mutants but remained below that of the parental AAV2_L1. This is a very important result as it highlights the synergism between capsid backbone and inserted peptide that ultimately governs the properties of the resulting synthetic viral particle.

A second example supporting this seminal conclusion is AAV9_BR1 that was designed here to combine the most potent serotype for brain-targeting, AAV9wt, with the brain-homing peptide BR1 from the directed evolution variant AAV2_BR1[218]. Also here, it was observed that the identical peptide led to strikingly different particle phenotypes depending on the capsid backbone it was presented in, again illustrating the complexity of AAV biology and the challenges in rational AAV capsid design.

A third consistent example from the literature is AAV2_MTP, a variant displaying a putative muscle-targeting peptide MTP that was isolated by phage display in muscle tissue. The paper reported an enhancement in the diaphragm, heart and gastrocnemius after systemic delivery in mice over AAV2wt, which is very inefficient in these tissues *in vivo* [201]. While these effects could be confirmed for all the muscle tissues in this doctoral work, 90% of vector activity was actually measured in the liver. This highlights the importance of performing comprehensive screens in a wide variety of tissues, as a prerequisite for drawing fair and unbiased conclusions about *in vivo* capsid efficiency and/or specificity.

Strikingly, we found that even single point mutations can significantly change the behavior of AAVs in the complex setting of a mouse. The first barcoded variant screening comprised capsids based on AAV4 with a K544E mutation, herein referred to as AAV4mut. All members of this family were less active in the lung in direct comparison to their wild type parent, AAV4wt. An example that the influence of a few mutations can also be negligible was AAVDJYF that was introduced to the 3rd generation library and bears three tyrosine-to-phenylalanine exchanges that could potentially improve particle stability by preventing proteasomal degradation. The beneficial effects of such tyrosine mutants was demonstrated previously[159,161–163,244] and now applied here by mimicking the respective residue alterations of an AAV2 triple mutant[161] in AAVDJ. Interestingly, the resulting mutant showed the same specificity (Figure 22 and Figure 23) and efficiency as AAVDJ, once more illustrating the complex interplay of capsid backbone and ectopic modifications, such as point mutations here or peptide insertions above. Last but not least, AAV9LD_P1 was generated to further enhance the effects of AAV9_P1, by including two point mutations P504A and G505A that led to a massive detargeting of the liver when integrated into AAV9wt[222]. This published mutant, AAV9LD, was already more specific for the muscle in comparison to its parental virus, implying that transfer of these two point mutations may improve our own variant. Remarkably, even though only minimal changes were made to AAV9_P1, its high efficiency in muscle was

actually decreased in exchange for a better liver-detargeting, as seen in the histology sections of the analyzed tissues (Figure 29).

In summary, rationally designing AAV capsids by transferring peptides or mutations between two capsids is challenging since the assembled AAV particle can be highly sensitive to even minimal variations, ultimately resulting in unpredictable phenotypes. This is perhaps best exemplified by our capsids resulting from the transfer of peptides that were pre-selected via random peptide display in AAV2, such as the P2, P4, P5, A1, A2 and A6, into another AAV serotype. In most cases, this integration into a slightly different AAV context largely changed particle behavior, most likely due to different steric requirements in the exposed capsid regions of closely-related serotypes. Even when *in vivo* selected peptides such as L1 and BR1 were transferred to the same integration site of another isolate, this typically yielded a phenotype that differed from the parental peptide-modified capsid. Still, two notable exceptions were observed in this work, namely AAV9_P1 and AAV9_P3, that both use peptides isolated through AAV2 peptide display. Both mutants exhibit a remarkable muscle-targeting that most likely results from the synergistic action of the capsid backbone and the peptide. Hence, they serve as very encouraging examples for how transfer of a given peptide between two serotypes can in fact create entirely novel and beneficial capsid features.

Taken together, the data presented here and literature findings show that AAV is a tremendously versatile and promising scaffold for the design of synthetic capsids and vectors with original features *in vitro* and *in vivo*. Concurrently, the presented results also support the notion that additional work and knowledge on AAV capsid biology are urgently needed in order to realize the potential of rational design, whereby the outcome of a capsid modulation including particle stability or receptor interaction can be fully predicted in advance.

5.3 DIRECTED EVOLUTION – THE HOLY GRAIL IN CAPSID ENGINEERING?

Until the field possesses sufficient knowledge to realize rational design of AAV capsids, two major capsid engineering techniques, DNA family shuffling and random peptide display, showed great promise for the identification of more efficient or specific vectors. Both strategies rely on the generation of AAV libraries with diversities of around $1x10^7$ novel synthetic variants that are subsequently used for transducing cells or animals. Viral genomes are then in

most cases PCR-rescued from successfully penetrated cells or organs of interest, therefore excluding variants incapable of reaching the target tissue. By repeating this procedure for several rounds, candidates are enriched that outperform their competitors in terms of e.g. efficiency. In this work, several capsids resulting from such a directed evolution scheme were included, either from the literature and then serving as benchmarks, or novel variants independently isolated by the group members Anne-Kathrin Herrmann and Jihad El Andari, which allowed us to assess the potency of these techniques.

The DNA family shuffled chimera AAVDJ is a very good example for a successful selection. The vector reported in 2008 by Grimm and colleagues was isolated from human hepatocytes after five consecutive rounds of screening of a shuffled capsid library in the presence of intravenous immunoglobulin and found to restrict the biodistribution to the liver[172]. Indeed, these data could be verified in the second and third barcode screening where AAVDJ demonstrated strong liver-targeting (Figure 14 and Figure 22). Concurrent with the original publication, AAVDJ was less efficient than AAV8wt in this organ (Figure 11) which is worth noting since directed evolution in theory mainly selects for a higher efficiency or in the case of AAVDJ in addition for antibody-evading features. Most impressively, within the liver the chimera was found almost exclusively in hepatocytes, reflecting the selection strategy used for its isolation (Figure 23). AAVLK03[191], AAVM41[189] and the lead candidates of the muscle selection (work of Jihad El Andari), AAVJEA3-S1, AAVJEA3-S10, AAVJEA3-H15 and AAVJEA3-D20 furthermore support the hypothesis that DNA family shuffling is oftentimes yielding vectors with an increased specificity. However, none of the mentioned chimeras demonstrated a higher efficiency compared to their parental counterparts. The fact that only wild type AAV isolates are used for shuffling can potentially explain the phenomenon since the resulting chimeras are restricted to the provided sequences of the parents. Hence, the chimeric sequences presumably possess a lesser chance to form motifs needed for a complete retargeting.

On the contrary, in random peptide display approaches, novel motifs are integrated into the capsids, thus offering the possibility to utilize a different entry mechanism. For instance, variants such as AAV9_PHP.B, AAV2_BR1 and AAV2_L1 were extracted from peptide displays and show a concurrent increase in specificity and efficiency (Figure 11 and Figure 14). Interestingly, the potency of the parental backbone for the insertion seems to play a minor role as illustrated by AAV2_L1 and AAV2_BR1. AAV2wt itself demonstrated a high liver specificity (Figure 13) and in general a weak efficiency in all tissues (Figure 10). Nevertheless, the incorporation of nine additional amino acids

converted the capsid to a particularly effective and specific mutant. Once more, the less important native activity of the parental wild type capsid for peptide display is in contrast to DNA family shuffling where enriched chimeras are oftentimes comprised of sequences from efficient serotypes in the respective organ or cells where they were selected in. For example, AAVDJ is a mixture of AAV2wt, AAV8wt and AAV9wt from which especially AAV8wt exhibited high efficiency in the liver (Figure 11). However, the chimera has the highest homology to AAV2wt, deviating in 60 amino acids. AAV2wt proved to be vastly effective *in vitro*, especially in the human hepatoma cell lines Huh7 and HepG2 (doctoral thesis of Eike Kienle), the cells AAVDJ was isolated from. Another example are the muscle-tropic chimeras of Jihad El Andari that comprise long stretches of the most potent wild type in the muscle, AAV9wt, at the C-terminus (data not shown). Finally, the shuffled variant AAVAH3-5, selected in stellate cells, consists of mainly AAV1wt (doctoral thesis of Anne-Kathrin Herrmann), a serotype demonstrating a surprisingly similar efficiency (4.5) and specificity (Figure 20 and Figure 21) in the on-target further supporting the abovementioned theory.

One very important aspect when choosing one of the two directed evolution approaches for the development of a tailored variant is the monitoring of the library during the selection. During every round, the library composition will change, which yields vital information about the enrichment of certain motifs. However, tracking a library created through DNA family shuffling is challenging since the whole 2.2 kb-long capsid gene undergoes alterations. Traditional Illumina sequencing cannot resolve these changes since homologous sequences are needed for the required alignment. Recently, another sequencing technology became available, namely PacBio sequencing[221]. The advantages are the particularly long read lengths covering the 2.2 kb with ease. Regardless, the system requires improvements since the total number of reads, roughly 50.000-100.000, cannot cover typical library diversities of up to $1x10^7$. In addition, the system is more expensive than traditional sequencing and, in its current iteration, introduces many insertions and deletions to the sequences based on our own experiences, which complicates the analysis. Random peptide libraries pose an advantage as one can readily exploit Illumina sequencing due to the only ~30 bp-long peptide-encoding DNA stretch that has to be resolved. Thereby, up to 450 million reads per sample can be generated when using the NextSeq500, which exceedingly covers a typical library. Accordingly, monitoring of every selection round can identify peptides with increasing abundance, as was perfectly exemplified in the original publication of AAV2_BR1[218]. By concomitantly also sequencing off-targets, the

collective information can be utilized to calculate the enrichment scores in the on-target while simultaneously determining tissue specificity[208].

In general, to obtain organ-specific variants, the number of selection rounds is crucial for both mentioned directed evolution strategies. Most tissue-specific published vectors resulted from five iterative rounds and therefore had to survive a strong selection pressure. Attempts to save time by lowering the repetitions increases the riskto obtain sub-optimal progeny, as exemplified by the chimera AAVB1, a variant that was selected for central nervous system-targeting after only one selection round[190]. Instead, the variant possesses a pronounced liver specificity (Figure 24). A second example are the chimeric muscle-tropic vectors from our laboratory that underwent either two or three selection rounds. All chimeras experiencing only two cycles were predominantly found in the liver followed by CD11b- and CD11c-positive cells as well as fat tissue. Notably, the four lead candidates with increased muscle-targeting, AAVJEA3-S1, AAVJEA3-S10, AAVJEA3-H15 and AAVJEA3-D20, were all extracted after three rounds. One exception is the brain-specific peptide display mutant, AAV9_PHP.B, that was isolated after the second round of selection. The fact that this relatively short selection scheme was still successful is perhaps explained by the use of the novel CREATE system, in which only those viral genomes that had undergone Cre-mediated recombination in astrocytes can be rescued, which substantially raised the stringency of the system[219].

In conclusion, directed evolution is currently arguably the most promising method to identify highly selective and effective capsids, especially after several selection rounds. Clearly and not surprisingly, a major challenge is to selectively and robustly target individual cells types. Although AAVDJ managed to preferentially hit hepatocytes, the lead candidate of our laboratory from a stellate cell selection, AAVAH3-5, additionally transduced Kupffer cells and LSECs (Figure 21) indicating a limit for a highly specific tissue-targeting when solely relying on the capsid. Nevertheless, specificity can be enhanced by combining a beneficial capsid with tissue-specific promoters or *cis*-acting elements for the transcriptional regulation with endogenous mRNAs. Last but not least, synthetic AAVs generated by directed evolution strategies are simultaneously selected for a potent production efficiency (experiences made in our laboratory), which is essential considering the immense manufacturing effort to yield high titers for clinical trials[245].

5.4 AAV9_P1 – AN UNEXPECTED NEWCOMER FOR MUSCLE GENE THERAPY

Genetic disorders leading to muscle diseases are ideal targets for an AAV-based gene therapy. In order to reach every affected muscle tissue in the whole body, an intravenous administration is required creating challenges concerning immune responses and off-targeting effects. To date, AAV9wt proved to be the most efficient wild type AAV in multiple animals[229] therefore qualifying it for the use in clinical trials. However, this work revealed a strong bias towards the liver for the naturally occurring isolate (Figure 13 and Figure 24), illustrating the existing room for improvement and the urgent need in the AAV field of new, potent and muscle-tropic vectors that could foster the implementation of gene therapies for many muscle diseases.

Surprisingly, a capsid added in the second barcode library, AAV9_P1, showed superior efficiency in the diaphragm, heart, biceps and quadriceps femoris (Figure 12 and Figure 18). In addition, it also exhibited a greatly improved specificity as compared to AAV9wt (Figure 15 and Figure 25). Importantly, separate validation of this capsid out of the library context could confirm the results on the cDNA (Figure 26) and protein level (Figure 29). In strong contrast to AAV9wt, the P1-displaying variant detargets the liver and many other organs, restricting ~70% of its activity to the muscles. These striking effects are caused by integrating the nine amino acid-long peptide GRGDLGLSA into the AAV9 capsid protein after position 588 (in VP1). The peptide, herein referred to as P1, was originally discovered during the screening of a random peptide display library based on AAV2wt inmurine breast cancer PymT cells[213]. In the course of his doctoral work in the Grimm laboratory, Eike Kienle had incorporated promising peptides from the literature, including P1, into AAV9 and 11 other AAV serotypes, with the aim to study the interplay of capsid and peptide concerning particle efficiency and specificity.

Indeed, this prior work resulted in the identification of numerous combinations of capsid and peptide that often surpassed the parental wild type in a vast collection of cell types in culture. This comprises AAV9_P1 which clearly outperformed AAV9wt and many other capsids in cultured human astrocytes where it was the lead candidate[202], as found and published together with a collaboration partner (Ruth Brack-Werner) at the Helmholtz Center in Munich. Intriguingly, the same capsid as well as the closely related AAV9_P3 behaved very differently in this work in peripherally injected mice, as

discussed above. Interestingly, P1 (GRGDLGLSA) and P3 (GRGDAVGVA) both comprise an RGD motif that was shown to interact with integrins[196,246,247] suggesting an alternative entry mechanism potentially independent of the used N-linked galactose[101,102], AAVR[87] or LamR[90]. Of note, P2 (GCDCRGDCFCA) displayed by AAV9wt was markedly less efficient than its parent and not found in the top 10 in the analyzed muscle tissues (Figure 7), despite containing an RGD motif. In contrast to P1 and P3, the motif starts three amino acids later potentially explaining the differences. This hypothesis is furthermore backed up by AAV9_K3 which comprises a peptide (GRGDLRVSA) that is highly similar to P1, deviating in only two amino acids. Stunningly, AAV9_K3 was predominantly detected in the liver (Figure 24) and demonstrated a ~20-fold reduction in efficiency as compared to AAV9_P1 in the skeletal muscles and 6-fold in the heart. Importantly, K3 was integrated after amino acid 589 thereby marginally altering the position of the RGD motif. In conclusion, the exact position of the motif seems to be of high importance whereas the successive amino acids can tolerate more changes without losing the beneficial effects, as exemplified by AAV9_P3. The exact role of the RGD-comprising peptide in the context of AAV9 and related capsids for muscle and astrocyte transduction is a matter of ongoing investigation in our group and the collaboration partner in Munich. By integrating P1, a ~10-fold reduction in the liver was observed in comparison to AAV9wt, indicating that fewer virus particles are trapped in the liver, which may add to the observed pronounced muscle activity. In fact, a double point mutation variant published by Adachi *et al.*, AAV9LD[222], largely detargets the liver by ~100-fold and concurrently showed higher specificity for the muscle tissues (Figure 24). However, transferring these two mutations to AAV9_P1 greatly diminished the expression of EGFP in the muscle sections (Figure 29), suggesting an even more complex correlation.

