Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht - Karls - University

Heidelberg

Presented by

M. Sc. Jonas Becker

Born in Malsch, Germany

Oral examination: 23.11.2023

# Pulling the right viral levers: Engineering, screening and application of next-generation combinatorial AAV vectors

Referees: Prof. Dr. Marc Freichel

Prof. Dr. Dirk Grimm

# Abstract

Adeno-associated viruses (AAVs) present powerful vectors for human gene therapy and biomedical research. They enable the delivery of transgenes to a broad range of target tissues. This allows persistent expression of reporter genes, therapeutic gene replacement and the delivery of control elements for manipulating endogenous gene expression. Transduction efficiency and specificity for on-target over off-target cells are critical factors driving vector safety and applicability. In this study, I engineered AAV capsids, promoters and knockdown tools with the goal of generating efficient and specific vector components for the next generation of cell-type-specific gene therapy vectors.

In the first part of this doctoral work, I utilized Cas13d (CasRx) and short-hairpin (sh)RNA effectors to assess AAV-induced RNA degradation (knockdown). While CasRx showed promising knockdown of a *Renilla* luciferase reporter target, it failed to silence endogenous *CD44*, a potential driver of metabolic (non-alcoholic) steatohepatitis. shRNA effectors, however, allowed robust target knockdown of cellular RNAs (*CD44* and *ACE2*) as well as SARS-CoV-2 viral RNA. Direct targeting of SARS-CoV-2 genomic RNA triggered the evolution of escape mutations within the viral target sites. This mutational escape was efficiently suppressed by multiplexing of three shRNAs in a single AAV vector, thereby allowing a sustained suppression of SARS-CoV-2 infection in Vero E6 cells.

In the second part, I focused on improving screening conditions for promoters and AAV capsids. By assessing *eYFP* reporter expression *in vivo* for four promoter constructs individually, I could validate the findings of a previous promoter screen. This screen had utilized high-throughput barcode sequencing for parallel readout of a library of AAV-promoter constructs. I then applied this barcoding technique to dissect the activity of the GFAP promoter and truncated versions thereof. The GFAP promoter has previously mostly been used to induce astrocyte-specific transgene expression in the central nervous system. Strikingly, though, my results demonstrate a highly efficient GFAP promoter-driven transgene expression in human and murine hepatocytes.

To optimize the directed evolution of AAV capsids, I modified conventional capsid library screening by altering selection parameters. This was achieved by (i) introducing a Cas9-based negative selection for the removal of unwanted variants from the capsid library, and (ii) by exploring and applying RNA-based functional selection. I could generate an RNA-based screening platform by driving the expression of *cap* from the ubiquitous CMV

promoter instead of the endogenous p40. Both screening approaches proved applicable in cell culture settings. As RNA-driven selection offers a functional readout from both on- and off-target cells, I applied this approach for *in vivo* screening of an AAV6 peptide display library in mouse non-parenchymal liver cells. CMV promoter-driven *cap* expression enabled RNA-based readout of variant enrichment for on- and off-target cell-types. This demonstrated improved selectivity for RNA- over conventional DNA-based screening and facilitated the identification of functional, cell-type-specific capsid candidates.

In conclusion, my results show the development of efficient combinatorial shRNA-based knockdown vectors for inhibiting *CD44* expression or SARS-CoV-2 infection *in vitro*. Furthermore, I could implement improvements in capsid and promoter screening. This allowed the detection of highly functional variants with potential future applications in the development of novel gene therapy vectors.

# Zusammenfassung

Adeno-assoziierte Viren (AAVs) sind potente Vektoren für Anwendungen in humaner Gentherapie und biomedizinischer Forschung. Sie ermöglichen die Übertragung von Transgenen in eine Vielzahl verschiedener Zielgewebe. Dadurch lassen sich beispielsweise Reportergene gezielt exprimieren, mutierte Gene therapeutisch ersetzen, oder Kontrollelemente zur Manipulation der endogenen Genexpression einbringen. Kritische Faktoren für die Anwendbarkeit und Sicherheit viraler Vektoren sind Transduktionseffizienz und -spezifität. In dieser Studie optimierte ich AAV-Kapside, Promotoren und Expressions-Kontrollelemente, um effiziente und spezifische Komponenten für die nächste Generation von AAV-Vektoren zu entwickeln.

Im ersten Teil dieser Doktorarbeit verwendete ich Kontrollelemente basierend auf Cas13d- (CasRx) und shRNA-Effektoren, um diese Systeme für AAV-induzierte Degradierung von RNAs zu vergleichen. Obwohl ein *Renilla*-Luciferase-Reporter durch CasRx effizient herunterreguliert werden konnte, gelang dies nicht für die endogene RNA von *CD44*, einem potenziellen Regulator von nichtalkoholischer (metabolischer) Steatohepatitis. Dies war stattdessen möglich mit Hilfe von shRNA-Effektoren, wodurch sowohl zelluläre RNAs (*CD44* und *ACE2*) als auch die virale genomische RNA von SARS-CoV-2 degradiert werden konnten. Dies führte allerdings bei SARS-CoV-2 zur Anreicherung von Punktmutationen in den shRNA-Bindestellen. Diese Mutationen auf der viralen genomischen RNA verhinderten die Wirksamkeit der shRNAs. Derartige Durchbruchmutationen konnten durch die simultane Anwendung von drei shRNA-Effektoren verhindert werden. Dadurch wurde die persistente Inhibierung der SARS-CoV-2-Infektion in Vero E6-Zellen ermöglicht.

Für den zweiten Teil dieser Thesis arbeitete ich an der Verbesserung von Testkonditionen für Promotoren und AAV-Kapside. Durch Messung der Expression eines *eYFP*-Reportergens *in vivo* vermittelt durch vier verschiedene Promotoren konnte ich die Ergebnisse einer vorangehenden Promotor-Studie validieren. Diese Studie hatte Hochdurchsatzsequenzierung von barcodierten Reportergenen zur parallelen Ermittlung von Promotor-Effizienzen angewandt. Ich nutzte daraufhin dieselbe Methode, um die Aktivität des GFAP-Promotors und trunkierter Varianten davon zu untersuchen. Dieser Promotor wurde zuvor hauptsächlich zur gezielten Expression in Astrozyten des zentralen Nervensystems genutzt. Meine Ergebnisse zeigten allerdings, dass der GFAP-

Promotor eine effiziente Transgenexpression in humanen und murinen Hepatozyten ermöglicht.

Um die gerichtete Evolution von AAV-Kapsiden zu verbessern, modifizierte ich die konventionelle Anreicherungsmethode durch Veränderung der Selektionsparameter. Dies gelang einerseits durch die Implementierung einer Cas9-basierten negativen Selektion, welche ungewünschte Varianten aus Kapsidbibliotheken entfernen kann, und andererseits durch die Anwendung von RNA-basierter funktioneller Kapsidselektion. Ich konnte eine RNA-basierte Selektion etablieren, indem ich die Expression des *cap*-Gens durch den ubiquitären CMV-Promotor statt des AAV p40-Promotors regulierte. Beide Selektionsansätze zeigten prinzipielle Anwendbarkeit in Zellkultursystemen. Da RNA-basierte Selektion allerdings eine funktionelle Auswertung in Ziel- und Nicht-Ziel-Zellen erlaubt, wendete ich diesen Ansatz für eine *in vivo* Studie an. Dabei wurde eine AAV6-Peptidbibliothek in nicht-parenchymalen Mausleberzellen selektiert. Die CMV-Promotor-basierte *cap*-Expression ermöglichte RNA-basierte Auswertung in verschiedenen Leberzelltypen. Dadurch konnte ich eine verbesserte Selektivität der RNA-Selektion im Vergleich zu konventioneller DNA-basierter Selektion feststellen. Dies führte außerdem zur Anreicherung von funktionellen, Zelltyp-spezifischen Kapsidkandidaten.

Zusammenfassend zeigen meine Ergebnisse die Entwicklung effizienter kombinatorischer Vektoren zur shRNA-basierten Inhibierung von *CD44*-Expression und SARS-CoV-2-Infektion *in vitro*. Weiterhin konnten Verbesserungen für Promotor- und Kapsidselektion entwickelt werden. Dies ermöglichte die Detektion spezifischer Varianten für zukünftige Anwendungen in neuartigen Gentherapievektoren.