In summary, based on the information provided by this work, the question remains whether the peptide-induced beneficial effects in the muscles are due to an improved muscle homing, the detargeting from the liver or a combination of both factors. Regardless of mechanism, the tremendous potential of AAV9_P1 as a candidate for muscle-directed gene therapy is obvious and undisputed. Currently, extremely high doses exceeding $1x10^{14}$ vg/kg are injected into animals in order to reach curative effects with the best vector on the market, AAV9wt (American Society of Gene and Cell Therapy, ASGCT 2018). By using AAV9_P1, the vector load could be reduced by at least 10-fold while maintaining high levels of transgene expression and additionally limiting the transcriptional activity predominantly to the muscle tissues, thus preventing unwanted off-targeting effects. Furthermore, since

AAV9_P1 produces as efficiently as AAV9wt, injecting lower doses would significantly decrease manufacturing costs and ultimately lower the prize for a related gene therapy product.

5.5 RESTRICTED TRANSFERABILITY – YOU GET WHAT YOU SCREEN FOR

Many synthetic AAV variants have been generated over the last 15 years by directed evolution approaches, trying to develop more efficient vectors for the transduction of cells or organs. This work showed that results obtained from published variants can indeed be reproduced when recapitulating the same experimental outlines. For instance, AAV2_L1 was selected for the murine lung after intravenous injection and demonstrated a drastically increased efficiency and specificity[208]. This could be confirmed after intravenously injecting the barcoded AAV library comprising AAV2_L1 into C57BL/6J mice. However, changing the injection route most likely alters the properties of a selected variant, as exemplified by AAV2_7m8[186]. Dalkara *et al.* intravitreally injected an AAV library in search of a lead candidate for the outer retina. The resulting hit, AAV2_7m8, was able to rescue two retinal diseases in mouse models[186]. Here, the peptide display mutant was added to the 2nd generation library. Notably, it showed a 90% specificity for the liver after systemic application, supporting the theory that the injection route should be kept constant to achieve the same result.

Another crucial aspect is the target tissue that was used for the selection. The shuffled chimera AAVLK03 was extracted from human hepatocytes in a xenograft mouse model and showed high efficiency in the human cells, in contrast to poor transduction of murine hepatocytes[191]. In this doctoral work, this variant was tested in a murine liver and exhibited a 200-fold lower activity than AAV8wt albeit 91% of LK03 targeted the liver. Hence, in the absence of its on-target (human hepatocytes), AAVLK03 exhibited a unique behavior. The same applies for the variants AAV9_K1 and AAV9_K3 that were selected on human coronary artery endothelial cells[215]. Both vectors were predominantly liver-tropic in the present work, barely showing any specificity for the heart *in vivo*.

Arguably the most important factor is the transferability of a selected AAV variant to clinically relevant animal models such as non-human primates and eventually to humans. The challenges are best exemplified by the brain-targeting peptide display mutant AAV9_PHP.B. Deverman and colleagues

isolated the capsid after two selection rounds in C57BL/6J mice utilizing their novel CREATE system and demonstrated superior efficiency and specificity compared to AAV9wt[219]. These effects could be fully confirmed in this work (Figure 11 and Figure 14). However, follow-up publications by Matsuzaki *et al.* and Hordeaux *et al.* could not prove the superiority of AAV9_PHP.B in the marmoset[239] and rhesus macaques[241] brain, respectively. Furthermore, unpublished work from the groups of Gray-Edwards and Sena-Esteves presented at the ASGCT conference in 2018 showed no transduction increase in the sheep and cat brain. Most strikingly, the abovementioned publication by Hordeaux *et al.* in addition demonstrated a discrepancy between the mouse strains C57BL/6J and BALB/cJ. The impressive features of AAV9_PHP.B were entirely absent in BALB/cJ mice, indicating a lack of transferability even within the same species. Nevertheless, it has to be noted that, for example, AAV2_L1 and AAV2_BR1 were selected in FVB/N mice and validated in this work in C57BL/6J mice, suggesting that AAV9_PHP.B may represent an exceptional case.

In summary, isolating and characterizing a variant for a specific application does not guarantee similar results when altering the injection route or the animal model, or when switching from *in vitro* to *in vivo*. Thus, the notable effects observed for AAV9_P1 in mice should be considered with caution. Nonetheless, it is certain that AAV9_P1 is a superior variant for muscle-targeting in C57BL/6J mice after intravenous injection. However, the transferability to higher animal models has to be elucidated first prior to considering AAV9_P1 as a potential vector for clinical trials. Generally, to increase chances to obtain capsids that are relevant for use in humans, it seems advisable to perform library selections directly in non-human primates. Although the initial costs would be significantly higher and ethical considerations will have to be made, costs may be saved in the long run since fewer validations have to be performed due to the increased chance for enhanced transferability of resulting lead candidates.

5.6 CONCLUSIONS AND PERSPECTIVES

In the course of this doctoral work, a barcode-based AAV capsid screening pipeline was established allowing the simultaneous tracking of over 100 variants in the context of a living organism. The applied comprehensive normalization strategy produces essential values denoting the specificity and efficiency of every analyzed barcoded AAV. The system turned out to be highly robust and especially helpful for the characterization of lead candidates after directed evolution approaches, consequently reducing animal numbers and downstream processing. These benefits not only promote a barcode-based capsid screening in higher animals such as non-human primates, but furthermore allow applying the pipeline to answer capsid-unrelated questions. For instance, different doses of the same vector, represented by corresponding barcodes, can be tested in one organism, once more omitting high animal numbers. Even the transduction differences[143,144] between barcode-labeled self-complementary and single-stranded AAV genomes could be assessed in more detail exemplifying the enormous potential of this technology.

The P1-displaying variant AAV9_P1 was an unexpected discovery of the herein used barcode-based capsid screenings, demonstrating superior efficiency in the muscle tissues and increased muscle-targeting on the transcript and protein level. The question remains how AAV9_P1 mediates the improved effects and how crucial the placement of the RGD motif is. One simple experiment might be to generate mutants comprising a slightly shifted P1 peptide by one amino acid upstream or downstream. Additionally, an alanine walk should elucidate the importance of each position. Although P1 was originally extracted by Michelfelder *et al.*[213] and found once more in 2016 by Körbelin *et al.*[208], the peptide was selected in the context of AAV2 and therefore not optimized for AAV9. Hence, based on the information provided by altering the RGD position and the alanine walk, amino acids that are essential for the improvements in the muscle tissues can be fixed while the remaining ones offer the possibility to be randomized. The resulting peptide-displaying AAV library can subsequently be screened in muscles tissues for new mutants with an improved P1 peptide. In conclusion, despite the already great promise of AAV9_P1, the development of an optimized muscle-tropic gene therapy vector will continue, raising hopes to eventually be able to cure patients with severe muscle diseases.

REFERENCES

1. Felgner, P. L. *et al.* Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure. *Proc. Natl. Acad. Sci.* **84,** 7413–7417 (1987).
2. San, H. *et al.* Safety and short-term toxicity of a novel cationic lipid formulation for human gene therapy. *Hum. Gene Ther.* **4,** 781–788 (1993).
3. Yuhong Xu and Francis C. Szoka, J., Transfection, C., Xu, Y. & Szoka, F. C. J. Mechanism of DNA Release from Cationic Liposome / DNA Complexes Used in. *Biochemistry* **2960,** 5616–5623 (1996).
4. Wu, G. Y. & Wu, C. H. Receptor-mediated in vitro gene transformation by a soluble DNA carrier system. *J. Biol. Chem.* **262,** 4429–4432 (1987).
5. Boussif, O. *et al.* A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine. *Proc. Natl. Acad. Sci.* **92,** 7297–7301 (1995).
6. Neumann, E., Schaefer-Ridder, M., Wang, Y. & Hofschneider, P. H. Gene transfer into mouse lyoma cells by electroporation in high electric fields. *EMBO J.* **1,** 841–845 (1982).
7. Titomirov, A. V., Sukharev, S. & Kistanova, E. In vivo electroporation and stable transformation of skin cells of newborn mice by plasmid DNA. *BBA - Gene Struct. Expr.* **1088,** 131–134 (1991).
8. ter Haar, G. Therapeutic applications of ultrasound. *Progress in Biophysics and Molecular Biology* **93,** 111–129 (2007).
9. Endoh, M. *et al.* Fetal gene transfer by intrauterine injection with microbubble-enhanced ultrasound. *Mol. Ther.* **5,** 501–508 (2002).
10. Kim, H. J., Greenleaf, J. F., Kinnick, R. R., Bronk, J. T. & Bolander, M. E. Ultrasound-Mediated Transfection of Mammalian Cells. *Hum. Gene Ther.* **7,** 1339–1346 (1996).
11. Klein, T. M., Wolf, E. D., Wu, R. & Sanford, J. C. High-velocity microprojectiles for delivering nucleic acids into living cells. *Nature* **327,** 70–73 (1987).
12. Liu, F., Song, Y. K. & Liu, D. Hydrodynamics-based transfection in animals by systemic administration of plasmid DNA. *Gene Ther.* **6,** 1258–1266 (1999).
13. Zhang, G., Song, Y. K. & Liu, D. Long-term expression of human alpha1-antitrypsin gene in mouse liver achieved by intravenous administration of plasmid DNA using a hydrodynamics-based procedure. *Gene Ther.* **7,** 1344–1349 (2000).
14. Zhang, G. *et al.* Hydroporation as the mechanism of hydrodynamic delivery. *Gene Ther.* **11,** 675–682 (2004).
15. Villemejane, J. & Mir, L. M. Physical methods of nucleic acid transfer: General concepts and applications. *British Journal of Pharmacology* **157,** 207–219 (2009).
16. Medina-Kauwe, L. K., Xie, J. & Hamm-Alvarez, S. Intracellular trafficking of nonviral vectors. *Gene Ther.* (2005). doi:10.1038/sj.gt.3302592

17. Lukacs, G. L. *et al.* Size-dependent DNA mobility in cytoplasm and nucleus. *J. Biol. Chem.* **275,** 1625–1629 (2000).

18. Danos, O. & Mulligan, R. C. Safe and efficient generation of recombinant retroviruses with amphotropic and ecotropic host ranges. *Proc. Natl. Acad. Sci.* **85,** 6460–6464 (1988).

19. Miller, A. D. *et al.* Construction and properties of retrovirus packaging cells based on gibbon ape leukemia virus. *J. Virol.* **65,** 2220–4 (1991).

20. Williams, D. A., Lemischka, I. R., Nathan, D. G. & Mulligan, R. C. Introduction of new genetic material into pluripotent haematopoietic stem cells of the mouse. *Nature* **310,** 476–80 (1984).

21. Riviere, I., Brose, K. & Mulligan, R. C. Effects of retroviral vector design on expression of human adenosine deaminase in murine bone marrow transplant recipients engrafted with genetically modified cells. *Proc. Natl. Acad. Sci.* **92,** 6733–6737 (1995).

22. Mavilio, F. *et al.* Peripheral blood lymphocytes as target cells of retroviral vector-mediated gene transfer. *Blood* **83,** 1988–1997 (1994).

23. Bunnell, B. A., Muul, L. M., Donahue, R. E., Blaese, R. M. & Morgan, R. A. High-efficiency retroviral-mediated gene transfer into human and nonhuman primate peripheral blood lymphocytes. *Proc. Natl. Acad. Sci. U. S. A.* **92,** 7739–43 (1995).

24. Aitui, A. *et al.* Gene Therapy for Immunodeficiency Due to Adenosine Deaminase Deficiency. *N. Engl. J. Med.* **360,** 447–458 (2009).

25. Wu, C. & Dunbar, C. E. Stem cell gene therapy: The risks of insertional mutagenesis and approaches to minimize genotoxicity. *Frontiers of Medicine in China* **5,** 356–371 (2011).

26. Montini, E. *et al.* The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy. *J. Clin. Invest.* **119,** 964–975 (2009).

27. Behavior, M. & Cortex, P. LMO2-Associated Clonal T Cell Proliferation in Two Patients after Gene Therapy for SCID-X1. *Science (80-.).* **304,** 1782–1787 (2004).

28. Turtle, C. J. *et al.* Immunotherapy of non-Hodgkin ′ s lymphoma with a defined ratio of CD8 + and CD4 + CD19-specific chimeric antigen receptor – modified T cells. *Sci. Transl. Med.* **8,** 355ra116 (2016).

29. Davila, M. L. *et al.* Efficacy and toxicity management of 19-28z CAR T cell therapy in B cell acute lymphoblastic leukemia. *Sci. Transl. Med.* **6,** (2014).

30. Ali, S. A. *et al.* T cells expressing an anti-B-cell maturation antigen chimeric antigen receptor cause remissions of multiple myeloma. *Blood* **128,** 1688–1700 (2016).

31. Brown, C. E. *et al.* Regression of Glioblastoma after Chimeric Antigen Receptor T-Cell Therapy. *N. Engl. J. Med.* **375,** 2561–2569 (2016).

32. Raper, S. E. *et al.* Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer. *Mol. Genet. Metab.* **80,** 148–158 (2003).

33. Capasso, C., Garofalo, M., Hirvinen, M. & Cerullo, V. The evolution of

adenoviral vectors through genetic and chemical surface modifications. *Viruses* **6,** 832–855 (2014).

34. Gaudet, D. *et al.* Efficacy and long-term safety of alipogene tiparvovec (AAV1-LPLS447X) gene therapy for lipoprotein lipase deficiency: an open-label trial. *Gene Ther.* **20,** 361–369 (2013).

35. Russell, S. *et al.* Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. *Lancet* **390,** 849–860 (2017).

36. Atchison, R. W., Casto, B. C. & Hammon, W. M. Adenovirus-Associated Virus Defective Virus Particles. *Science* **149,** 754–755 (1965).

37. Weindler, F. W. & Heilbronn, R. A Subset of Herpes Simplex Virus Replication Genes Provides Helper Functions for Productive Adeno-Associated Virus Replication. *J. Virol.* **65,** 2476–2483 (1991).

38. You, H. *et al.* Multiple human papillomavirus genes affect the adeno-associated virus life cycle. *Virology* **344,** 532–540 (2006).

39. Rose, J. A., Berns, K. I., Hoggan, M. D. & Koczot, F. J. Evidence for a single-stranded adenovirus-associated virus genome: formation of a DNA density hybrid on release of viral DNA. *Proc. Natl Acad. Sci. USA* **64,** 863–9 (1969).

40. Nault, J. C. *et al.* Recurrent AAV2-related insertional mutagenesis in human hepatocellular carcinomas. *Nat. Genet.* **47,** 1187–1193 (2015).

41. Russell, D. W. & Grompe, M. Adeno-associated virus finds its disease. *Nature Genetics* **47,** 1104–1105 (2015).

42. Logan, G. J. *et al.* Identification of liver-specific enhancer-promoter activity in the 3′ untranslated region of the wild-type AAV2 genome. *Nat. Genet.* **49,** 1267–1273 (2017).

43. Hoggan, M. D., Blacklow, N. R. & Rowe, W. P. 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 (1966).

44. Parks, W. P., Melnick, J. L., Rongey, R. & Mayor, H. D. Physical Assay and Growth Cycle Studies of a Defective Adeno-Satellite Virus. *J. Virol.* **1,** 171–180 (1967).

45. Bantel-Schaal, U. & Zur Hausen, H. Characterization of the DNA of a defective human parvovirus isolated from a genital site. *Virology* (1984). doi:10.1016/0042-6822(84)90271-X

46. Rutledge, E. A., Halbert, C. L. & Russell, D. W. Infectious Clones and Vectors Derived from Adeno-Associated Virus (AAV) Serotypes Other Than AAV Type 2. *J. Virol.* **72,** 309–319 (1998).

47. Gao, G.-P. *et al.* Novel adeno-associated viruses from rhesus monkeys as vectors for human gene therapy. *Proc. Natl. Acad. Sci.* (2002). doi:10.1073/pnas.182412299

48. Gao, G. *et al.* Clades of Adeno-Associated Viruses Are Widely Disseminated in Human Tissues. *J. Virol.* **78,** 6381–6388 (2004).

49. Schmidt, M. *et al.* Adeno-Associated Virus Type 12 (AAV12): a Novel AAV Serotype with Sialic Acid- and Heparan Sulfate Proteoglycan-Independent

Transduction Activity. *J. Virol.* (2008). doi:10.1128/JVI.02012-07

50. Bello, A. *et al.* Isolation and evaluation of novel adeno-associated virus sequences from porcine tissues. *Gene Ther.* **16,** 1320–1328 (2009).

51. Koczot, F. J., Carter, B. J., Garon, C. F. & Rose, J. A. Self-complementarity of terminal sequences within plus or minus strands of adenovirus-associated virus DNA. *Proc. Natl. Acad. Sci. U. S. A.* **70,** 215–219 (1973).

52. Balakrishnan, B. & Jayandharan, G. Basic Biology of Adeno-Associated Virus (AAV) Vectors Used in Gene Therapy. *Curr. Gene Ther.* **14,** 86–100 (2014).

53. Mendelson, E., Trempe, J. P. & Carter, B. J. Identification of the trans-acting Rep proteins of adeno-associated virus by antibodies to a synthetic oligopeptide. *J. Virol.* **60,** 823–32 (1986).

54. Janik, J. E., Huston, M. M. & Rose, J. A. Adeno-associated virus proteins: origin of the capsid components. *J. Virol.* **52,** 591–7 (1984).

55. Becerra, S. P., Koczot, F., Fabisch, P. & Rose, J. A. Synthesis of adeno-associated virus structural proteins requires both alternative mRNA splicing and alternative initiations from a single transcript. *J. Virol.* **62,** 2745–2754 (1988).

56. Sonntag, F., Schmidt, K. & Kleinschmidt, J. A. A viral assembly factor promotes AAV2 capsid formation in the nucleolus. *Proc. Natl. Acad. Sci.* **107,** 10220–10225 (2010).

57. Trempe, J. P. & Carter, B. J. Alternate mRNA splicing is required for synthesis of adeno-associated virus VPA capsid protein. *J. Virol.* **62,** 3356–3363 (1988).

58. Johnson, F. B., Ozer, H. L. & Hoggan, M. D. Structural proteins of adenovirus-associated virus type 3. *J. Virol.* **8,** 860–63 (1971).

59. Sonntag, F. *et al.* The Assembly-Activating Protein Promotes Capsid Assembly of Different Adeno-Associated Virus Serotypes. *J. Virol.* **85,** 12686–12697 (2011).

60. Earley, L. F. *et al.* 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,** 1–21 (2017).

61. Große, S. *et al.* Relevance of assembly-activating protein for Adeno-associated virus vector production and capsid protein stability in mammalian and insect cells. *J. Virol.* JVI.01198-17 (2017). doi:10.1128/JVI.01198-17

62. Tse, L. V, Moller-Tank, S., Meganck, R. M. & Asokan, A. Mapping and Engineering Functional Domains of the Assembly Activating Protein of Adeno-Associated Viruses. *J. Virol.* **92,** JVI.00393-18 (2018).

63. Maurer, A. C. *et al.* The Assembly-Activating Protein Promotes Stability and Interactions between AAV's Viral Proteins to Nucleate Capsid Assembly. *Cell Rep.* **23,** 1817–1830 (2018).

64. Miller, E. B. *et al.* Production, purification and preliminary X-ray crystallographic studies of adeno-associated virus serotype 1. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **62,** 1271–1274 (2006).

65. Kronenberg, S., Kleinschmidt, J. A. & Böttcher, B. Electron cryo-microscopy and image reconstruction of adeno-associated virus type 2 empty capsids. *EMBO Rep.* **2,** 997–1002 (2001).

66. Xie, Q. *et al.* The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. *Proc. Natl. Acad. Sci.* (2002). doi:10.1073/pnas.162250899

67. Lerch, T. F., Xie, Q. & Chapman, M. S. The structure of adeno-associated virus serotype 3B (AAV-3B): Insights into receptor binding and immune evasion. *Virology* **403,** 26–36 (2010).

68. Padron, E. *et al.* Structure of Adeno-Associated Virus Type 4. *J. Virol.* **79,** 5048–5058 (2005).

69. Govindasamy, L. *et al.* Structurally Mapping the Diverse Phenotype of Adeno-Associated Virus Serotype 4. *J. Virol.* **80,** 11556–11570 (2006).

70. Walters, R. W. *et al.* Structure of adeno-associated virus serotype 5. *J. Virol.* **78,** 3361–71 (2004).

71. Xie, Q., Ongley, H. M., Hare, J. & Chapman, M. S. Crystallization and preliminary X-ray structural studies of adeno-associated virus serotype 6. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **64,** 1074–1078 (2008).

72. Quesada, O. *et al.* Production, purification and preliminary X-ray crystallographic studies of adeno-associated virus serotype 7. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **63,** 1073–1076 (2007).

73. Nam, H.-J. *et al.* Structure of Adeno-Associated Virus Serotype 8, a Gene Therapy Vector. *J. Virol.* **81,** 12260–12271 (2007).

74. Mitchell, M. *et al.* Production, purification and preliminary X-ray crystallographic studies of adeno-associated virus serotype 9. *Acta Crystallogr. Sect. F. Struct. Biol. Cryst. Commun.* **65,** 715–8 (2009).

75. DiMattia, M. A. *et al.* Structural Insight into the Unique Properties of Adeno-Associated Virus Serotype 9. *J. Virol.* **86,** 6947–6958 (2012).

76. Tseng, Y. S. & Agbandje-McKenna, M. Mapping the AAV capsid host antibody response toward the development of second generation gene delivery vectors. *Frontiers in Immunology* **5,** (2014).

77. Bleker, S., Pawlita, M. & Kleinschmidt, J. A. Impact of Capsid Conformation and Rep-Capsid Interactions on Adeno-Associated Virus Type 2 Genome Packaging Impact of Capsid Conformation and Rep-Capsid Interactions on Adeno-Associated Virus Type 2 Genome Packaging. *J. Virol.* **80,** 810–820 (2006).

78. Girod, A. *et al.* The VP1 capsid protein of adeno-associated virus type 2 is carrying a phospholipase A2 domain required for virus infectivity. *J. Gen. Virol.* **83,** 973–978 (2002).

79. Kronenberg, S. *et al.* A Conformational Change in the Adeno-Associated Virus Type 2 Capsid Leads to the Exposure of Hidden VP1 N Termini. *J. Virol.* **79,** 5296–5303 (2005).

80. Bleker, S., Sonntag, F. & Kleinschmidt, J. A. Mutational Analysis of Narrow Pores at the Fivefold Symmetry Axes of Adeno-Associated Virus Type 2 Capsids Reveals a Dual Role in Genome Packaging and Activation of Phospholipase A2 Activity. *J. Virol.* **79,** 2528–2540 (2005).

81. Sonntag, F., Bleker, S., Leuchs, B., Fischer, R. & Kleinschmidt, J. A. 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,** 11040–11054 (2006).

82. Steinbach, S., Wistuba, A. & Bock, T. Assembly of adeno-associated virus type 2 capsids. *J. Gen. Virol.* **78,** 1453–1462 (1997).

83. Ruffing, M., Zentgraf, H. & Kleinschmidt, J. A. Assembly of viruslike particles by recombinant structural proteins of adéno-associated virus type 2 in insect cells. *J. Virol.* **66,** 6922–6930 (1992).

84. Summerford, C. & Samulski, R. J. Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions. *J. Virol.* **72,** 1438–1445 (1998).

85. Wu, Z., Miller, E., Agbandje-McKenna, M. & Samulski, R. J. alpha2,3 and alpha2,6 N-Linked Sialic Acids Facilitate Efficient Binding and Transduction by Adeno-Associated Virus Types 1 and 6. *J. Virol.* **80,** 9093–9103 (2006).

86. Ng, R. *et al.* Structural Characterization of the Dual Glycan Binding Adeno-Associated Virus Serotype 6. *J. Virol.* **84,** 12945–12957 (2010).

87. Pillay, S. *et al.* An essential receptor for adeno-associated virus infection. *Nature* **530,** 108–112 (2016).

88. Qing, K. *et al.* Human fibroblast growth factor receptor 1 is a co-receptor for infection by adeno-associated virus 2. *Nat Med* **5,** 71–77 (1999).

89. Kashiwakura, Y. *et al.* Hepatocyte Growth Factor Receptor Is a Coreceptor for Adeno-Associated Virus Type 2 Infection. *J. Virol.* **79,** 609–614 (2005).

90. Akache, B. *et al.* The 37/67-Kilodalton Laminin Receptor Is a Receptor for Adeno-Associated Virus Serotypes 8, 2, 3, and 9. *J. Virol.* **80,** 9831–9836 (2006).

91. Kurzeder, C. *et al.* CD9 promotes adeno-associated virus type 2 infection of mammary carcinoma cells with low cell surface expression of heparan sulphate proteoglycans. *Int J Mol Med* **19,** 325–333 (2007).

92. Summerford, C., Bartlett, J. S. & Samulski, R. J. AlphaVbeta5 integrin: a co-receptor for adeno-associated virus type 2 infection. *Nat. Med.* **5,** 78–82 (1999).

93. Asokan, A., Hamra, J. B., Govindasamy, L., Agbandje-McKenna, M. & Samulski, R. J. Adeno-Associated Virus Type 2 Contains an Integrin alpha5beta1 Binding Domain Essential for Viral Cell Entry. *J. Virol.* **80,** 8961–8969 (2006).

94. Rabinowitz, J. E. *et al.* Cross-packaging of a single adeno-associated virus (AAV) type 2 vector genome into multiple AAV serotypes enables transduction with broad specificity. *J. Virol.* **76,** 791–801 (2002).

95. Blackburn, S. D., Steadman, R. A. & Johnson, F. B. Attachment of adeno-associated virus type 3H to fibroblast growth factor receptor 1. *Arch. Virol.* **151,** 617–623 (2006).

96. Ling, C. *et al.* Human hepatocyte growth factor receptor is a cellular coreceptor for adeno-associated virus serotype 3. *Hum. Gene Ther.* **21,** 1741–

1747 (2010).

97. Kaludov, N., Brown, K. E., Walters, R. W., Zabner, J. & Chiorini, J. A. Adeno-associated virus serotype 4 (AAV4) and AAV5 both require sialic acid binding for hemagglutination and efficient transduction but differ in sialic acid linkage specificity. *J. Virol.* **75,** 6884–93 (2001).

98. Walters, R. W. *et al.* Binding of Adeno-associated Virus Type 5 to 2,3-Linked Sialic Acid is Required for Gene Transfer. *J. Biol. Chem.* **276,** 20610–20616 (2001).

99. Di Pasquale, G. *et al.* Identification of PDGFR as a receptor for AAV-5 transduction. *Nat. Med.* **9,** 1306–1312 (2003).

100. Weller, M. L. *et al.* Epidermal growth factor receptor is a co-receptor for adeno-associated virus serotype 6. *Nat. Med.* **16,** 662–664 (2010).

101. Shen, S., Bryant, K. D., Brown, S. M., Randell, S. H. & Asokan, A. Terminal n-linked galactose is the primary receptor for adeno-associated virus. *J. Biol. Chem.* **286,** 13532–13540 (2011).

102. Bell, C. L. *et al.* The AAV9 receptor and its modification to improve in vivo lung gene transfer in mice. *J. Clin. Invest.* **121,** 2427–2435 (2011).

103. Dudek, A. M. *et al.* An Alternate Route for Adeno-associated Virus (AAV) Entry Independent of AAV Receptor. *J. Virol.* **92,** (2018).

104. Mizukami, H., Young, N. S. & Brown, K. E. Adeno-associated virus type 2 binds to a 150-kilodalton cell membrane glycoprotein. *Virology* **217,** 124–130 (1996).

105. Pillay, S. *et al.* AAV serotypes have distinctive interactions with domains of the cellular receptor AAVR. *J. Virol.* JVI.00391-17 (2017). doi:10.1128/JVI.00391-17

106. Gurda, B. L. *et al.* Capsid Antibodies to Different Adeno-Associated Virus Serotypes Bind Common Regions. *J. Virol.* **87,** 9111–9124 (2013).

107. Rapberger, R., Lukas, A. & Mayer, B. Identification of discontinuous antigenic determinants on proteins based on shape complementarities. *J. Mol. Recognit.* **20,** 113–121 (2007).

108. Monteilhet, V., Veron, P., Leborgne, C. & Benveniste, O. 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.* **712,** 704–712 (2010).

109. Calcedo, R., Vandenberghe, L. H., Gao, G., Lin, J. & Wilson, J. M. Worldwide Epidemiology of Neutralizing Antibodies to Adeno-Associated Viruses. *J. Infect. Dis.* **199,** 381–390 (2009).

110. Wobus, C. E. *et al.* Monoclonal Antibodies against the Adeno-Associated Virus Type 2 (AAV-2) Capsid: Epitope Mapping and Identification of Capsid Domains Involved in AAV-2-Cell Interaction and Neutralization of AAV-2 Infection. *J. Virol.* **74,** 9281–9293 (2000).

111. Harbison, C. E. *et al.* Examining the cross-reactivity and neutralization mechanisms of a panel of mabs against adeno-associated virus serotypes 1 and 5. *J. Gen. Virol.* **93,** (2012).

112. Duan, D. *et al.* Dynamin Is Required for Recombinant Adeno-Associated

Virus Type 2 Infection. *J. Virol.* **73,** 10371–10376 (1999).

113. Bartlett, J. S., Wilcher, R. & Samulski, R. J. Infectious Entry Pathway of Adeno-Associated Virus and Adeno-Associated Virus Vectors. *J. Virol.* **74,** 2777–2785 (2000).

114. Nonnenmacher, M. & Weber, T. Adeno-associated virus 2 infection requires endocytosis through the CLIC/GEEC pathway. *Cell Host Microbe* **10,** 563–576 (2011).

115. Hansen, J., Qing, K., Kwon, H. J., Mah, C. & Srivastava, A. Impaired intracellular trafficking of adeno-associated virus type 2 vectors limits efficient transduction of murine fibroblasts. *J. Virol.* **74,** 992–996 (2000).

116. Ding, W., Zhang, L. N., Yeaman, C. & Engelhardt, J. F. rAAV2 traffics through both the late and the recycling endosomes in a dose-dependent fashion. *Mol. Ther.* **13,** 671–682 (2006).

117. Ding, W., Zhang, L., Yan, Z. & Engelhardt, J. F. Intracellular trafficking of adeno-associated viral vectors. *Gene Therapy* **12,** 873–880 (2005).

118. Stahnke, S. *et al.* Intrinsic phospholipase A2 activity of adeno-associated virus is involved in endosomal escape of incoming particles. *Virology* **409,** 77–83 (2011).

119. Grieger, J. C., Snowdy, S. & Samulski, R. J. Separate Basic Region Motifs within the Adeno-Associated Virus Capsid Proteins Are Essential for Infectivity and Assembly. *J. Virol.* **80,** 5199–5210 (2006).

120. Johnson, J. S. *et al.* Mutagenesis of Adeno-Associated Virus Type 2 Capsid Protein VP1 Uncovers New Roles for Basic Amino Acids in Trafficking and Cell-Specific Transduction. *J. Virol.* **84,** 8888–8902 (2010).

121. Kelich, J. M. *et al.* Super-resolution imaging of nuclear import of adeno-associated virus in live cells. *Mol. Ther. - Methods Clin. Dev.* **2,** 15047 (2015).

122. Kotin, R. M. *et al.* Site-specific integration by adeno-associated virus. *Proc. Natl. Acad. Sci. U. S. A.* **87,** 2211–5 (1990).

123. Samulski, R. J. *et al.* Targeted integration of adeno-associated virus (AAV) into human chromosome 19. *EMBO J.* **10,** 3941–50 (1991).

124. Weitzman, M. D., Kyöstiö, S. R., Kotin, R. M. & Owens, R. A. Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. *Proc. Natl. Acad. Sci. U. S. A.* **91,** 5808–5812 (1994).

125. Linden, R. M., Ward, P., Giraud, C., Winocour, E. & Berns, K. I. Site-specific integration by adeno-associated virus. *Proc Natl Acad Sci U S A* **93,** 11288–11294 (1996).

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

127. Daya, S., Cortez, N. & Berns, K. I. Adeno-Associated Virus Site-Specific Integration Is Mediated by Proteins of the Nonhomologous End-Joining Pathway. *J. Virol.* **83,** 11655–11664 (2009).