# Table of contents

# List of figures

# List of tables

# List of abbreviations

| Abbreviation | Full expression |
|---|---|
| **AAV** | Adeno-associated virus |
| **AAV-GPseq** | AAV-genome population sequencing |
| **AdH** | Adenoviral helper plasmid |
| **ANOVA** | Analysis of variance |
| **BC** | Barcode |
| **bgh-pA** | Bovine growth hormone polyadenylation signal |
| **bp** | Base pairs |
| **Cas** | CRISPR-associated protein |
| **CasRx** | *Ruminococcus flavefaciens* Cas13d |
| **cDNA** | Complementary DNA (*i.e.*, reversely transcribed RNA) |
| **Chol** | Cholangiocytes |
| **CNS** | Central nervous system |
| **COVID-19** | Coronavirus disease 2019 |
| **CRISPR** | Clustered regularly interspaced short palindromic repeats |
| **CRISPRi** | CRISPR interference |
| **crRNA** | CRISPR RNA |
| **dCas9** | Dead Cas9 |
| **ddPCR** | Droplet digital PCR |
| **DEPOOL** | Depletion of off-targeting AAVs from on-target libraries |
| **DMEM** | Dulbecco's modified eagle medium |
| **DR** | Direct repeat |
| **dsRNA** | Double-stranded RNA |
| **EC** | Endothelial cells |
| **EMA** | European Medicines Agency |
| **FACS** | Fluorescence-activated cell sorting |
| **FBS** | Fetal bovine serum |
| **FDA** | U.S. Food and Drug Administration |
| ***Fluc*** | Firefly luciferase |
| **gDNA** | Genomic DNA |
| **gRNA** | Guide RNA |
| **HCC** | Hepatocellular carcinoma |
| **HCV** | Hepatitis C virus |
| **Hep** | Hepatocytes |
| **HIV** | Human immunodeficiency virus |
| **HSC** | Hepatic stellate cell |
| **HSPG** | Heparan-sulfate proteoglycan |
| **IDT** | Integrated DNA Technologies |
| **IF** | Immunofluorescence |
| **IVIG** | Intravenous immunoglobulins |
| **ITR** | Inverted terminal repeat |
| **kb** | Kilobases |
| **LSEC** | Liver sinusoidal endothelial cells |
| **Mac** | Macrophages |
| **MAFLD** | Metabolic dysfunction-associated fatty liver disease (see NAFLD) |
| **MASH** | Metabolic dysfunction-associated steatohepatitis (see NASH) |
| **min** | Minutes |
| **mini-pA** | Minimal polyadenylation signal |
| **miRNA** | micro-RNA |
| **MOI** | Multiplicity of infection |
| **MPRA** | Massively parallel reporter assay |
| **mRNA** | Messenger RNA |
| **NAb** | Neutralizing antibody |

| Abbreviation | Full expression |
|---|---|
| **NAFLD** | Non-alcoholic fatty liver disease (see MAFLD) |
| **NASH** | Non-alcoholic steatohepatitis (see MASH) |
| **NEB** | New England Biolabs |
| **NGS** | Next generation sequencing (high-throughput sequencing) |
| **NLS** | Nuclear localization signal |
| **NPC** | Non-parenchymal (liver) cell |
| **NSP** | Non-structural protein |
| **nt** | Nucleotide |
| **NTC** | No-template control |
| **ORF** | Open reading frame |
| **pA** | Polyadenylation signal |
| **PAGE** | Polyacrylamide gel electrophoresis |
| **PBS** | Phosphate-buffered saline |
| **PEI** | Polyethyleneimine |
| **PHH** | Primary human hepatocytes |
| **pre-miRNA** | Precursor-miRNA |
| **pri-miRNA** | Primary miRNA |
| **qPCR** | Quantitative PCR |
| **rAAV** | Recombinant AAV |
| **RdRP** | RNA-dependent RNA polymerase |
| **RISC** | RNA-induced silencing complex |
| **RLC** | RISC-loading complex |
| **RLU** | Relative luminescence units |
| ***Rluc*** | *Renilla* luciferase |
| **RNAi** | RNA interference |
| **rpm** | Revolutions per minute |
| **rt** | Room temperature |
| **RT** | Reverse transcription |
| **RTqPCR** | Reverse transcription quantitative PCR |
| **SARS-CoV-2** | Severe acute respiratory syndrome coronavirus 2 |
| **scAAV** | Self-complementary AAV |
| **SD** | Standard deviation |
| **sec** | Seconds |
| **sgRNA** | Single guide RNA |
| **shRNA** | Short hairpin RNA |
| **siRNA** | Short interfering RNA |
| **SMRT-seq** | Single-molecule, real-time sequencing |
| **ssAAV** | Single-stranded AAV |
| **ssDNA** | Single-stranded DNA |
| **TF** | Transcription factor |
| **TFBS** | Transcription factor binding site |
| **tracrRNA** | *trans*-activating CRISPR RNA |
| **TRS** | Terminal resolution site |
| **UTR** | Untranslated region |
| **vg** | Vector genomes |
| **vg/dg** | Vector genomes per diploid genome |
| **VR** | Variable region |
| **WPRE** | Woodchuck hepatitis virus post-transcriptional regulatory element |

# 1 Introduction

Adeno-associated viruses (AAVs) (section 1.1) have been modified into powerful vectors for gene therapy (section 1.1.1-1.1.3). The creation of efficient and specific AAV vectors heavily relies on capsid engineering (section 1.1.4) and the assembly of optimal transgene expression cassettes (section 1.1.5). Apart from therapeutic gene replacement, AAVs also enable the delivery of genetic components that allow gene editing, silencing, or, specifically, the targeted degradation of cellular or pathogenic RNAs (section 1.2). Such knockdown tools are valuable instruments for studying or countering various diseases, including viral infections (section 1.2.3.1) and metabolic diseases (section 1.2.3.2). This doctoral thesis focuses on the improvement of AAV vector tools on both capsid and transgene level. Furthermore, AAV-induced RNA knockdown tools are evaluated and applied for downregulation of different targets.

## 1.1 Adeno-associated virus (AAV)

First described in 1965, AAVs were initially identified as contaminating agents in simian adenovirus cultures [1,2]. Their inability to replicate in the absence of adenovirus led to their description as “defective” virus particles. Although not defective, the replication of AAVs is indeed dependent on co-infection with helper viruses. These include adenoviruses, herpesviruses or papillomaviruses [3-5]. Induction of genotoxic stresses was also shown to drive AAV replication [6,7]. This dependency led to the classification of AAVs into the genus *dependoparvoviruses* within the *parvoviridae* family. As such, they present as non-enveloped capsids containing a 4.7 kb single-stranded (ss)DNA genome, equally packaging sense and antisense strands (Figure 1A) [8]. The AAV genome is flanked by inverted terminal repeats (ITRs) and contains the two major genes *rep* (non-structural) and *cap* (structural), followed by a polyadenylation signal (pA) [9]. The *rep* gene encodes the non-structural proteins Rep78, Rep68, Rep52 and Rep40, which are expressed from the p5 promoter (large Rep proteins, *i.e*., Rep78 and Rep68) and p19 promoter (small Rep proteins) [10,11]. Splicing the intron in the 3’ region of *rep* yields Rep68 and Rep40, while the other two isoforms are not spliced. The unique N-terminal region of the large Rep proteins contains a DNA-binding domain with site-specific recognition, while the region common to all four Reps contains helicase and ATPase domains [12,13].

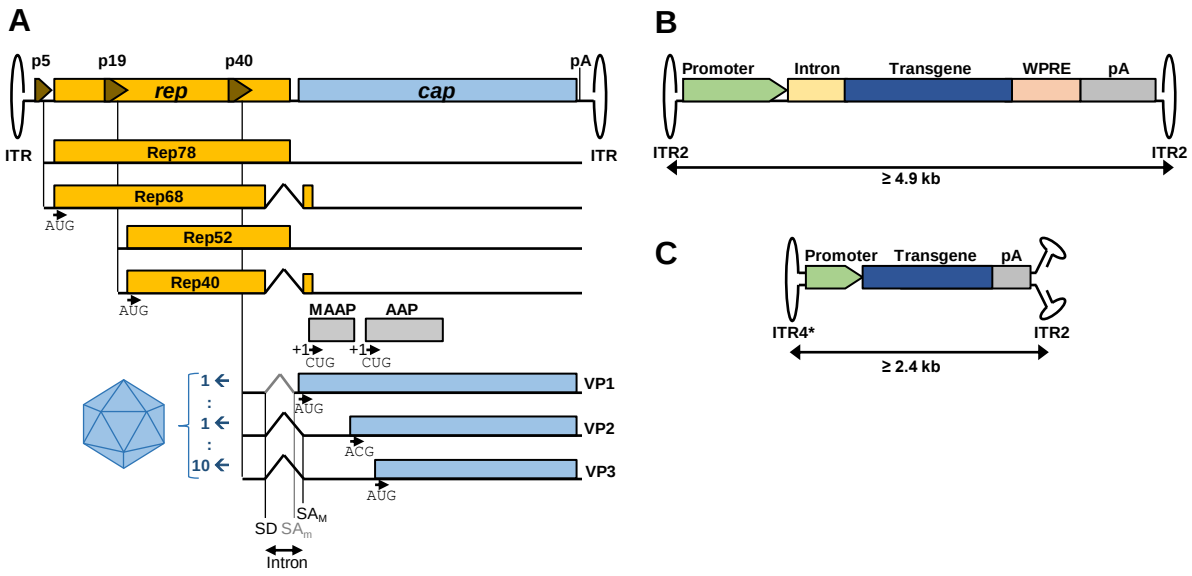



**Figure 1: Genome structure of wild-type AAV and recombinant (r)AAV. (A)** The 4.7 kb genome of wild-type AAV contains *rep* and *cap* genes flanked by inverted terminal repeats (ITRs). Expression is driven by p5, p19 and p40 promoters and terminated by a single polyadenylation signal (pA). Transcription from p5 yields the large Rep proteins Rep78 (non-spliced) and, through splicing, Rep68. Small Rep proteins Rep52 (non-spliced) and Rep40 (spliced) are expressed from the p19 promoter. The p40 promoter drives expression of the capsid proteins VP1-3. These are generated through differential splicing from a single splice donor (SD) to a minor and a major splice acceptor site ($SA_m$ and $SA_M$, respectively) and differential start codon usage including a non-canonical ACG for VP2. Accordingly, VP1-3 are expressed in a 1:1:10 ratio and assemble as such into the final capsid. The accessory proteins MAAP (membrane-associated accessory protein) and AAP (assembly-activating protein) are encoded from a +1 frameshift within *cap* and utilize CUG start codons. **(B-C)** For recombinant (r)AAV vectors, the viral genes are replaced by transgene expression cassettes. This is possible since the ITRs are the only *cis* genomic element required for packaging into AAV capsids. For single-stranded (ss)AAV vectors (**B**), packaging of up to 4.9 kb is possible [14]. Mutation (*) of one of the ITRs (often using ITR4 to prevent recombination) within the terminal resolution site leads to the arrest of vector DNA replication after one round of duplication, resulting in packaging of self-complementary (sc)AAV genomes containing two inverted copies of an insert with up to 2.4 kb packaging capacity (**C**). These scAAVs do not rely on second-strand synthesis upon transduction and thus allow for a more rapid onset of transgene expression.

Expression of *cap* is driven by the p40 promoter (lying within *rep*), which yields two messenger (m)RNAs by differential splicing [15,16]. The intron is defined by a single splice donor (SD) site and two splice acceptors, the minor ($SA_m$) and the major splice acceptor ($SA_M$). Splicing to $SA_m$ allows usage of the conventional AUG start codon, thus resulting in the large VP1 protein. Splicing to $SA_M$, on the other hand, skips this start codon, thus only enabling translation for VP2 (ACG start codon) and VP3 (AUG start codon). The differential splicing and start-codon usage lead to an approximate ratio of 1:1:10 for VP1:VP2:VP3, which is reflected in the assembled capsid [17,18]. Within the *cap* gene, two

additional open reading frames encode the membrane-associated accessory protein (MAAP) [19] and the assembly-activating protein (AAP) [20] in a +1 frameshift. MAAP is membrane-bound and likely assists in viral egress and exosome formation [19,21,22], while AAP is essential for capsid assembly of most serotypes [20,23]. The assembled AAV capsid contains 60 subunits of VP1/VP2/VP3 forming a T=1 icosahedral structure with a diameter of approximately 26 nm [24,25]. Currently, 13 primate AAV serotypes are differentiated [26,27], which exist next to a plethora of other naturally occurring variants. Based on capsid similarity, these can be structured into six clades A-F [28]. The human seroprevalence for AAV-binding antibodies is estimated to be between 30-60% and varies between population and serotype [29,30], with high cross-serotype reactivity being observed [31]. AAVs are commonly regarded as non-pathogenic [32] and even display beneficial properties such as the inhibition of infections with its helper viruses [5]. However, the recent association of AAV2 infection with severe acute non-A-E hepatitis suggests conditional pathogenesis in previously non-exposed children, which is still poorly understood [33-35]. Apart from humans and non-human primates, AAVs have been discovered in multiple other mammalian species, with close relatives also being found in birds and snakes [36-38].

### 1.1.1 AAV infection pathway

AAV tropism varies between serotypes and includes a wide range of infectable dividing and non-dividing cell types. These tropisms are defined by the capsid structures and their interactions with different attachment factors and co-receptors expressed by host cells [39,40]. AAV2, the best studied and most prevalent serotype in the human population, uses heparan-sulfate proteoglycan (HSPG) as its primary attachment factor [41-43]. Known co-receptors include fibroblast growth factor receptor 1 and integrins [44-46]. Other AAVs bind to O- and N-linked sialic acid or N-linked galactose moieties for attachment [47], and show co-receptor usage of, *e.g.*, platelet-derived growth factor receptor (PDGFR; AAV5) [48] or epidermal growth factor receptor (EGFR; AAV6) [49]. With a few exceptions such as AAV4 and AAVrh32.33 [50], most AAVs rely on binding to the common AAV receptor AAVR for establishing infection [51]. After attachment to their primary receptor/co-receptor and AAVR, AAVs are internalized by clathrin-mediated [52] or clathrin-independent endocytosis [53], or macropinocytosis [54] (Figure 2). Following uptake, the acidification in early to late endosomes triggers a conformational change in the AAV capsid that releases the previously hidden VP1/VP2 regions [55]. AAVs may then travel to the *trans*-Golgi network

[56]. Mediated by the phospholipase A2 domain in the VP1-unique region [57], capsids escape from vesicles (endo-/lysosomes or *trans*-Golgi network) into the cytosol. In the cytosol, some viruses are lost due to ubiquitination and proteasomal degradation [58]. The remaining AAVs travel through the cytosol *via* interactions with microtubules [59]. They then enter the nucleus through a nuclear pore complex [60] using a nuclear localization signal (NLS) present within the VP1/VP2 common region [61].

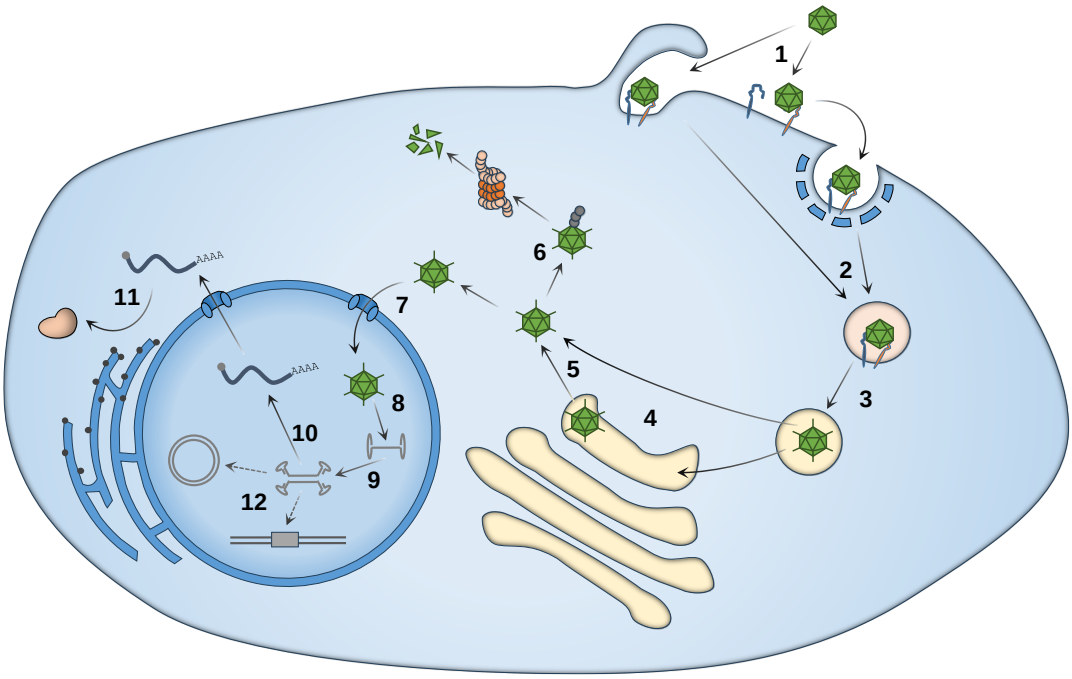



**Figure 2: AAV infection and transduction pathway.** Depending on capsid variant, AAVs attach to different cell surface receptors and co-receptors such as AAVR (1), after which they enter cells through endocytosis or macropinocytosis (2). The change in pH from early to late endosomes triggers a conformational change that releases the hidden VP1/VP2-unique region from within the AAV capsid (3). Next, AAVs can travel to the *trans*-Golgi network (4) and escape into the cytosol using the phospholipase A2 domain of the VP1-unique region. Within the cytosol, ubiquitination can lead to proteasomal degradation of AAV capsids (6). If this is not the case, the nuclear localization signal in the VP1/VP2-unique region allows trafficking through the nuclear pore complex into the nucleus (7), where AAV capsids release their genome (8). After second-strand DNA synthesis (9), both wild-type and recombinant AAVs can now express encoded genes, although expression is strongly reduced for wild-type AAVs in the absence of a helper virus co-infection (10). Transcribed mRNA is exported into the cytosol for translation into proteins (11). Recombinant AAV vectors mostly persist as mono- or concatemeric episomes, whereas the large Rep proteins of wild-type AAVs facilitate integration into the AAVS1 locus on chromosome 19 (12).