128. Nash, K., Chen, W., Salganik, M. & Muzyczka, N. Identification of Cellular Proteins That Interact with the Adeno-Associated Virus Rep Protein. *J.*

Virol. **83,** 454–469 (2009).

129. Gonçalves, M. A. F. V. Adeno-associated virus: From defective virus to effective vector. *Virology Journal* **2,** (2005).

130. Brister, J. R. & Muzyczka, N. Mechanism of Rep-mediated adeno-associated virus origin nicking. *J. Virol.* **74,** 7762–71 (2000).

131. Dubielzig, R., King, J. A., Weger, S., Kern, A. & Kleinschmidt, J. A. Adeno-associated virus type 2 protein interactions: formation of pre-encapsidation complexes. *J. Virol.* **73,** 8989–8998 (1999).

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

133. Samulski, R. J., Chang, L. S. & Shenk, T. Helper-free stocks of recombinant adeno-associated viruses: normal integration does not require viral gene expression. *J. Virol.* **63,** 3822–8 (1989).

134. Matsushita, T. *et al.* Adeno-associated virus vectors can be efficiently produced without helper virus. *Gene Ther.* **5,** 938–945 (1998).

135. Kearns, W. *et al.* Recombinant adeno-associated virus (AAV-CFTR) vectors do not integrate in a site-specific fashion in an immortalized epithelial cell line. *Gene Ther.* **3,** 748–755 (1996).

136. Nakai, H. *et al.* Extrachromosomal Recombinant Adeno-Associated Virus Vector Genomes Are Primarily Responsible for Stable Liver Transduction In Vivo. *J. Virol.* **75,** 6969–6976 (2001).

137. Riviere, C., Danos, O. & Douar, A. M. Long-term expression and repeated administration of AAV type 1, 2 and 5 vectors in skeletal muscle of immunocompetent adult mice. *Gene Ther* **13,** 1300–1308 (2006).

138. Buie, L. K. K. *et al.* Self-complementary AAV virus (scAAV) safe and long-term gene transfer in the trabecular meshwork of living rats and monkeys. *Investig. Ophthalmol. Vis. Sci.* **51,** 236–248 (2010).

139. Sehara, Y. *et al.* Persistent Expression of Dopamine-Synthesizing Enzymes 15 Years After Gene Transfer in a Primate Model of Parkinson's Disease. *Hum. Gene Ther. Clin. Dev.* **28,** 74–79 (2017).

140. Nathwani, A. C. *et al.* Long-Term Safety and Efficacy of Factor IX Gene Therapy in Hemophilia B. *N. Engl. J. Med.* **371,** 1994–2004 (2014).

141. Inagaki, K., Piao, C., Kotchey, N. M., Wu, X. & Nakai, H. Frequency and spectrum of genomic integration of recombinant adeno-associated virus serotype 8 vector in neonatal mouse liver. *J. Virol.* **82,** 9513–24 (2008).

142. Li, H. *et al.* Assessing the potential for AAV vector genotoxicity in a murine model. *Blood* **117,** 3311–3319 (2011).

143. Fisher, K. J. *et al.* Transduction with recombinant adeno-associated virus for gene therapy is limited by leading-strand synthesis. *J. Virol.* **70,** 520–532 (1996).

144. Ferrari, F. K., Samulski, T., Shenk, T. & Samulski, R. J. Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. *J. Virol.* **70,** 3227–34 (1996).

145. Wang, Z. *et al.* Rapid and highly efficient transduction by double- stranded

adeno-associated virus vectors in vitro and in vivo. *Gene Ther.* **10,** 2105–2111 (2003).

146. McCarty, D. M. *et al.* Adeno-associated virus terminal repeat (TR) mutant generates self-complementary vectors to overcome the rate-limiting step to transduction in vivo. *Gene Ther.* **10,** 2112–2118 (2003).

147. McCarty, D. M., Monahan, P. E. & Samulski, R. J. Self-complementary recombinant adeno-associated virus (scAAV) vectors promote efficient transduction independently of DNA synthesis. *Gene Ther.* **8,** 1248–1254 (2001).

148. Duan, D., Yue, Y., Yan, Z. & Engelhardt, J. F. A new dual-vector approach to enhance recombinant adeno-associated virus-mediated gene expression through intermolecular cis activation. *Nature Medicine* **6,** 595–598 (2000).

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

150. Yan, Z., Zhang, Y., Duan, D. & Engelhardt, J. F. Trans-splicing vectors expand the utility of adeno-associated virus for gene therapy. *Proc. Natl. Acad. Sci.* **97,** 6716–6721 (2000).

151. Duan, D., Yue, Y. & Engelhardt, J. F. Expanding AAV packaging capacity with Trans-splicing or overlapping vectors: A quantitative comparison. *Mol. Ther.* **4,** 383–391 (2001).

152. Keeler, A. M., ElMallah, M. K. & Flotte, T. R. Gene Therapy 2017: Progress and Future Directions. *Clin. Transl. Sci.* **10,** 242–248 (2017).

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

154. Vandenberghe, L. *et al.* Naturally occurring singleton residues in AAV capsid impact vector performance and illustrate structural constraints. *Gene Ther.* **16,** 1416–1428 (2009).

155. Perabo, L. *et al.* Combinatorial engineering of a gene therapy vector: Directed evolution of adeno-associated virus. *J. Gene Med.* **8,** 155–162 (2006).

156. Maheshri, N., Koerber, J. T., Kaspar, B. K. & Schaffer, D. V. Directed evolution of adeno-associated virus yields enhanced gene delivery vectors. *Nat. Biotechnol.* **24,** 198–204 (2006).

157. Maersch, S., Huber, A., Büning, H., Hallek, M. & Perabo, L. Optimization of stealth adeno-associated virus vectors by randomization of immunogenic epitopes. *Virology* **397,** 167–175 (2010).

158. Pulicherla, N. *et al.* Engineering liver-detargeted AAV9 vectors for cardiac and musculoskeletal gene transfer. *Mol. Ther.* **19,** 1070–1078 (2011).

159. Zhong, L. *et al.* Next generation of adeno-associated virus 2 vectors: point mutations in tyrosines lead to high-efficiency transduction at lower doses. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 7827–32 (2008).

160. Petrs-Silva, H. *et al.* High-efficiency transduction of the mouse retina by tyrosine-mutant AAV serotype vectors. *Mol. Ther.* **17,** 463–471 (2009).

161. Li, M. *et al.* High-Efficiency Transduction of Fibroblasts and Mesenchymal

Stem Cells by Tyrosine-Mutant AAV2 Vectors for Their Potential Use in Cellular Therapy. *Hum. Gene Ther.* **21,** 1527–1543 (2010).

162. Markusic, D. M. *et al.* High-efficiency transduction and correction of murine hemophilia B using AAV2 vectors devoid of multiple surface-exposed tyrosines. *Mol. Ther.* **18,** 2048–2056 (2010).

163. Ku, C. A. *et al.* Gene therapy using self-complementary Y733F capsid mutant AAV2/8 restores vision in a model of early onset Leber congenital amaurosis. *Hum. Mol. Genet.* **20,** 4569–4581 (2011).

164. Mao, Y. *et al.* Single point mutation in adeno-associated viral vectors -DJ capsid leads to improvement for gene delivery in vivo. *BMC Biotechnol.* **16,** 1 (2016).

165. Limberis, M. P., Vandenberghe, L. H., Zhang, L., Pickles, R. J. & Wilson, J. M. Transduction efficiencies of novel AAV vectors in mouse airway epithelium in vivo and human ciliated airway epithelium in vitro. *Mol. Ther.* **17,** 294–301 (2009).

166. Santiago-Ortiz, J. *et al.* AAV ancestral reconstruction library enables selection of broadly infectious viral variants. *Gene Ther.* **22,** 934–946 (2015).

167. Zinn, E. *et al.* In silico reconstruction of the viral evolutionary lineage yields a potent gene therapy vector. *Cell Rep.* **12,** 1056–1068 (2015).

168. Suzuki, J., Hashimoto, K., Xiao, R., Vandenberghe, L. H. & Liberman, M. C. Cochlear gene therapy with ancestral AAV in adult mice: Complete transduction of inner hair cells without cochlear dysfunction. *Sci. Rep.* **7,** (2017).

169. Landegger, L. D. *et al.* A synthetic AAV vector enables safe and efficient gene transfer to the mammalian inner ear. *Nat. Biotechnol.* **35,** 280–284 (2017).

170. Pan, B. *et al.* Gene therapy restores auditory and vestibular function in a mouse model of Usher syndrome type 1c. *Nat. Biotechnol.* **35,** 264–272 (2017).

171. Stemmer, W. P. DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc. Natl. Acad. Sci.* **91,** 10747–10751 (1994).

172. Grimm, D. *et al.* In Vitro and In Vivo Gene Therapy Vector Evolution via Multispecies Interbreeding and Retargeting of Adeno-Associated Viruses. *J. Virol.* **82,** 5887–5911 (2008).

173. Li, W. *et al.* Engineering and selection of shuffled AAV genomes: A new strategy for producing targeted biological nanoparticles. *Mol. Ther.* **16,** 1252–1260 (2008).

174. Koerber, J. T., Jang, J. H. & Schaffer, D. V. DNA shuffling of adeno-associated virus yields functionally diverse viral progeny. *Mol. Ther.* **16,** 1703–1709 (2008).

175. Lerch, T. F. *et al.* Structure of AAV-DJ, a retargeted gene therapy vector: Cryo-electron microscopy at 4.5 Å resolution. *Structure* **20,** 1310–1320 (2012).

176. Melo, S. P. *et al.* Somatic correction of junctional epidermolysis bullosa by a highly recombinogenic AAV variant. *Mol. Ther.* **22,** 725–733 (2014).

177. Lakhan, R. *et al.* Local administration of AAV-DJ pseudoserotype

expressing COX2 provided early onset of transgene expression and promoted bone fracture healing in mice. *Gene Ther.* **22,** 721–728 (2015).

178. Liu, J. & Moon, Y. A. Simple purification of adeno-associated virus-DJ for liver-specific gene expression. *Yonsei Med. J.* **57,** 790–794 (2016).

179. Klimczak, R. R., Koerber, J. T., Dalkara, D., Flannery, J. G. & Schaffer, D. V. A novel adeno-associated viral variant for efficient and selective intravitreal transduction of rat Müller cells. *PLoS One* **4,** (2009).

180. Ward, P. & Walsh, C. E. Chimeric AAV Cap sequences alter gene transduction. *Virology* **386,** 237–248 (2009).

181. Koerber, J. T. *et al.* Molecular evolution of adeno-associated virus for enhanced glial gene delivery. *Mol. Ther.* **17,** 2088–2095 (2009).

182. Excoffon, K. J. D. A. *et al.* Directed evolution of adeno-associated virus to an infectious respiratory virus. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 3865–70 (2009).

183. Gray, S. J. *et al.* Directed evolution of a novel adeno-associated virus (AAV) vector that crosses the seizure-compromised blood-brain barrier (BBB). *Mol. Ther.* **18,** 570–578 (2010).

184. Jang, J. H. *et al.* An evolved adeno-associated viral variant enhances gene delivery and gene targeting in neural stem cells. *Mol. Ther.* **19,** 667–675 (2011).

185. Asuri, P. *et al.* Directed Evolution of Adeno-associated Virus for Enhanced Gene Delivery and Gene Targeting in Human Pluripotent Stem Cells. *Mol. Ther.* **20,** 329–338 (2012).

186. Dalkara, D. *et al.* In Vivo – Directed Evolution of a New Adeno-Associated Virus for Therapeutic Outer Retinal Gene Delivery from the Vitreous. *Sci. Transl. Med.* **5,** (2013).

187. Powell, S. K. *et al.* Characterization of a novel adeno-associated viral vector with preferential oligodendrocyte tropism. *Gene Ther.* **23,** 807–814 (2016).

188. Albright, B. H. *et al.* Mapping the Structural Determinants Required for AAVrh.10 Transport across the Blood-Brain Barrier. *Mol. Ther.* (2018). doi:10.1016/j.ymthe.2017.10.017

189. Yang, L. *et al.* A myocardium tropic adeno-associated virus (AAV) evolved by DNA shuffling and in vivo selection. *Proc. Natl. Acad. Sci.* **106,** 3946–3951 (2009).

190. Choudhury, S. R. *et al.* In vivo selection yields AAV-B1 capsid for CNS and muscle gene therapy. *Mol. Ther.* **24,** 1247–1257 (2016).

191. Lisowski, L. *et al.* Selection and evaluation of clinically relevant AAV variants in a xenograft liver model. *Nature* **506,** 382–386 (2014).

192. Paulk, N. K. *et al.* Bioengineered AAV Capsids with Combined High Human Liver Transduction In Vivo and Unique Humoral Seroreactivity. *Mol. Ther.* **26,** 289–303 (2018).

193. Girod, A. *et al.* Genetic capsid modifications allow efficient re-targeting of adeno- associated virus type 2. *Nat. Med.* **5,** 1052–1056 (1999).

194. Grifman, M. *et al.* Incorporation of tumor-targeting peptides into recombinant adeno-associated virus capsids. *Mol. Ther.* **3,** 964–975 (2001).

195. Nicklin, S. A. *et al.* Efficient and selective AAV2-mediated gene transfer directed to human vascular endothelial cells. *Mol. Ther.* **4,** 174–81 (2001).

196. Shi, W. & Bartlett, J. S. RGD inclusion in VP3 provides adeno-associated virus type 2 (AAV2)-based vectors with a heparan sulfate-independent cell entry mechanism. *Mol. Ther.* **7,** 515–525 (2003).

197. Work, L. M. *et al.* Development of efficient viral vectors selective for vascular smooth muscle cells. *Mol. Ther.* **9,** 198–208 (2004).

198. White, S. J. *et al.* Targeted Gene Delivery to Vascular Tissue In Vivo by Tropism-Modified Adeno-Associated Virus Vectors. *Circulation* **109,** 513–519 (2004).

199. Work, L. M. *et al.* Vascular Bed-Targeted in Vivo Gene Delivery Using Tropism-Modified Adeno-associated Viruses. *Mol. Ther.* **13,** 683–693 (2006).

200. White, K. *et al.* Engineering adeno-associated virus 2 vectors for targeted gene delivery to atherosclerotic lesions. *Gene Ther.* **15,** 443–451 (2008).

201. Yu, C. Y. *et al.* A muscle-targeting peptide displayed on AAV2 improves muscle tropism on systemic delivery. *Gene Ther.* **16,** 953–962 (2009).

202. Kunze, C. *et al.* Synthetic AAV/CRISPR vectors for blocking HIV-1 expression in persistently infected astrocytes. *Glia* **66,** 413–427 (2018).

203. Opie, S. R., Warrington, K. H., Agbandje-McKenna, M., Zolotukhin, S. & Muzyczka, N. 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 (2003).

204. Kern, A. *et al.* Identification of a Heparin-Binding Motif on Adeno-Associated Virus Type 2 Capsids. *J. Virol.* **77,** 11072–11081 (2003).

205. Michelfelder, S. *et al.* Peptide ligands incorporated into the threefold spike capsid domain to re-direct gene transduction of AAV8 and AAV9 in vivo. *PLoS One* **6,** (2011).

206. Perabo, L. *et al.* In vitro selection of viral vectors with modified tropism: The adeno-associated virus display. *Mol. Ther.* **8,** 151–157 (2003).

207. Müller, O. J. *et al.* Random peptide libraries displayed on adeno-associated virus to select for targeted gene therapy vectors. *Nat. Biotechnol.* **21,** 1040–1046 (2003).

208. Körbelin, J. *et al.* Pulmonary Targeting of Adeno-associated Viral Vectors by Next-generation Sequencing-guided Screening of Random Capsid Displayed Peptide Libraries. *Mol. Ther.* **24,** 1050–1061 (2016).

209. Waterkamp, D. A., Müller, O. J., Ying, Y., Trepel, M. & Kleinschmidt, J. A. Isolation of targeted AAV2 vectors from novel virus display libraries. *J. Gene Med.* **8,** 1307–1319 (2006).

210. Michelfelder, S. *et al.* Vectors selected from adeno-associated viral display peptide libraries for leukemia cell-targeted cytotoxic gene therapy. *Exp. Hematol.* **35,** 1766–1776 (2007).

211. Stiefelhagen, M. *et al.* Application of a haematopoetic progenitor cell-targeted adeno-associated viral (AAV) vector established by selection of an AAV random peptide library on a leukaemia cell line. *BioMed Cent.* **6,** (2008).

212. Sellner, L. *et al.* Generation of efficient human blood progenitor-targeted recombinant adeno-associated viral vectors (AAV) by applying an AAV random peptide library on primary human hematopoietic progenitor cells. *Exp. Hematol.* **36,** 957–964 (2008).

213. Michelfelder, S. *et al.* Successful expansion but not complete restriction of tropism of adeno-associated virus by in vivo biopanning of random virus display peptide libraries. *PLoS One* **4,** (2009).

214. Ying, Y. *et al.* Heart-targeted adeno-associated viral vectors selected by in vivo biopanning of a random viral display peptide library. *Gene Ther.* (2010). doi:10.1038/gt.2010.44

215. Varadi, K. *et al.* Novel random peptide libraries displayed on AAV serotype 9 for selection of endothelial cell-directed gene transfer vectors. *Gene Ther.* **19,** 800–809 (2012).

216. Sallach, J. *et al.* Tropism-modified AAV vectors overcome barriers to successful cutaneous therapy. *Mol. Ther.* **22,** 929–939 (2014).

217. Cronin, T. *et al.* Efficient transduction and optogenetic stimulation of retinal bipolar cells by a synthetic adeno-associated virus capsid and promoter. *EMBO Mol Med* **6,** 1175–1190 (2014).

218. Körbelin, J. *et al.* A brain microvasculature endothelial cell-specific viral vector with the potential to treat neurovascular and neurological diseases. *EMBO Mol Med* **8,** 1409–1420 (2016).

219. Deverman, B. E. *et al.* Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. *Nat. Biotechnol.* **34,** 204–209 (2016).

220. Chan, K. Y. *et al.* Engineered AAVs for efficient noninvasive gene delivery to the central and peripheral nervous systems. *Nat. Neurosci.* **20,** 1172–1179 (2017).

221. Rhoads, A. & Au, K. F. PacBio Sequencing and Its Applications. *Genomics, Proteomics and Bioinformatics* **13,** 278–289 (2015).

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

223. Marsic, D. *et al.* Vector Design Tour de Force Integrating combinatorial and rational approchaes to derive novel adeno associated virus variants. **22,** 1900–1909 (2014).

224. Marsic, D., Méndez-Gómez, H. R. & Zolotukhin, S. High-accuracy biodistribution analysis of adeno-associated virus variants by double barcode sequencing. *Mol. Ther. - Methods Clin. Dev.* **2,** 15041 (2015).

225. Davidsson, M. *et al.* A novel process of viral vector barcoding and library preparation enables high-diversity library generation and recombination-free paired-end sequencing. *Sci. Rep.* **6,** (2016).

226. Nakai, H. *et al.* Unrestricted Hepatocyte Transduction with Adeno-Associated Virus Serotype 8 Vectors in Mice. *J. Virol.* **79,** 214–224 (2005).

227. Davidoff, A. M. *et al.* Comparison of the ability of adeno-associated viral vectors pseudotyped with serotype 2, 5, and 8 capsid proteins to mediate efficient transduction of the liver in murine and nonhuman primate models.

Mol. Ther. **11,** 875–888 (2005).

228. Bello, A. *et al.* Novel adeno-associated viruses derived from pig tissues transduce most major organs in mice. *Sci. Rep.* **4,** (2014).

229. Wang, D., Zhong, L., Nahid, M. A. & Gao, G. The potential of adeno-associated viral vectors for gene delivery to muscle tissue. *Expert Opin Drug Deliv* **11,** 345–364 (2014).

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

231. Inagaki, K. *et al.* Robust systemic transduction with AAV9 vectors in mice: efficient global cardiac gene transfer superior to that of AAV8. *Mol. Ther.* (2006). doi:10.1016/j.ymthe.2006.03.014

232. Bostick, B., Ghosh, A., Yue, Y., Long, C. & Duan, D. Systemic AAV-9 transduction in mice is influenced by animal age but not by the route of administration. *Gene Ther.* (2007). doi:10.1038/sj.gt.3303029

233. Bish, L. T. *et al.* Adeno-Associated Virus (AAV) Serotype 9 Provides Global Cardiac Gene Transfer Superior to AAV1, AAV6, AAV7, and AAV8 in the Mouse and Rat. *Hum. Gene Ther.* (2008). doi:10.1089/hum.2008.123

234. Xu, L. *et al.* Adeno-associated Virus 9 mediated FKRP gene therapy restores functional glycosylation of α-dystroglycan and improves muscle functions. *Mol. Ther.* (2013). doi:10.1038/mt.2013.156

235. Schmidt, E. V, Christoph, G., Zeller, R. & Leder, P. The cytomegalovirus enhancer: a pan-active control element in transgenic mice. *Mol. Cell. Biol.* **10,** 4406–4411 (1990).

236. Kouadjo, K. E., Nishida, Y., Cadrin-Girard, J. F., Yoshioka, M. & St-Amand, J. Housekeeping and tissue-specific genes in mouse tissues. *BMC Genomics* **8,** (2007).

237. Zeng, J. *et al.* Identification and analysis of house-keeping and tissue-specific genes based on RNA-seq data sets across 15 mouse tissues. *Gene* **576,** 560–570 (2016).

238. Löser, P., Jennings, G. S., Strauss, M. & Sandig, V. Reactivation of the previously silenced cytomegalovirus major immediate-early promoter in the mouse liver: involvement of NFkappaB. *J. Virol.* **72,** 180–90 (1998).

239. Matsuzaki, Y. *et al.* Intravenous administration of the adeno-associated virus-PHP.B capsid fails to upregulate transduction efficiency in the marmoset brain. *Neurosci. Lett.* **665,** 182–188 (2018).

240. Rincon, M. Y. *et al.* Widespread transduction of astrocytes and neurons in the mouse central nervous system after systemic delivery of a self-complementary AAV-PHP.B vector. *Gene Therapy* 1–10 (2018). doi:10.1038/s41434-018-0005-z

241. Hordeaux, J. *et al.* The Neurotropic Properties of AAV-PHP.B Are Limited to C57BL/6J Mice. *Molecular Therapy* **26,** 664–668 (2018).

242. Nathwani, A. *et al.* Enhancing transduction of the liver by adeno-associated viral vectors. *Gene Ther.* **16,** 60–9 (2009).

243. Ellis, B. L. *et al.* A survey of ex vivo/in vitro transduction efficiency of

mammalian primary cells and cell lines with Nine natural adeno-associated virus (AAV1-9) and one engineered adeno-associated virus serotype. *Virology Journal* **10,** (2013).

244. Zhong, L. *et al.* Tyrosine-phosphorylation of AAV2 vectors and its consequences on viral intracellular trafficking and transgene expression. *Virology* **381,** 194–202 (2008).

245. Clément, N. & Grieger, J. C. Manufacturing of recombinant adeno-associated viral vectors for clinical trials. *Mol. Ther. - Methods Clin. Dev.* **3,** 16002 (2016).

246. Plow, E. F., Haas, T. A., Zhang, L., Loftus, J. & Smith, J. W. Ligand binding to integrins. *Journal of Biological Chemistry* **275,** 21785–21788 (2000).

247. Ruoslahti, E. RGD and other recognition sequences for integrins. *Annu. Rev. Cell Dev. Biol.* **12,** 697–715 (1996).

SUPPLEMENTARY INFORMATION

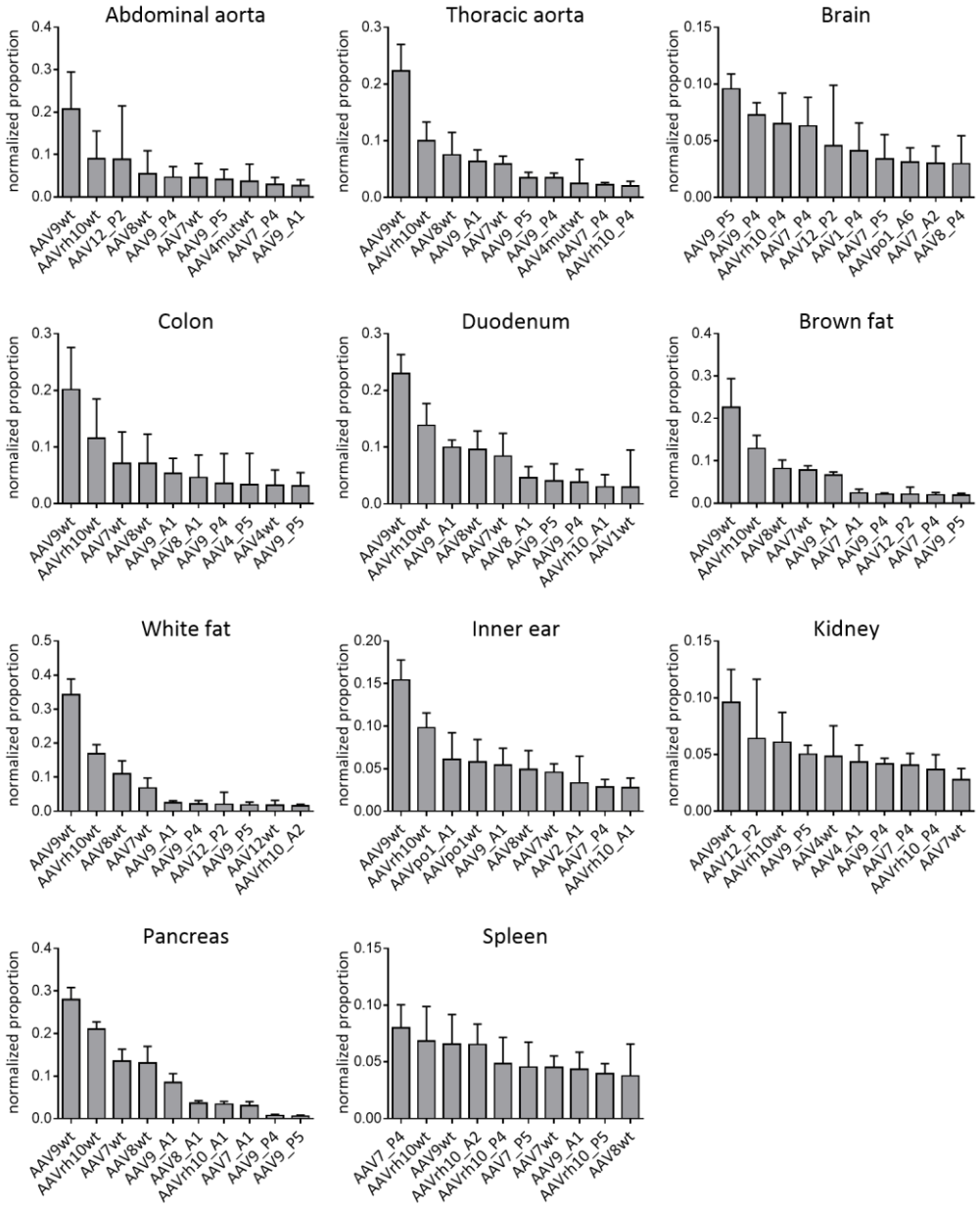

Figure 30: Transcriptional efficiency in various tissues

Bar plots show the transcriptional efficiency as normalized proportion of the top 10 AAV variants of the 1st generation library in the abdominal aorta, thoracic aorta, brain, colon, duodenum, brown fat, white fat, inner ear, kidney, pancreas and spleen. The cDNA values are the average from six C57BL/6J mice with SD.

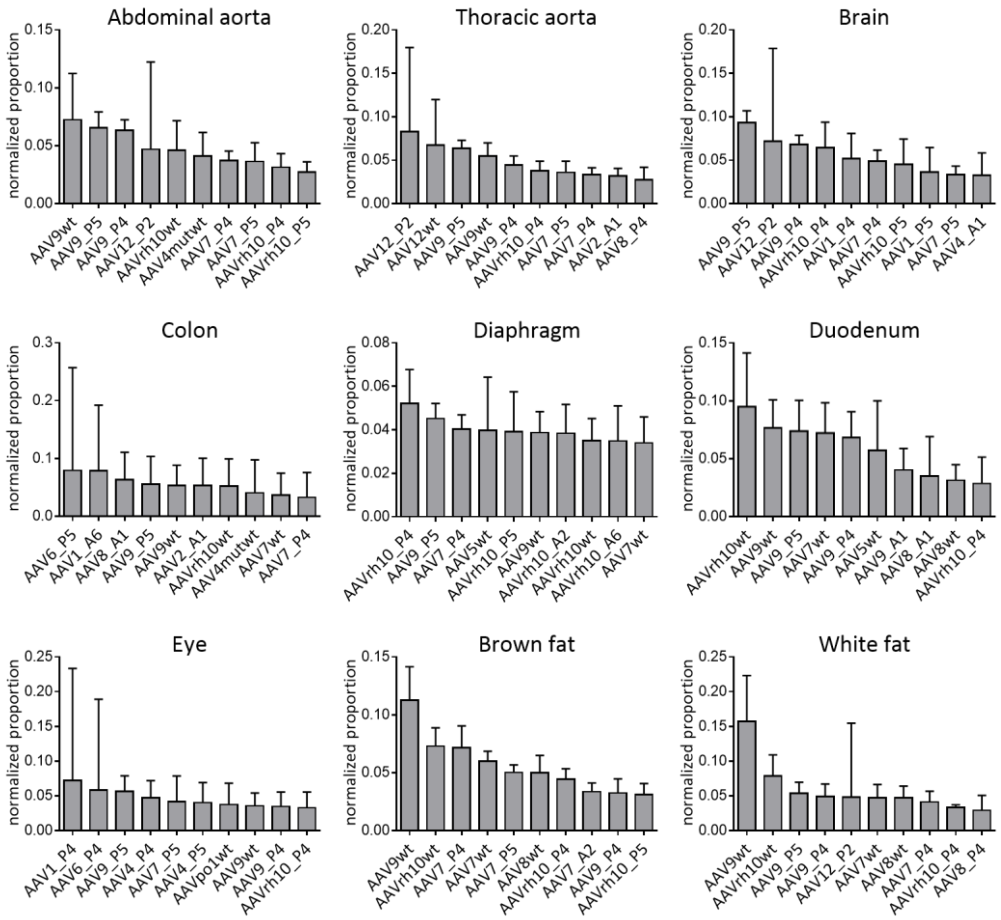

Figure 31: Transduction efficiency in various tissues

Bar plots show the transduction efficiency as normalized proportion of the top 10 AAV variants of the 1st generation library in the abdominal aorta, thoracic aorta, brain, colon, diaphragm, duodenum, eye, brown fat and white fat. The gDNA values are the average from six C57BL/6J mice with SD.