In the nucleus, capsids travel through the nucleolus and finally release their genome [62]. The single-stranded AAV genomes undergo second-strand synthesis by host polymerases, which is primed by the 5' ITR [63]. In the absence of helper viruses, AAVs establish latent infections. Initially expressed Rep78 and Rep68 proteins cooperate with

host factors to induce silencing of the p5 and p19 promoters [64]. During latency, the AAV genome can persist either episomally as monomers or concatemers [65], or is integrated into the AAVS1 locus on chromosome 19 through the large Rep proteins [66,67]. The autorepression of the p5 promoter through large Rep proteins prevents excision of the integrated virus and expression of its proteins.

Co-infection with a helper virus induces the lytic phase of AAV infection by *trans*-activating its promoters [5,68]. This promoter regulation is multifaceted: as mentioned, p5 and p19 are inactivated in the absence of helper virus through binding of Rep78 and Rep68 [64,69]. Several host factors such as the YY1 transcriptional regulator are involved in repression as well [40]. In the presence of adenovirus coinfection, the adenoviral E1A protein relieves YY1-induced repression [70], thus allowing expression of large Rep proteins. These in turn reduce p5 activity and *trans*-activate both p19 (together with the host SP1 transcription factor) and p40 promoters [71]. Inactivation of p5 by Rep78 and Rep68 is lifted through Rep52 and Rep40, thereby creating a tight control of *rep* expression [68]. During helper virus co-infection, the p40 promoter generates more transcripts than p5 and p19 [72], but is barely active in the absence of a helper virus [73,74]. The induced expression of Rep proteins is necessary to facilitate excision of integrated AAV genomes. Large Rep proteins can bind to the Rep-binding site located within the ITR and induce nicking at the terminal resolution site (TRS) [75]. This nicking allows the host polymerase machinery to replicate the AAV genome including the ITRs using the generated 3' hydroxyl group at the TRS as primer [5,76]. The *cap* proteins VP1-3 (expressed from p40) assemble in the nucleoplasm or nucleolus with the assistance of AAP (depending on serotype) [20,23,77]. Next, the small Rep proteins (Rep52, Rep40) facilitate packaging of newly formed genomes into the assembled capsids through their helicase activity [78]. Rep52/40 pump the single-stranded genome through a pore in the 5-fold symmetry axis of the AAV capsid [79]. Deletions in the AAV genome demonstrated that only the ITR sequences are essentially required for genome packaging [80]. Therefore, packaging is still possible when the entire coding sequence of AAV is replaced by a sequence of choice. This facilitates the applicability of AAVs as viral vectors.

### 1.1.2 Recombinant AAVs as viral vectors

The cloning of the AAV2 genome into a bacterial plasmid was first reported in 1982 [81]. This enabled production of infectious AAV2 virus by plasmid transfection into cells that were superinfected with adenovirus, and led to a better understanding of the virus'

biology. Soon after, packaging of AAV genomes with major deletions in coding sequences was achieved [80]. Without the presence of functional *rep* and *cap* genes, the AAV replication and infection phases are decoupled [40]. The first usage of this, and thus the first application of a recombinant (r)AAV as a viral vector, was achieved by replacing the *cap* gene with a neomycin resistance cassette and using the resulting vector for transduction of cultured mammalian cells [82]. Several critical improvements helped developing rAAV into an efficient vector system [83], including complete gutting of *rep* and *cap* genes and improved production systems [84,85]. Production initially involved co-transfection of AAV vector and *rep/cap* plasmids into cells, followed by infection with adenovirus. Adenovirus had to be removed from purified lysates by heat-inactivation to generate helper-free rAAV stocks [85]. As the necessary helper functions were narrowed down, it was discovered that the only adenovirus helper genes necessary for rAAV production were E1A, E1B, E2A, E4 and VA RNA [5]. Since Hek293T cells already express E1A and E1B, rAAV production is feasible for instance by triple-transfection of (i) an AAV-helper plasmid supplying the *rep* and *cap* genes without ITRs, (ii) an adenovirus helper (AdH) plasmid supplying E2A, E4 and VA, and (iii) the vector plasmid containing a transgene cassette flanked by ITRs [86,87]. Other systems use a combined plasmid containing both AdH and AAV-helper genes [88], producer cell lines with permanently integrated helper genes [89], or even production in Sf9 insect cells through baculovirus-based delivery of vector genome and helper genes [90]. Each production system has specific advantages [91], but common problems include (i) difficulty to achieve sufficient yields for clinical application (highly depending on the employed capsid), (ii) presence of large fractions of empty capsids in the final product (depending on the purification process) [92], and (iii) mis-packaging of helper or cellular genes, or truncated vector products (especially in cases of strong secondary transgene structures) [93].

While wild-type AAVs undergo site-directed integration into AAVS1, rAAVs lack Rep proteins and therefore mostly persist episomally [94]. However, rare events of undirected integration into the host genome have been observed [95]. Although these random integrations are a major concern for the safety of AAV gene therapies, they can also contribute to stable and persistent therapeutic transgene expression [96].

The design of transgene cassettes for AAV vectors has one inevitable restriction: vector packaging is strictly limited to genome sizes of up to 4.9 kb (Figure 1B) [14]. This limits the number of transgene expression cassettes, as well as transgene size and utilization of modulator sequences (discussed further in section 1.1.5). Once packaged, the AAV

vector can infect (transduce) a host cell equally to its wild-type counterpart (Figure 2). A major rate-limiting step of this transduction is the second-strand synthesis of the vector genome, which is necessary to enable transgene expression [97]. This limitation can be circumvented by employing modified vector constructs that contain a deletion of the TRS of one of the two ITRs [98,99]. Preventing terminal resolution promotes the formation of a double-stranded back-fold intermediate during genome replication that can be efficiently packaged instead of the single-stranded (ss)AAV genome. The resulting self-complementary (sc)AAV vector (Figure 1C) produces a more rapid onset and overall higher transgene expression, as it does not require second-strand synthesis prior to initiation of transgene expression. The drawback of scAAVs is a reduction in genome size to 50% (ca. 2.4 kb) as compared to ssAAVs, although reports of using helper plasmids with lower Rep expression have shown successful packaging of up to 3.3 kb [100]. In order to prevent recombination of the TRS-mutated ITR with the wild-type AAV2 ITR, the sequence of the AAV4 ITR (ITR4) with TRS-deletion can be used instead.

Of all serotypes, AAV2 is the most researched and best studied variant. ITR and *rep* sequences of AAV2 are most widely applied for vector production. However, early observations demonstrated that AAV pseudotyping is possible, as genomes flanked with AAV2 ITR can be packaged into the capsids of other AAV serotypes by inserting their *cap* gene instead of *cap2* in the AAV-helper plasmid [101-103]. As mentioned above, different AAV capsids bind to different host receptors and thus exhibit different *in vivo* tropisms. Therefore, pseudotyping allows the creation of gene therapy vectors with transduction properties matching the required indication. As the naturally occurring AAV capsids have not evolved to meet specialized tropism requirements (*e.g.*, transduction limited to the central nervous system for gene therapy of neurodegenerative diseases), various capsid modification approaches were pursued to create capsid variants with desired transduction properties (see section 1.1.4).

### 1.1.3 AAV gene therapy

After the first use of AAVs as vectors for gene transfer [82], its subsequent application *in vivo* quickly demonstrated successful and long-lasting overexpression of a therapeutic transgene, the cystic fibrosis transmembrane conductance regulator (CFTR) [104]. This soon entailed the first clinical trial with an AAV2 vector for replacing CFTR in patients with cystic fibrosis, a disease in which inherited mutations in the *CFTR* gene cause a lethal multi-symptom disease [105]. Unfortunately, this vector only achieved low levels of gene

transfer and failed to induce therapeutic efficacy [106], demonstrating needs for optimized delivery. The number of clinical trials with AAV vectors has since increased massively, with over 350 reported so far [107]. Therefore, AAV vectors are amongst the most commonly applied gene therapy transfer agents, only surpassed by adenovirus, retrovirus, naked plasmid DNA and lentivirus [107]. So far, seven AAV gene therapy products have achieved approval by U.S. Food and Drug Administration (FDA), European Medicines Agency (EMA) or both (Table 1).