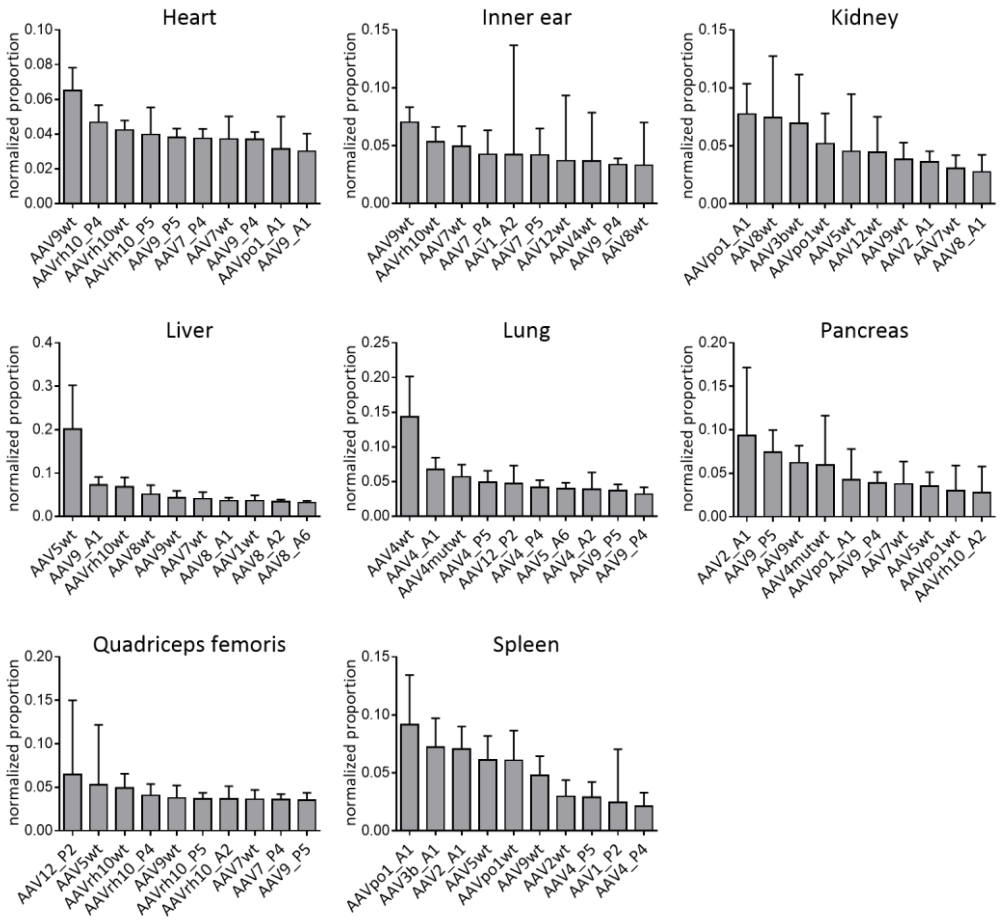

Figure 32: Transduction efficiency in various tissues

Bar plots show the transduction efficiency as normalized proportion of the top 10 AAV variants of the 1st generation library in the heart, inner ear, kidney, liver, lung, pancreas, quadriceps femoris and spleen. The gDNA values are the average from six C57BL/6J mice with SD.

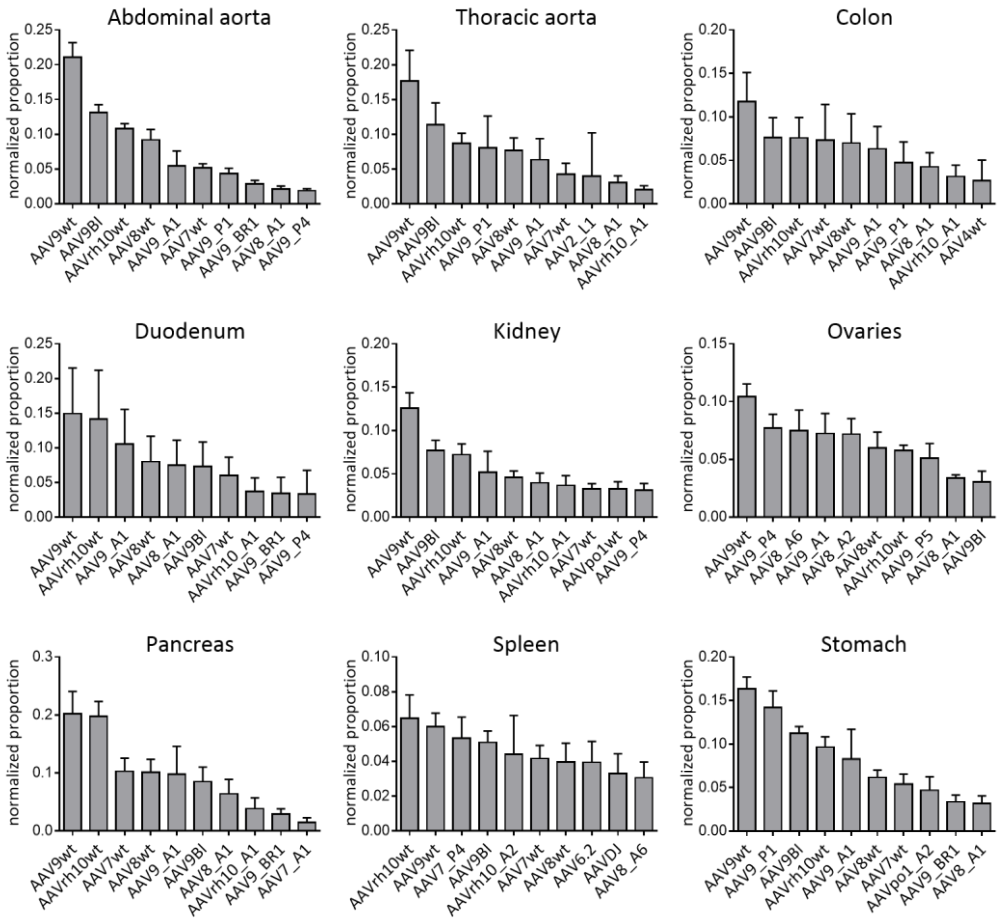

Figure 33: Transcriptional efficiency in various tissues
Bar plots show the transcriptional efficiency as normalized proportion of the top 10 AAV variants of the 2nd generation library in the abdominal aorta, thoracic aorta, brain, colon, duodenum, brown fat, white fat, inner ear, kidney, pancreas and spleen. The cDNA values are the average from six C57BL/6J mice with SD.

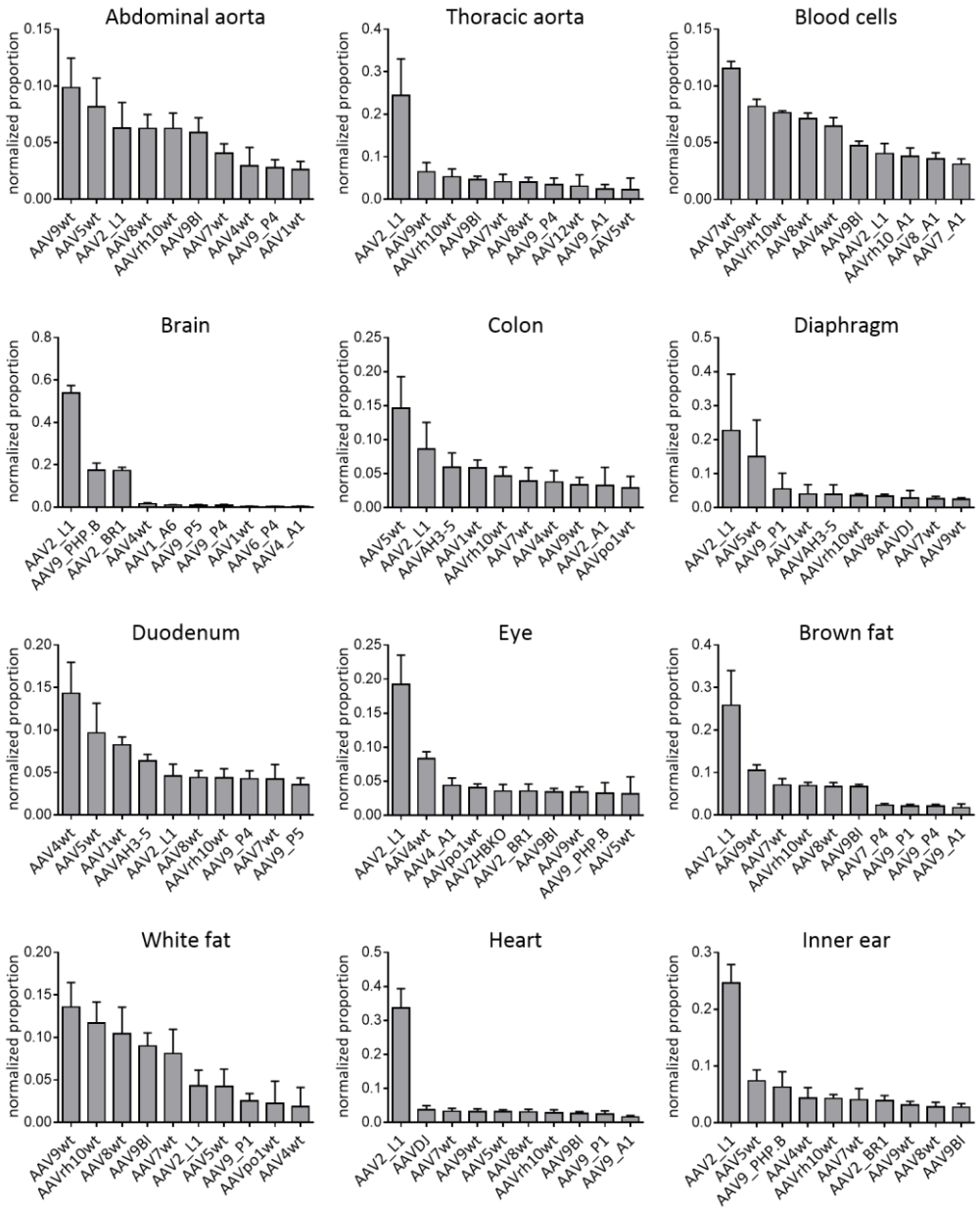

Figure 34: Transduction efficiency in various tissues

Bar plots show the transduction efficiency as normalized proportion of the top 10 AAV variants of the 2nd generation library in the abdominal aorta, thoracic aorta, blood cells, brain, colon, diaphragm, duodenum, eye, brown fat, white fat, heart and inner ear. The gDNA values are the average from six C57BL/6J mice with SD.

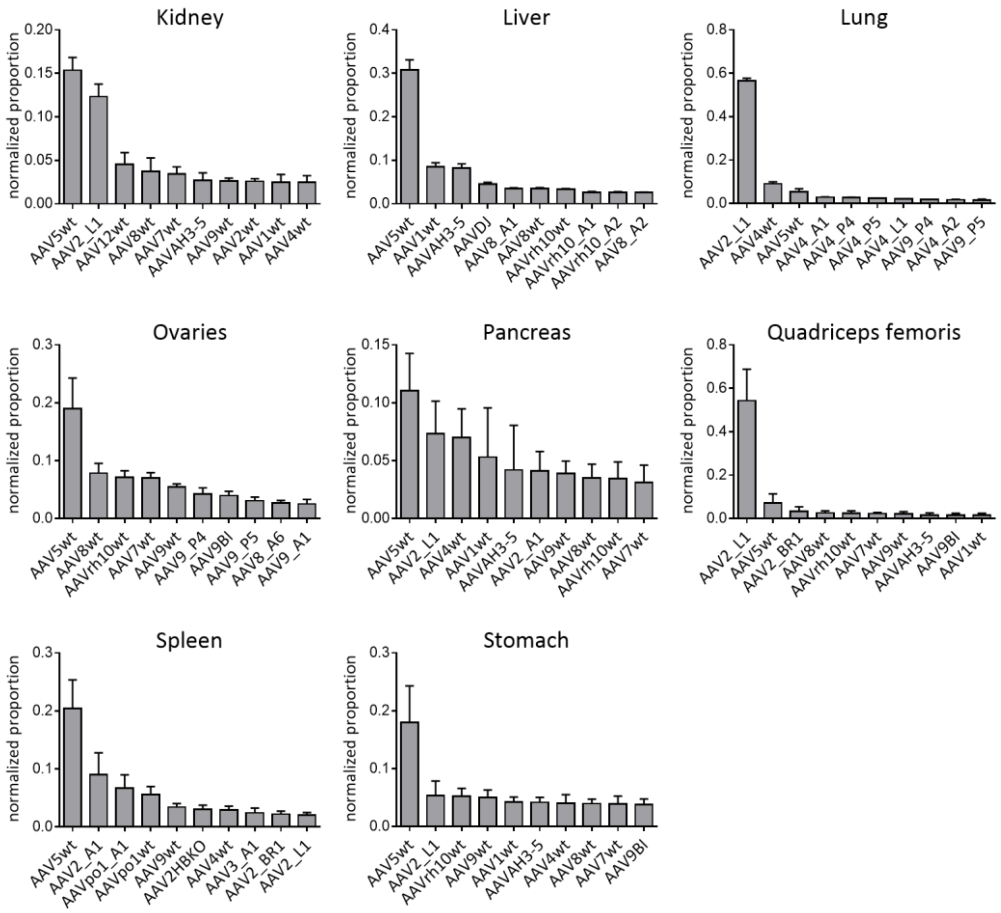

Figure 35: Transduction efficiency in various tissues

Bar plots show the transduction efficiency as normalized proportion of the top 10 AAV variants of the 2nd generation library in the kidney, liver, lung, ovaries, pancreas, quadriceps femoris, spleen, stomach. The gDNA values are the average from six C57BL/6J mice with SD.