**Table 1: Approved AAV gene therapy products.** Approval by U.S. Food and Drug Administration (FDA) and/or European Medicines Agency (EMA).

| First approval | Trade name | Full name | AAV capsid | Transgene | Route of administr. | Disease | Ref. |
|---|---|---|---|---|---|---|---|
| 2012* | Glybera | Alipogene tiparvovec | AAV1 | *LPL$^{S477X}$* | intra-muscular | lipoprotein lipase deficiency | [108] |
| 2017 | Luxturna | Voretigene neparvovec | AAV2 | *RPE65* | subretinal | RPE65-mediated inherited retinal dystrophy | [109] |
| 2019 | Zolgensma | Onasemnogene abeparvovec | AAV9 | *SMN1* | intravenous | Spinal muscular atrophy | [110] |
| 2022 | Upstaza | Eladocagene exuparvovec | AAV2 | *AADC* | intra-putaminal | aromatic L-amino acid decarboxylase deficiency | [111] |
| 2022 | Roctavian | Valoctocogene roxaparvovec | AAV5 | Factor VIII | intravenous | Hemophilia A | [112] |
| 2022 | Hemgenix | Etranacogene dezaparvovec | AAV5 | Factor IX | intravenous | Hemophilia B | [113] |
| 2023 | Elevidys | Delandistrogene moxeparvovec | AAV rh74 | Micro-dystrophin | intravenous | Duchenne muscular dystrophy | [114] |

*Market authorization not renewed.

Authorized in 2012 by the EMA, Glybera (alipogene tiparvovec) achieved the first ever market authorization of an AAV gene therapy product in the Western world [115]. It utilized an AAV1 capsid for intramuscular delivery of the *LPL$^{S477X}$* transgene, a gain-of-function variant of lipoprotein lipase, to treat lipoprotein lipase deficiency [108,116]. Although Glybera showed favorable safety and efficacy outcomes, its market authorization was not renewed in 2017 due to economic concerns [117]. Such concerns exist for most AAV gene therapy products [83,118], with the highest cost for an individual treatment being set by Hemgenix with 3.5 million U.S. dollars per dose. However, the benefits these gene therapies can offer is immense, including retention of eyesight in patients with RPE65-

mediated inherited retinal dystrophy (Luxturna) [109], or survival and gain of basic muscle functions for infants with spinal muscular atrophy (Zolgensma) [110].

Clinical studies over the last decades demonstrated several difficulties for AAV gene therapy. These include high levels of neutralizing antibodies (NAbs) in the human population [29,30]. This is most often solved by either immunosuppression *via* glucocorticoids or exclusion of patients with high NAb titers [119]. The potential risk of capsid- or transgene-directed immune response after administration requires close monitoring of patients after treatment [119,120]. As transduction efficiency of employed capsids is often low, high vector titers are required to achieve therapeutic efficacy. The highest doses applied so far were $3\times10^{14}$ vector genomes (vg) per kilogram bodyweight for intravenous administration in a clinical trial using AAV8-based delivery of the therapeutic *MTM1* gene for treatment of X-linked myotubular myopathy [121]. Unfortunately, of the 17 patients that received this dosage, three succumbed to hepatotoxicity-induced fatality (also 1/7 patients treated with a lower dose of $1\times10^{14}$ vg/kg). So far, this marks the highest number of deaths for any AAV gene therapy trial. Apart from hepatotoxicity, dorsal root ganglion toxicity, T-cell derived toxicity and complement activation have all been observed in clinical trials [120]. In order to overcome these limitations and to not depend on potentially toxic high-dose administrations, novel AAV vectors are required with lower immunogenicity, better on- over off-target specificity and higher transduction efficiency [122].

### 1.1.4 Capsid engineering

Most clinical applications of AAV vectors employ systemic (intravenous) administration of capsids from naturally occurring variants [119]. Target tissues for intravenous administration often include liver (*e.g.*, for overexpression of Factor VIII/IX in hemophilia A/B), muscle (*e.g.*, for treatment of muscular dystrophies) or central nervous system (CNS; *e.g.*, for treatment of neurodegenerative diseases). The indiscriminate broad transduction profile of wild-type AAV variants commonly does not match the requirements for such specialized targeted tropism. Although still underrepresented in clinical trials [123], engineered capsids can display enhanced or otherwise modified tropisms as compared to naturally occurring AAV variants [122,124,125]. This enables the creation of rAAV vectors with enhanced transduction of on-target tissues, de-targeting from off-target tissues, or evasion of neutralizing antibodies. Modifications can be implemented through either post-translational capsid changes, such as chemical fusion to receptor-specific targeting

moieties [126], or through changes in the amino-acid composition of the capsid proteins VP1-3 encoded through modifications in the *cap* gene (Figure 3A).

As mentioned before, pseudotyping allows the utilization of different capsid variants for packaging of ITR2-flanked vector genomes by employing a *cap* gene of choice in the AAV-helper plasmid during vector production [101-103]. A large set of naturally occurring capsid variants have already been discovered and characterized, showing a range of distinct transduction profiles [127]. In addition to this, different capsid mutagenesis approaches were employed to the *cap* gene early on with the goal of understanding and modifying capsid features. Modification strategies can be separated into rational mutagenesis and random diversification. Rational modifications introduce a limited set of defined alterations to the *cap* gene, *e.g*., changes of individual amino acid residues [19,41], insertion of receptor-binding peptides [128] or even fusion to receptor-targeting nanobodies [129] or DARPins [130]. Random diversification, on the other hand, is not limited to prior knowledge and instead utilizes randomization of *cap* by error-prone PCR [131], insertion of random peptide sequences [132], or DNA family shuffling [133]. These processes generate diverse libraries of *cap* variants (1.1.4.1, Figure 3A). During virus production, each *cap* variant is then ideally packaged into its respective capsid. The resulting capsid libraries are then screened iteratively in an appropriate animal or tissue culture model to enrich variants with the desired transduction properties (1.1.4.2, Figure 3B-C).

#### *1.1.4.1 Capsid diversification*

Comparisons between capsid sequences from different AAV serotypes revealed nine regions with low conservation (Figure 3A) [134]. These variable regions (VR) I-IX tolerate modifications especially well [19], and can enable efficient capsid re-targeting due to their partial surface localization and accessibility to host receptors. Mutations in VR XIII helped to identify the HSPG-binding motif of AAV2 [41]. Insertional mutagenesis of VR XIII [132], and to a lesser extent VR IV [128], is commonly used for capsid engineering due to their tolerance for mutation and re-targeting capability.