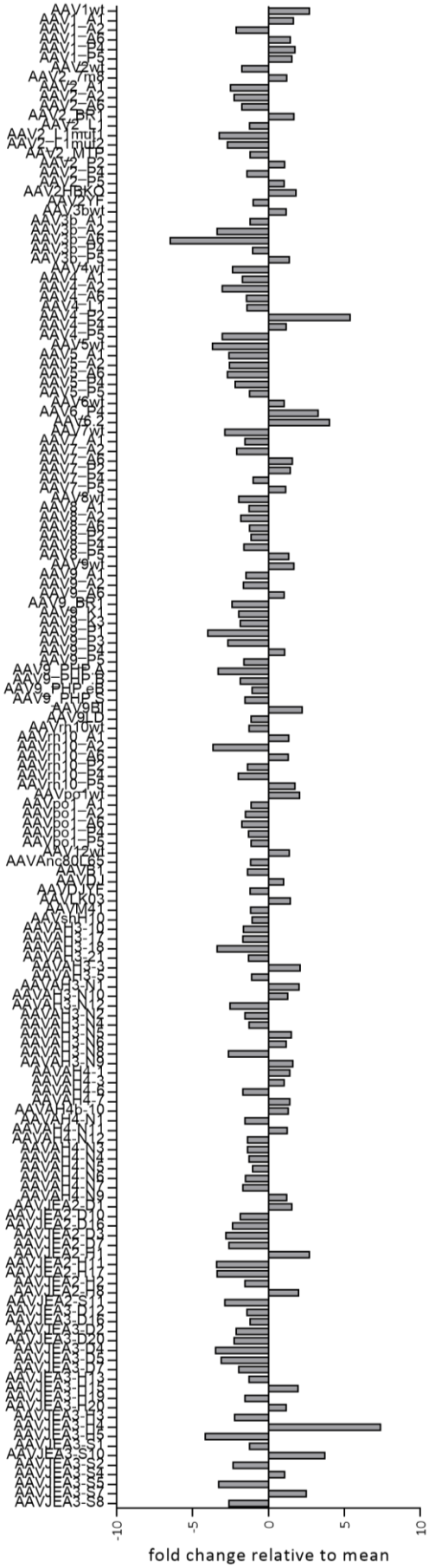

Figure 36: Composition of 3rd generation library

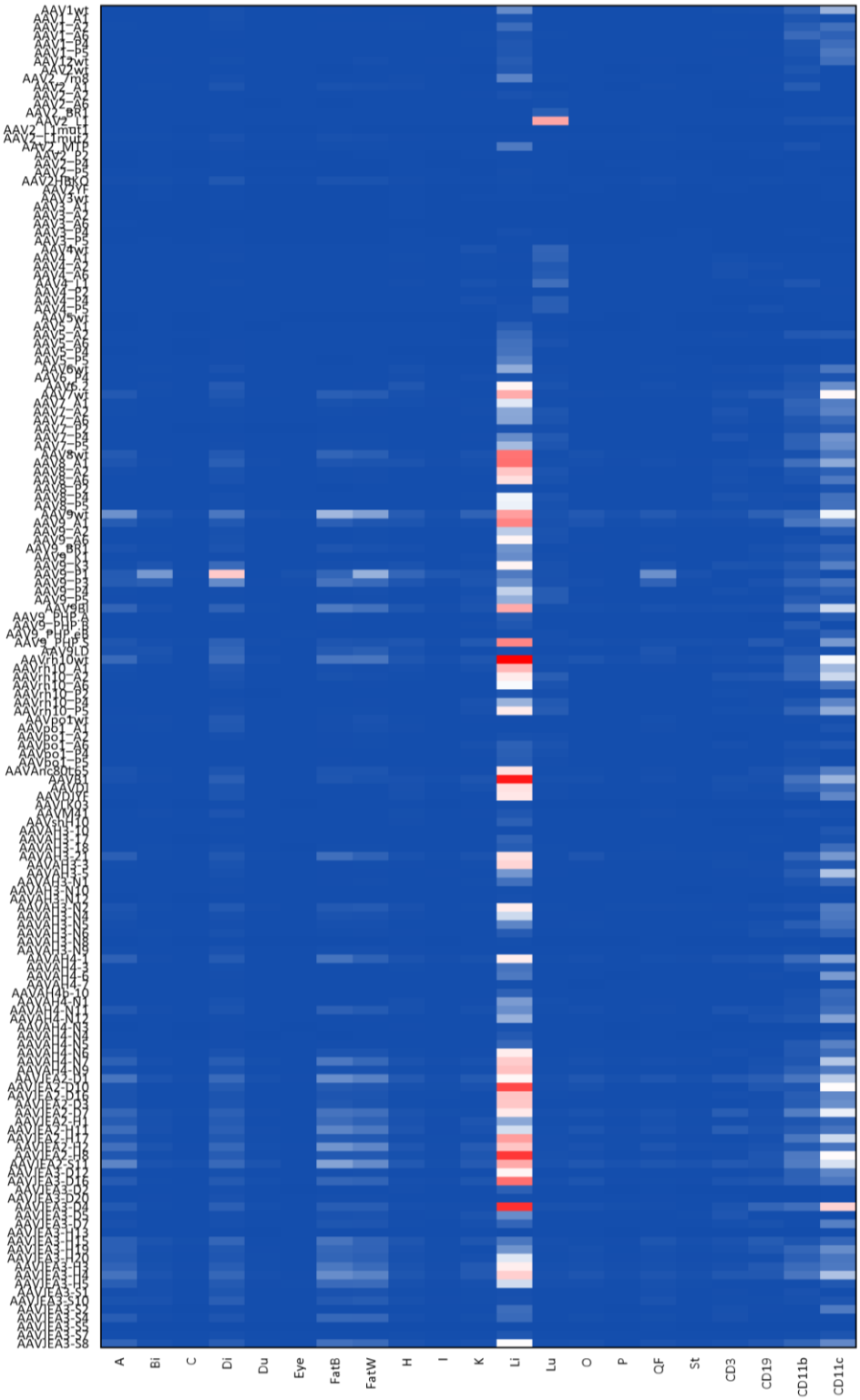

Figure 37: Biodistribution of all variants of the 3rd generation library

Calculated $B_{\alpha\beta}$ values are depicted as a heat map simultaneously illustrating the transcriptional efficiency and specificity of all variants in the 3rd generation library in the aorta (A), biceps (Bi), colon (C), diaphragm (Di), duodenum (Du), eye, brown fat (FatB), white fat (FatW), heart (H), inner ear (I), kidney (K), liver (Li), lung (Lu), ovaries (O), pancreas (P), quadriceps femoris (QF), stomach (St) and CD3-, CD19-, CD11b- as well as CD11c-positive cells. A logarithmic scale is used with blue representing the value 0, white 0.37 and red 3.71.

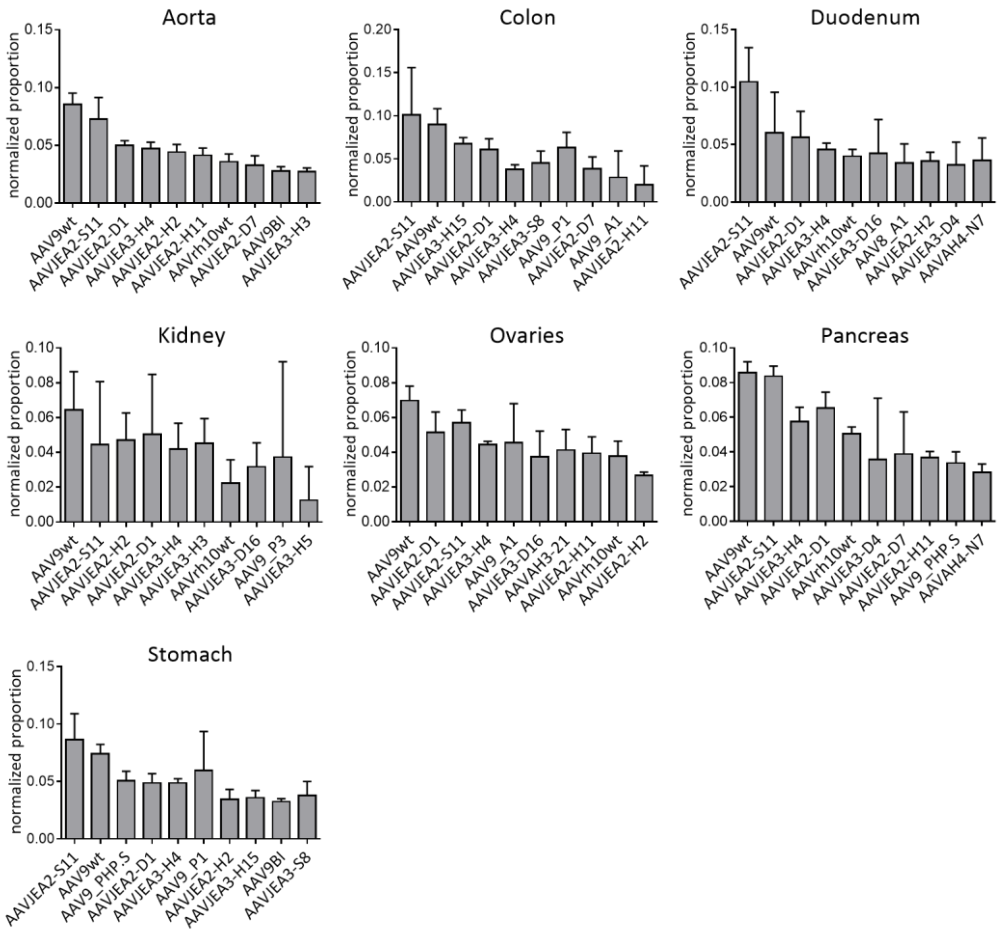

Figure 38: Transcriptional efficiency in various tissues

Bar plots show the transcriptional efficiency as normalized proportion of the top 10 AAV variants of the 3rd generation library in the aorta, colon, duodenum, kidney, ovaries, pancreas and stomach. The cDNA values are the average from four C57BL/6J mice with SD.

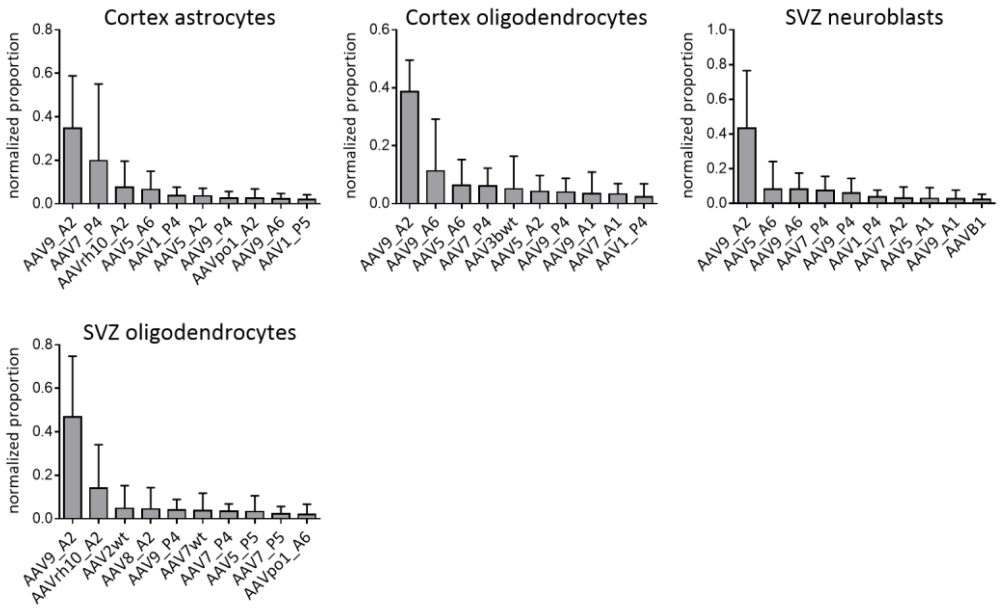

Figure 39: Transcriptional efficiency in brain cells
Bar plots show the transcriptional efficiency as normalized proportion of the top 10 AAV variants of the 3rd generation library in the astrocytes and oligodendrocytes of the cortex as well as the neuroblasts and oligodendrocytes in the subventricular zone (SVZ). The cDNA values are the average from six C57BL/6J mice with SD.

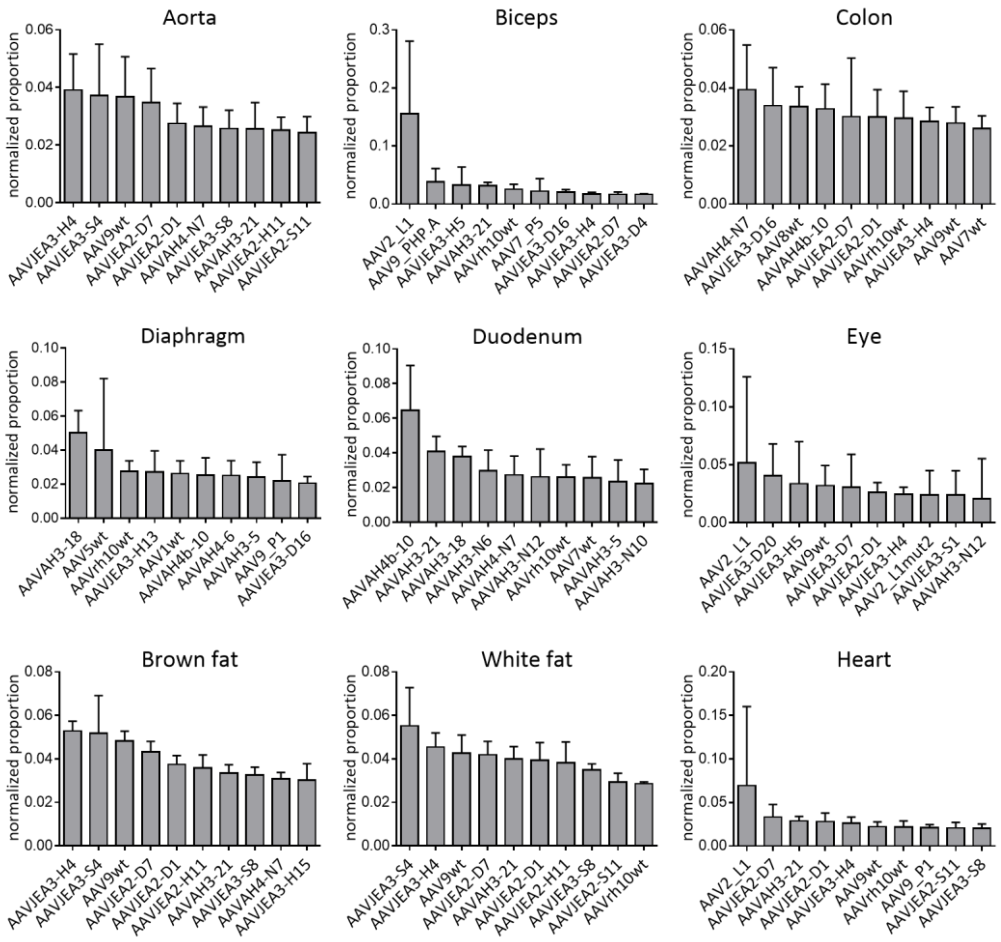

Figure 40: Transduction efficiency in various tissues

Bar plots show the transduction efficiency as normalized proportion of the top 10 AAV variants of the 3rd generation library in the aorta, biceps, colon, diaphragm, duodenum, eye, brown fat, white fat and heart. The gDNA values are the average from four C57BL/6J mice with SD.

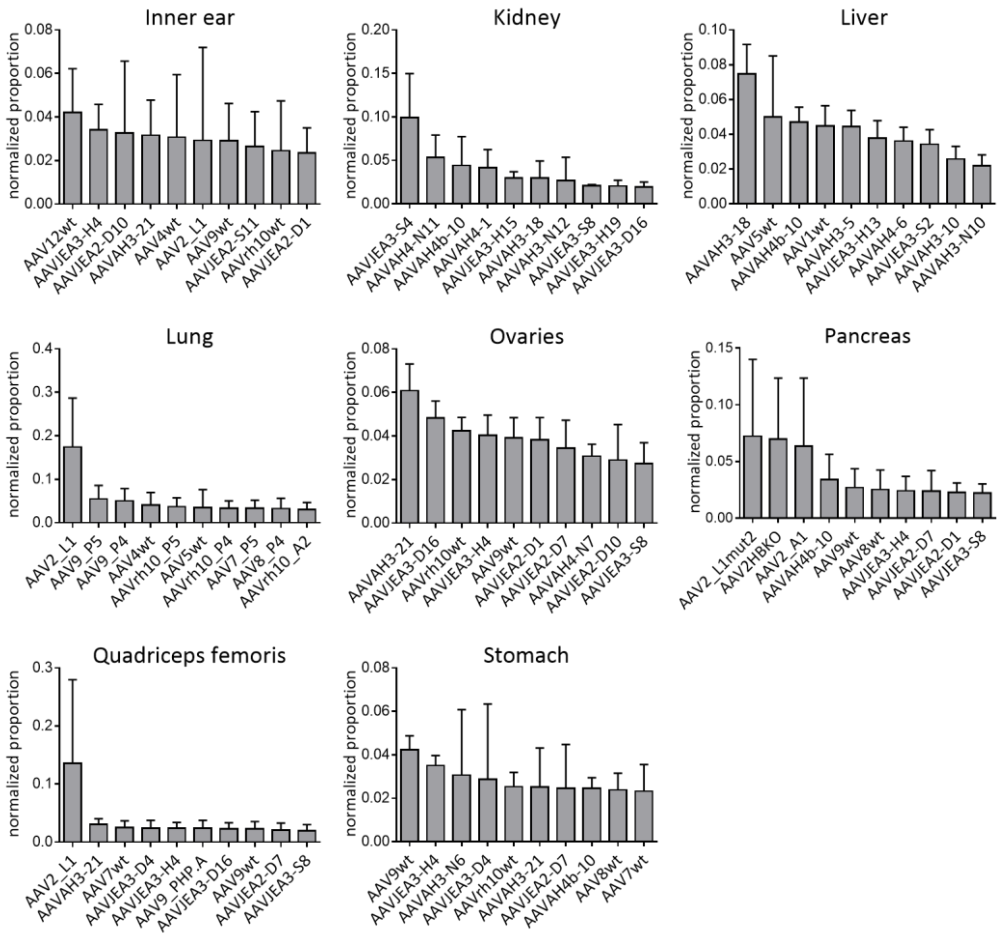

Figure 41: Transduction efficiency in various tissues

Bar plots show the transduction efficiency as normalized proportion of the top 10 AAV variants of the 3rd generation library in the inner ear, kidney, liver, lung, ovaries, pancreas, quadriceps femoris and stomach. The gDNA values are the average from four C57BL/6J mice with SD.

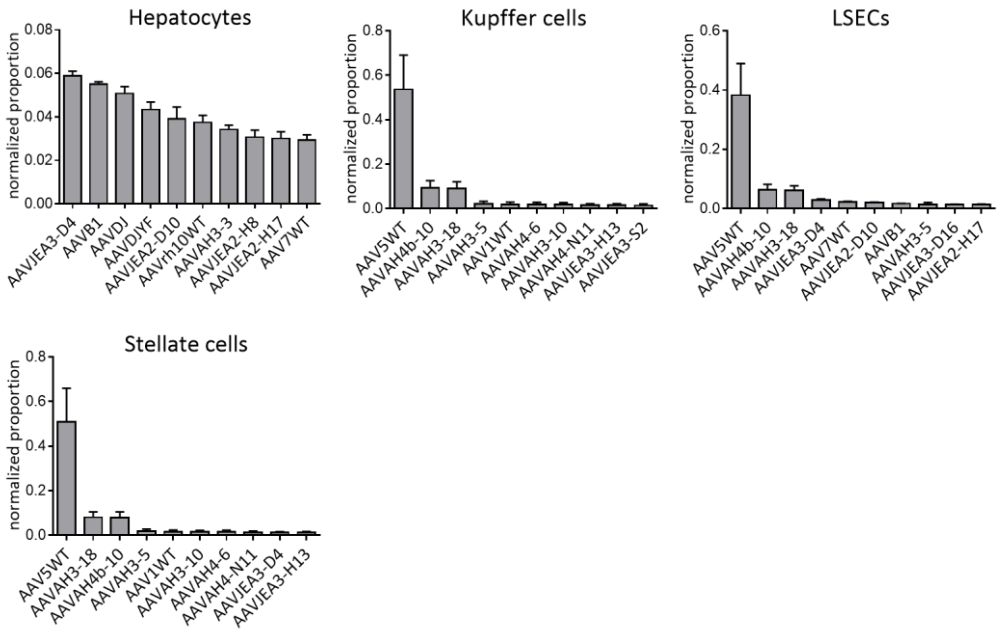

Figure 42: Transduction efficiency in liver cell types

Bar plots show the transduction efficiency as normalized proportion of the top 10 AAV variants of the 3rd generation library in hepatocytes, Kupffer cells, liver sinusoidal endothelial cells (LSECs) and stellate cells. The gDNA values are the average from four BALB/c mice with SD.

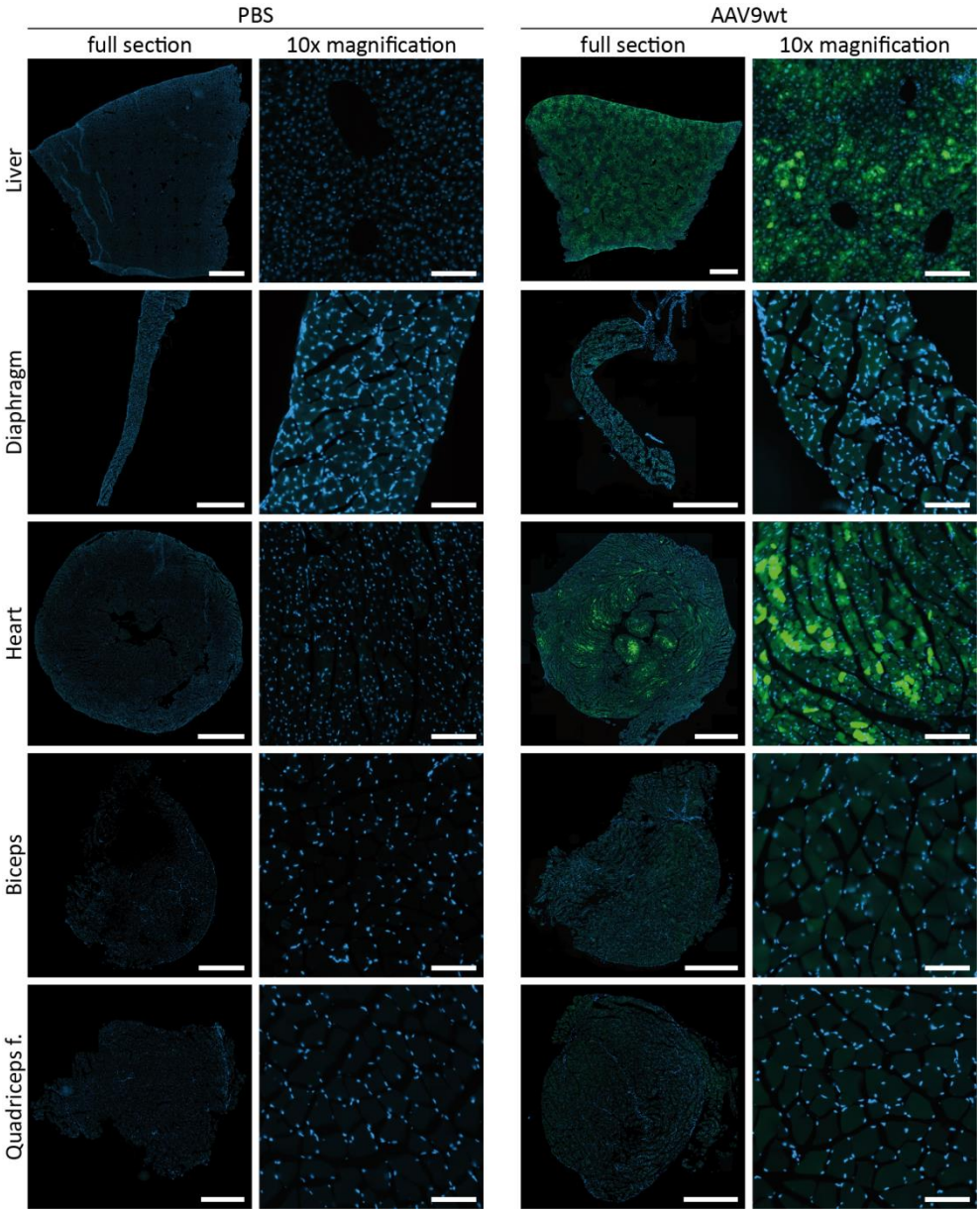

Figure 43: EGFP fluorescence of PBS and AAV9wt group
Images show 10 µm cryosections of the liver, diaphragm, heart, biceps and quadriceps femoris. Representative sections were chosen from C57BL/6J mice injected with $5x10^{11}$ vg/mouse of AAV9wt or PBS as a control. Direct EGFP fluorescence was detected (green) together with the DAPI signal (blue). Scale bar in the full section is 1 mm and 100 µm for the 10x magnification. Exposure was normalized to the liver of the AAV9wt group.

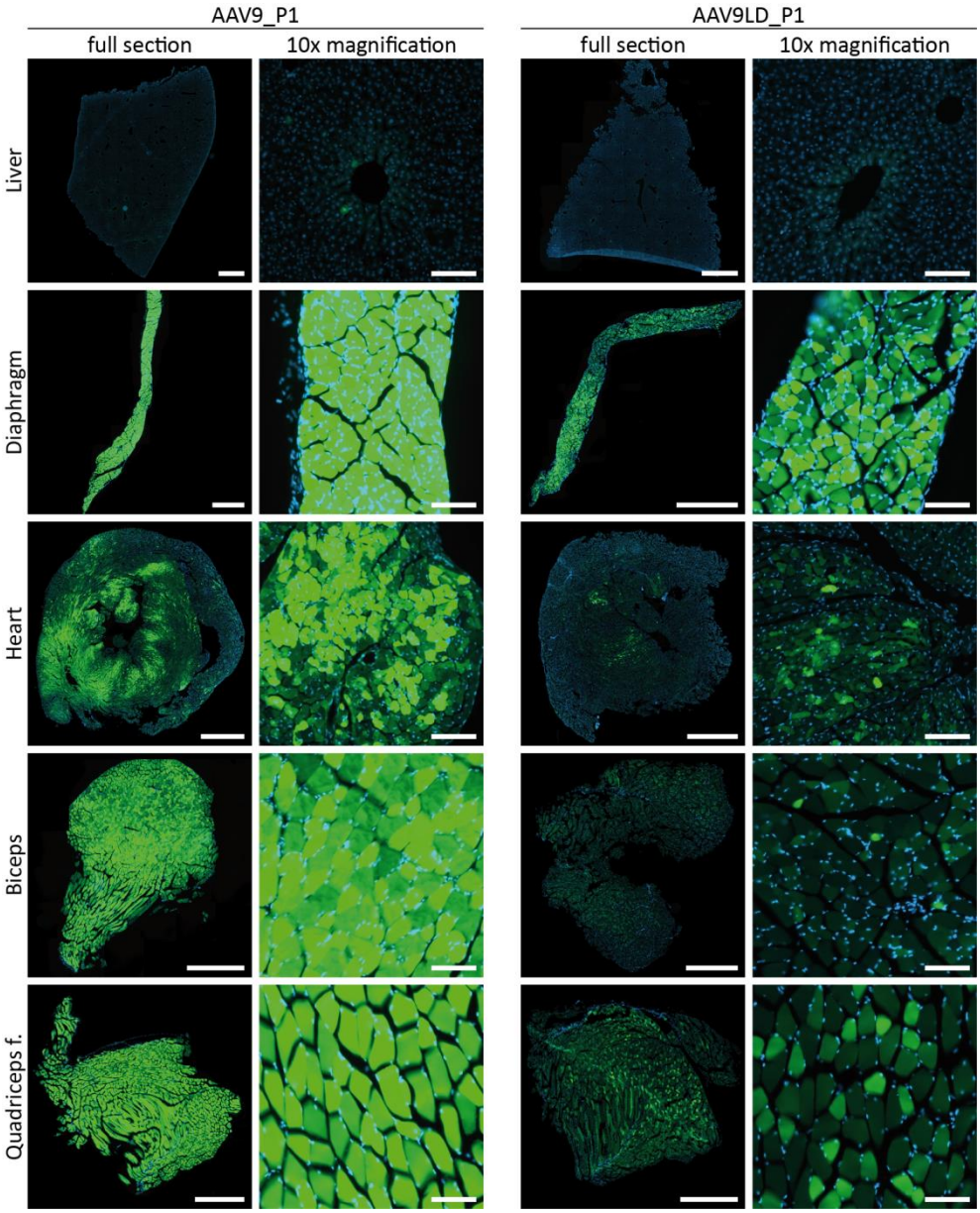

Figure 44: EGFP fluorescence of AAV9_P1 and AAV9LD_P1 group
Images show 10 µm cryosections of the liver, diaphragm, heart, biceps and quadriceps femoris. Representative sections were chosen from C57BL/6J mice injected with $5x10^{11}$ vg/mouse of AAV9_P1 or AAV9LD_P1. Direct EGFP fluorescence was detected (green) together with the DAPI signal (blue). Scale bar in the full section is 1 mm and 100 µm for the 10x magnification. Exposure was normalized to the liver of the AAV9wt group (Figure 43).

ACKNOWLEDGEMENTS

First of all, I would like to thank my supervisor Prof. Dr. Dirk Grimm for giving me the opportunity to become a part of his renowned research group and therefore experience the highly exciting field of gene therapy. I truly appreciate his trust to hire me for such an intense industry project that has been perfectly structured due to his thorough project planning beforehand. In my belief this is one of the most important prerequisites for a successful and efficient PhD, thus it cannot be stressed enough. I am also grateful for the profound scientific input throughout all the years that helped me dramatically to progress in the project and furthermore enhanced my knowledge regarding AAVs.

Secondly, I want to thank the other members of my thesis advisory committee, Prof. Dr. Ana Martin-Villalba and Prof. Dr. Oliver Müller for providing essential input for this work. Once more, I would like to explicitly thank Prof. Dr. Ana Martin-Villalba for being my primary examiner for this thesis and Prof. Dr. Oliver Müller for sharing his mice for my first barcode-based capsid screening. In addition, I appreciate having Prof. Dr. Martin Müller and Prof. Dr. Marc Freichel as my third and fourth examiner in my defense committee.

Many thanks go to my collaborators from Boehringer Ingelheim in particular Dr. Thorsten Lamla for coordinating this exciting joint project from the BI side and always supporting and motivating me during the course of my PhD. I also want to express gratitude to the whole Lamla lab, especially Kai Zuckschwerdt (ciao ragazzo) who helped me tremendously to get familiar with the new environment and was willing to answer my endless questions. Moreover, I want to thank Dr. Tanja Schönberger and Martina Steinrock for undertaking the long travel to Heidelberg in order to assist during the mice harvests. Many thanks as well to the sequencing lab, namely, Werner Rust, Dagmar Knebel and Eleonora Capitolo. It was always a pleasure for me to work with you and experience your positive approach to life. I also would like to say thank you to Dr. Birgit Stiersdorfer, Dr. Hannah Wyatt and Maria-Theresia Trinz for introducing me to the at first difficult but eventually beautiful world of histology. Furthermore, I want to acknowledge the bioinformaticians Dr. Holger Klein for verifying our analysis script and Dr. Germán Leparc for keeping the NGS pipeline running. Finally, my gratitude goes to Dr. Martin Lenter for his contributions in the three years and Silke

Heusel-Stütz for the perfect organization of my accommodation, even though I was pushing my luck most of the time by being super late.

On the campus in Heidelberg, I want to thank my collaborators Sascha Dehler for the complicated brain cell isolation and for the beer & talk sessions we had as well as Martin Busch for showing me how to isolate immune cells by MACS.

I would also like to mention the people behind the scenes, Martina Galvan from HBIGS for answering all my questions and the Bioquant staff, namely, Monika Beilharz, Yvonne Hess, Peter Schawerna and Daniel Browne for enabling a smooth daily routine.

A very big thanks goes to Sabrina Weis! Without her help I would still fiddle with gigantic excel sheets always worrying that there might me a tiny mistake somewhere in row 49311. Your script not only massively alleviated my work but will most likely be used regularly in our lab in the future. I promise, I will protect our bible and never forget the "Bäumchen". In this context, I also want to thank Josefine Sippel who adapted and designed the first analysis script, although she was working on her master's degree at that time.

One of the richest experiences I have made in this chapter of my life was the great atmosphere in AG Grimm. You guys not only helped me to survive the difficult phases of my PhD but in addition constantly supported me during a typical working day, for instance, by replying to my nonverbal demands. I would like to particularly mention Ellen Wiedtke and thank you for all the little things here and there which made my life considerably easier. Moreover, thanks to Janina Haar, Claire Domenger and Chiara Krämer for helping with the tissue isolation and Anne-Kathrin Herrmann as well as Jihad El Andari for working with me on the third library screening. I would also like to thank my practical student Jannik Traut for assisting me during the very labor-intensive and repetitive beginning of my PhD. Although you are not strictly a part of AG Grimm, I would still like to thank you, Manuela Nickl, for just being a nice personality and for the all the talks we had. Much appreciation to the infamous Team Ehrgeiz! Valle (Valerie Oberhardt), Tommy (Thomas Kehrer), Josel (Josefine Sippel) and Re-Pete (Peter Breunig), you were responsible for a truly great spirit in the S2 lab. I will keep our pictographic excursions in my mind. Thanks as well to my beloved A-Team. Claire (Domenger) and Adrian (Westhaus), it was a pleasure to share a bench with you and our passion for snowmen. However, I am still wondering, "Wo ist das A-Team?". Ben (Kachel), my voice is still scratchy from the high pitches in Lemon Tree. Thank

you buddy for waking up my at most tolerable singing qualities and also thank you for everything else.

Furthermore, I want to thank my DKFZ gang. “Wednesday is DKFZ friends day” was always a perfect opportunity to catch up with you guys and simultaneously briefly leave the own lab behind and focus on something else. Very importantly, a massive thank you to my old friends from Rottweil and Würzburg as well as to my siblings. I feel honored to have such amazing persons around me. Although we barely see each other, I have the feeling that nothing changed when we finally do and I can just be myself which was especially vital during this challenging period of my life.

From deep in my heart, I thank you, Britta, for your unconditional support, love and shared experiences. I will forever keep the memories in my mind.

Last but definitely not least, I want to express my highest gratitude to my parents for the continuous assistance and encouragement. I learned crucial life lessons in these three years but the most important things you learn in your childhood, and for that, you were the best teachers that I can imagine.