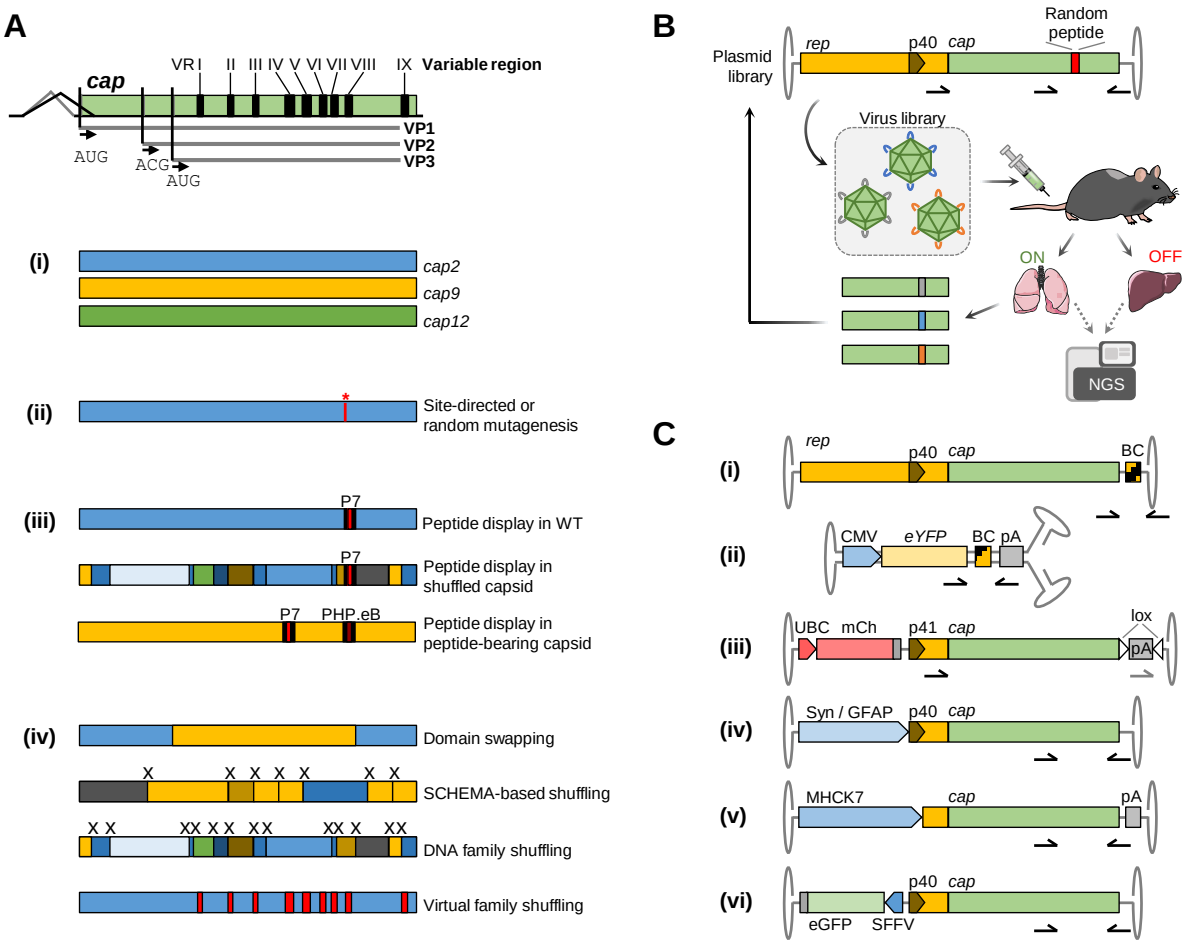



**Figure 3: Capsid engineering – *cap* diversification and selection designs. (A)** The AAV *cap* gene encodes the capsid proteins VP1-3 and contains nine variable regions (VR I-IX) with low conservation across serotypes. Capsids from naturally occurring AAVs can be utilized directly (i) or modified to alter capsid features such as tropism. Common capsid modifications include site-directed or random mutagenesis (ii) [41,131], peptide insertion or random peptide display (iii) [132,133,135], and recombination of *cap* regions from different serotypes [133,136-138]. **(B)** Common *cap* selection scheme [139]. To enrich capsid variants with desired tropisms from diverse *cap* libraries (here: random peptide display), virus pools are produced where each *cap* variant is packaged by its self-encoded capsid. After screening of virus libraries in the absence of a helper virus in an animal or cell culture model, enriched *cap* variants can be rescued by PCR from DNA of on-target tissues (*e.g.*, lung) and cloned back into the *rep-cap* context for iterative selection. For random peptide libraries, next-generation sequencing (NGS) allows tracking of variant enrichment. This setting only allows DNA-based tracking and does not interrogate functional transduction. **(C)** Alternative screening settings. To allow NGS-based readout of variants with more broadly distributed modifications than peptide display, barcoding of *rep-cap* libraries has been employed (i) [140,141]. Barcoding can also be performed for reporter transgenes, where each barcode is coupled with one specific capsid variant (ii) [142-144]. This setting enables readout for both DNA and RNA, thus allowing tracking of functional transduction. The “CREATE” and similar systems track functional transduction by using the Cre-lox system, where rescue PCR amplifies only *cap* genes that have undergone Cre-based recombination in the nucleus (iii) [145]. The “TRACER” (iv) [146], “DELIVER” [147] (v) and “FT” platforms (vi) [74] use rescue from RNA to enrich functional capsids. This is facilitated by increasing *cap* expression through tissue-specific promoters (iv and v) or by boosting p40 activity through a reverse SFFV promoter (vi). A modified version of panel A has been published in Pathogens [124].

Capsid re-targeting by insertion of peptide sequences has been performed with pre-selected or pre-defined peptides, such as hits from phage display [148] or insulin-mimetic peptides for targeting of the insulin receptor [128]. More commonly, however, randomized peptide sequences are inserted to generate diverse capsid libraries in a process termed AAV peptide display [132,143,146,147,149]. This enables targeting of undefined (and mostly unknown) receptors by directed evolution with a suitable iterative selection model. For instance, screening of a randomized 7mer peptide library inserted into the AAV2 capsid allowed the identification of a highly specific variant (AAV2_L1) with potent transduction of lung endothelial cells [149]. Insertion has been performed for AAV2 initially [132], but is feasible for any natural [148] or synthetic [133,135] capsid variant as long as it is performed in a favorable insertion site.

Apart from peptide display, modification of VR sequences has also been employed for virtual family shuffling, a process where the residues across different VRs are diversified [138]. Other options of broader diversifications addressing a larger region of the AAV capsids integrate *cap* sequence information from multiple parental serotypes. This integration is possible in two different manners, namely, either *in silico* by using sequence information of different capsid variants to calculate and generate a pool of their most likely common ancestor sequences [150,151], or by creating chimeric *cap* genes of a set of closely related parental sequences through random DNA family shuffling [133] or site-directed SCHEMA-based shuffling [137]. As opposed to peptide display, diversification of multiple VRs [152] and DNA family shuffling [153] also allows the creation of immune-evading capsid properties (*i.e.*, NAb evasion).

#### *1.1.4.2 Capsid library screening*

Once diversified, capsids with favorable features (*e.g.*, transduction profile) need to be enriched and identified. Conventionally, modified *cap* gene libraries are cloned into an AAV2-like *rep-cap* genome (containing ITR2 and *rep2*; Figure 3B) [139]. Highly diverse capsid libraries are produced and then screened iteratively in cell culture or animal models. After transduction, variants enriched in target cells are rescued through PCR-based amplification from genomic DNA, cloned back into the *rep-cap* genome, and screened again. For peptide display screens, variant enrichment can be monitored through Next-Generation Sequencing (NGS) of peptide insertion sites, comparing variant enrichment over iterative rounds in target tissues [149]. Other options are Sanger or single-

molecule, real-time (SMRT) sequencing, which offer readout of longer sequences but lower read depths.

This conventional screening approach has several pitfalls, including variant translatability (a), immunogenicity (b), off-targeting (c), and lack in functionality (d) [122,154]. (a) Translatability of capsid performance is a general problem, as screening in cell culture models often does not predict the *in vivo* behavior of capsids, while screening in mouse models often does not translate to non-human primates or humans. This issue can be overcome by screening in humanized mice [155] or by screening iteratively in several animal models to enrich cross-active candidates [156]. (b) Another problem with conventional selection is that enriched candidates are usually still sensitive to neutralization by NAbs present within the human population. Evasion of neutralizing antibodies can be implemented, however, by including negative selection pressure through incubation with intravenous immunoglobulins (IVIGs) [153]. (c) Furthermore, although variants are selected in a target tissue, candidates may evolve to bind non-specific but highly prevalent receptors. These capsids would present a general, non-selective tropism, and therefore be useless for targeted applications. Such candidates can be excluded by performing variant tracing across different tissues. This off-target monitoring is best achieved through NGS readout of variant composition in on- and off-target tissues [149], which greatly benefits from the immense read-depths achieved with high-throughput (Illumina) sequencing. For randomization strategies that alter longer stretches of the *cap* sequence (*e.g.*, DNA family shuffling), variant readout by Illumina sequencing is not an option due to the limitation in read length. A work-around has been proposed by Pekrun *et al.*, who employed saturated barcoding of shuffled *cap* libraries [141]. This enables tracing of variant enrichment and off-targeting through high-throughput sequencing of barcodes. The *cap* sequences of selected candidates can then be identified through Sanger sequencing, using the DNA barcode as amplification primer (Figure 3C, (i)). Similar barcoding strategies generate lookup-tables, where each cloned *cap* variant is assigned to a specific barcode [19,140]. (d) Lastly, another issue with conventional capsid library screening is the enrichment of non-functional candidates. If enriched *cap* variants are selected or traced through PCR-amplification from tissue DNA, then their actual capacity to functionally transduce cells is not interrogated. PCR will not only amplify vector genomes from variants that successfully delivered their genome into the nucleus of a cell, but also from capsids that only attached to the cell surface without entering or that failed during intracellular trafficking [154]. Unfortunately, readout of *cap* RNA is not feasible for *rep-cap* constructs, as the p40

promoter does not allow sufficient transcription of *cap* in the absence of helper-virus co-infection [74]. To implement a screening for functional candidates, several modified selection schemes have been developed (Figure 3C, (ii)-(vi)). These include barcoding of reporter transgenes [144], which allows readout of DNA and RNA but requires a strategy for barcode/capsid association. Another option is Cre recombinase-based selection, which relies on a modified rescue-PCR. Here, only *cap* variants are amplified that have entered nuclei of Cre-expressing cells and have undergone Cre-lox recombination [145]. Furthermore, several groups have recently employed modified library constructs that offer a boosted *cap* RNA expression for enabling RNA-based enrichment of functional variants. This was achieved by either employing tissue-specific promoters that replace the *rep* gene (TRACER [146] and DELIVER [147] platforms), or by boosting p40 expression through the insertion of an inverted reporter cassette [74]. All of these alternative schemes have demonstrated advantages over conventional DNA-based selection. Apart from vector delivery, however, the efficiency and specificity of transgene expression can also be modulated through optimization of expression cassettes.

### 1.1.5 AAV-based transgene expression

In rAAVs, the *rep* and *cap* genes are removed from the ITR-flanked genome, allowing packaging of up to 4.9 kb of transgenic cargo (2.4 kb for scAAV; Figure 1B-C) [14]. Minimal requirements for transgene expression aside from the coding sequence are the promoter and polyadenylation elements, but modulators that control expression or stabilize mRNA can be included as well. Stabilization of transgene mRNA and prolonged expression can be achieved, for instance, by including a synthetic intron [157] and/or a woodchuck hepatitis virus post-transcriptional regulatory element (WPRE) [158,159]. Three main concerns for rAAV-based transgene expression will be addressed in the following paragraphs: (i) the limited packaging capacity of AAV vectors, (ii) achieving sufficient transgene expression to reach therapeutic efficacy, and (iii) restricting transgene expression to the intended target cells.

A common problem of AAV vectors is the insufficient packaging size for inclusion of large transgenes, large modulator sequences, or multiple expression cassettes [160]. This is commonly faced for transgenes such as dystrophin, CFTR or Cas9, whose cDNAs either exceed the packaging limit, or hinder inclusion of efficient modulator elements or other essential components (*e.g.*, strong sgRNA expression cassettes for Cas9). Several solutions have been proposed, including decreasing the size of the employed modulator

sequences. Employing smaller versions of WPRE (WPRE3) [159], small synthetic polyadenylation signals [161] and truncated promoters can already free enough packaging space for larger transgenes [159]. Shorter but highly functional promoter variants were derived by truncating full-length counterparts, generating small promoters such as mini-CMV [162], EF1α short promoter (EFS; EF1α promoter devoid of the 3' intron) [163], and gfaABC1D (derived from the human GFAP promoter) [164]. Another possibility is the application of smaller functional versions of the therapeutic transgene itself. For instance, mini- and micro-versions of the 11.5 kb coding sequence of dystrophin have been developed successfully [165]. Elevidys, the first authorized gene therapy for treatment of Duchenne muscular dystrophy, employs such a micro-dystrophin with a smaller coding sequence that enables packaging in ssAAV [114,166]. When multiple transgenes are to be expressed from a single vector (*e.g.*, heavy and light chains of a therapeutic anti-HIV antibody) [167], polycistronic expression cassettes can be implemented by employing 2A self-processing sequences [168]. These allow ribosome skipping, and thus expressing two proteins within a single expression cassette without relying on two sets of promoters and polyadenylation signals. If these options still do not grant sufficient packaging space, another solution is the utilization of dual-vector systems [160]. These allow packaging of coding sequences that exceed the packaging limit through mechanisms such as homologous recombination between two vectors [169], or RNA *trans*-splicing across two vectors that have formed a combined episome [170]. Another option is the intein-based recombination of two separate protein halves expressed from the dual vectors [171-173]. All dual vector systems are, however, inherently limited by the necessity for both vector parts to transduce the same cell, and thus require highly efficient transduction or high-dose administration.

Another issue for transgene expression is the requirement to generate sufficiently high and persistent levels of the gene product to achieve lasting therapeutic efficacy. Apart from stabilizing mRNA through efficient polyadenylation signals, synthetic introns, and WPRE as discussed above, transgene expression is mostly influenced by the transcriptional efficiency driven through enhancer and promoter sequence [174]. In clinical studies with AAV vectors, the most frequently employed promoters are the ubiquitous CMV, CAG and CBA [123]. The CMV promoter contains the immediate/early enhancer and promoter elements of cytomegalovirus [175-177]. The CAG promoter employs the same enhancer, but utilizes a chicken beta-actin promoter element and a synthetic intron [178]. CBA is essentially the same as CAG, but lacks the 3' intron [179]. All three are known for

enabling strong and ubiquitous transgene expression, but especially CMV is reported to undergo gradual silencing and thus shows decreased expression levels over time in some tissues [176,180-182]. For systemic administration of AAV vectors, the capsid transduction profile is usually not limited to the target tissue due to broad vector tropism. Thus, the application of ubiquitous promoters can induce transgene expression in off-target tissues as well. This can entail problems such as toxicity in dorsal root ganglia [183] or transgene-induced immune responses [184]. To overcome this, expression can be modulated depending on cell-type identity by employing tissue-specific promoters or by including repressive elements. An example of such a repressor is a micro-RNA-122 (miR-122) switch [174,185], which allows downregulation of transgene expression in liver tissue. This is achieved by including binding sites for miR-122 into the 3'-untranslated region (UTR). In cells expressing high levels of miR-122 (*i.e.*, liver cells), miR-122 will bind to the transgene mRNA and cause its degradation. Tissue-specific promoters are also capable of localizing transgene expression to intended target cells. For instance, following transduction of the CNS, the human synapsin I (SYN) promoter can drive neuron-specific expression [186], while the GFAP promoter can limit transgene expression to astrocytes [164]. Other promoters offer muscle-specific expression, such as MHCK7 (employed in Elevidys [114]) and SPc5-12, or liver-specific expression, such as LP1 [187-189]. Although a large toolset of promoter and modulator sequences exists, it is often difficult to identify the optimal components for a given task since detailed comparisons of expression profiles are tedious.

## 1.2 Inducing RNA knockdown *via* AAV vectors

While the currently authorized AAV gene therapy products (see Table 1) all focus on overexpression of a therapeutic transgene (*e.g.*, *SMN1* overexpression in patients with spinal muscular atrophy caused by *SMN1* mutation), other indications may instead require downregulating the expression of a cellular or pathogenic gene. This can be achieved by permanently altering the genome sequence in order to prevent expression entirely, a process termed gene "knockout". Alternatively, if modifications in the cellular genome are not desired, "knockdown" systems may be applied. These systems downregulate gene expression by preventing transcription or by destroying the transcribed RNA. Two knockdown systems compatible with AAV-based delivery will be discussed in the following chapters, namely RNA interference (RNAi, 1.2.1) and Cas13d (1.2.2).

### 1.2.1 RNA interference (RNAi)

RNAi is an endogenous pathway driving post-transcriptional regulation of eukaryotic gene expression [190]. By this pathway, micro (mi)RNAs regulate a large set of eukaryotic gene products. miR-122, for instance, is essentially involved in liver function and development [191]. miRNAs are endogenously expressed, small non-coding RNAs with a prominent imperfect hairpin structure. They are transcribed by RNA polymerase II as primary (pri-)miRNAs and undergo several processing steps before acting as post-transcriptional modulators [192]. First, (i) pri-miRNAs are cleaved in the nucleus by the microprocessor complex involving Drosha and DGCR8 [193]. (ii) The resulting precursor (pre-)miRNA is then transported into the cytoplasm by Exportin 5 [194]. (iii) There, after binding to Dicer and TRBP, the pre-miRNA is cleaved at the hairpin structure, generating a sense-antisense duplex [195]. (iv) Following cleavage by Dicer, the recruitment of an Argonaute protein (AGO1-4) yields the RISC-loading complex (RLC) [196,197]. (v) Within the RLC, the RNA duplex is disengaged, and its guide strand is loaded onto an AGO protein, while the passenger strand is discarded. (vi) This generates the RNA-induced silencing complex (RISC) containing the guide strand bound by AGO, as well as GW182 protein [198]. Binding of the guide strand to a complementary mRNA induces silencing thereof either through mRNA degradation or translational repression [190]. If sufficient sequence complementarity between guide strand and mRNA target occurs, AGO2 can induce targeted cleavage and thus most efficient degradation of the bound mRNA [199].

After the discovery of double-stranded RNAs efficiently interfering with gene expression [200], short interfering (si)RNAs were quickly adopted as a tool for studying and manipulating eukaryotic gene expression [201]. siRNAs are synthetic sense-antisense duplex RNAs with a size of ~21 nt that can be introduced into cells *in vivo* through excipients such as (lipid) nanoparticles or polymers [202]. They enter the miRNA pathway in the cytosol at step (v), where they are loaded onto RLC and later RISC. This offers a simple way of introducing targeted transcript repression defined by sequence alone. Besides target sequence specificity, key considerations for siRNA applications are nucleotide modifications, excipient formulation and administration route [202].

Even when modified to achieve better stability, siRNAs are only able to induce transient silencing of target genes. To achieve lasting post-transcriptional regulation as is induced by miRNAs, the interfering RNA needs to be expressed within the cell. This can be achieved by using AAV vectors for ideally tissue-specific delivery of transgene cassettes expressing miRNA-like effectors [203]. The most basic design for such effectors is

presented in short hairpin (sh)RNAs. They are designed as stem-loop-stem hairpins to resemble the structure of a miRNA tip, but can skip processing by Drosha [204]. Similar to siRNAs, the 5' stem of shRNAs directly matches 17-23 nt of the target mRNA sequence (sense), whereas the 3' stem (antisense) is complementary to the target sequence and will function as the guide strand in RISC. Transcription of shRNAs is commonly driven by strong, constitutive RNA polymerase III promoters such as U6, 7SK or H1 [205,206]. Following transcription, shRNAs are exported from the nucleus to the cytosol *via* Exportin 5 (step (ii) in the miRNA pathway described above) [194]. There, shRNAs are processed by Dicer before acting in RISC like miRNAs and siRNAs. Although enabling persistent and efficient silencing through knockdown of target mRNAs, shRNAs have demonstrated toxicity due to saturation of the endogenous miRNA machinery [207]. To overcome this, several alternative routes were pursued, including modified effectors that enable Dicer-independent processing by AGO2 (AgoshRNAs/agshRNAs) [208]. Furthermore, miRNA-like effectors can be expressed from ubiquitous or tissue-specific RNA polymerase II promoters (discussed in section 1.1.5), which offer reduced toxic side effects [209]. Another option to circumvent unspecific effects due to interference in the endogenous miRNA pathway is the application of entirely exogenous systems that still allow targeted RNA knockdown.

### 1.2.2 Expression control *via* CRISPR/Cas systems

An exogenous pathway allowing sequence-directed RNA degradation is presented by the clustered regularly interspaced short palindromic repeats (CRISPR) system, specifically its sub-system CRISPR class 2 type VI-D and the effector Cas13d [210].

CRISPR systems are a form of prokaryotic adaptive immunity towards infections with phages [211]. They function in three consecutive steps: (i) adaptation, (ii) transcription and (iii) interference [212]. (i) Upon infection, CRISPR-associated proteins Cas1 and Cas2 acquire protospacer sequences of the invading virus and integrate it into the host CRISPR locus between adjacent repeat sequences, one of which is duplicated in this process [213]. (ii) The CRISPR locus is then transcribed and subsequently processed into individual CRISPR (cr)RNAs [214]. (iii) crRNAs assemble with Cas effectors into ribonucleoprotein (RNP) complexes, where the spacer sequence mediates binding to a complementary target DNA or RNA. After binding, the Cas effector induces target cleavage, thus preventing virus infection [212].

Two classes of CRISPR systems exist, class 1 with CRISPR types I, III and IV, and class 2 with types II, V and VI [215]. Class 1 systems utilize multiple Cas proteins as effector modules, while class 2 systems employ a single Cas effector protein for interference and in some cases crRNA processing. Of all CRISPR variants, the class 2 type II system and its effector Cas9 have been applied first for directed DNA targeting in mammalian cells [216]. Cas9 can be targeted through a ~20 nt spacer within the crRNA that is complementary to the target sequence. For target binding, Cas9 requires the presence of a protospacer-adjacent motif (PAM) in the target DNA [217]. Although naturally depending on both a target-specific crRNA and a universal *trans*-activating CRISPR (tracr)RNA, the two have been engineered into a single guide (sg)RNA for improved applicability [216]. CRISPR/Cas9 has been applied widely to introduce targeted double-strand breaks in genomic DNA. These are repaired by the host cell either through non-homologous end-joining, leaving random insertions or deletions (indels), or through homology-directed repair (HDR) in the presence of a homology repair template [218]. The CRISPR/Cas9 system is compatible with AAV vectors, where the Cas9 effector is typically expressed from an RNA polymerase II cassette (as described in section 1.1.5) and the sgRNA is expressed from a secondary RNA polymerase III expression cassette [219]. As such, an AAV5-Cas9 vector (EDIT-101) has been developed for treatment of Leber congenital amaurosis type 10, a type of eye disorder which is caused by a mutation in CEP290 that induces aberrant splicing and premature translation termination [220]. Subretinal delivery of EDIT-101 demonstrated successful correction *via* Cas9-based targeting of the cryptic splice donor in pre-clinical testing and has since traversed into a clinical trial (ClinicalTrials.gov Identifier: NCT03872479).

Apart from inducing double-strand breaks, modified versions of CRISPR/Cas9 have been adopted for a wide variety of applications. Among others, these include expression modulation and base-editing, which is achieved by implementing modifications of Cas9 such as full or partial catalytical inactivation (dead (d)Cas9 or Cas9 nickase) and fusion to other effector proteins [221,222]. This can mitigate risks for dire side effects like chromosomal rearrangements [223]. A prominent example for mutation-free silencing is CRISPR interference (CRISPRi). CRISPRi achieves transcriptional repression by directing a catalytically dead (d)Cas9 to the promoter or 5'-UTR region of a target gene, thereby blocking transcription initiation or elongation [224]. This repression can be enhanced by fusing a secondary effector domain to dCas9. Fusion to a KRAB domain (Krüppel-associated box domain of transcriptional repressor Kox1), for instance, allows

for more efficient transcriptional repression by recruiting chromatin-modifying complexes that induce formation of heterochromatin and transcriptional repression [225]. By using a transcriptional activation domain instead, such as VP64, the opposite effect can be achieved, inducing chromatin accessibility and enhanced transcription [225-227]. These strategies are compatible with AAV vectors but are often limited by the vectors packaging capacity [228,229]. Especially for large Cas9 variants (*e.g.*, the first described and most widely applied *Streptococcus pyogenes* (*Sp*)Cas9) and Cas9 fusion effectors, dual vector systems are necessary for packaging their large cDNA sizes and including appropriate modulator sequences [173,230,231].

An option for achieving RNA-guided silencing by CRISPR effectors with single AAV vectors uses targeting of transcribed RNA for post-transcriptional silencing instead of blocking of transcription. This is achieved by utilizing a different type of CRISPR system, *i.e.*, Cas13 effectors of the class 2 type VI CRISPR family that bind to and cleave target RNA instead of DNA [215]. Of these effectors, Cas13d and specifically the Cas13d ortholog of *Ruminococcus flavefaciens* (CasRx) is highly promising for AAV-based applications [210]. Due to its small size compared to other Cas effectors, Cas13d and an accompanying gRNAs cassette can be packaged into single AAV vectors. Like other Cas13 variants, Cas13d can process crRNA arrays into individual gRNAs, thus simplifying multiplexed targeting. Directed by a target-specific gRNA, the HEPN domains of Cas13d induce cleavage of the bound RNA target [232]. Potentially outperforming silencing by RNAi and CRISPRi, Cas13d presents a useful tool for AAV-based transcription manipulation [210]. Since its discovery in 2018 [210], AAV-delivered Cas13d has been used for multiple applications [233-235].

### 1.2.3 Applications

In the following subchapters, two applications will be briefly introduced that require or benefit from AAV-mediated degradation of targeted RNAs. In both cases, the targeted RNAs are involved in pathogenesis. Thus, degradation of these targets can assist in understanding and potentially countering the respective disease. For the first sub-chapter, the target RNA is derived from a viral infection (SARS-CoV-2), whereas the second section focuses on a cellular RNA target that is involved in metabolic disease progression (CD44).

#### *1.2.3.1 Targeting RNAs of pathogenic viruses: SARS-CoV-2*

In 2019, the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) emerged as a novel human airway infection [236]. It quickly spread across the human population and soon reached pandemic status [237]. Although mostly causing acute infections with mild to moderate disease symptoms of the upper airway tract, patients can develop severe disease including acute respiratory distress syndrome (ARDS) [238] and lasting post-acute sequelae [239]. Vaccinations and neutralizing antibodies against SARS-CoV-2 are available, but achieving lasting efficacy thereof is difficult due to rapid antigenic evolution and rapid emergence of novel viral variants [240,241].

As a member of *sarbecoviruses*, SARS-CoV-2 is an enveloped virus with a 30 kb positive-stranded RNA genome [242]. Infectious virus is transmitted through aerosols and initially infects the upper airway tract [243,244]. Here, SARS-CoV-2 attaches *via* its spike protein (S) to the human angiotensin-converting enzyme 2 (ACE2) on the host cell surface [245]. Next, the host proteinase TMPRSS2 cleaves S, causing a conformational change that enables fusion of viral and cell membranes and deposition of viral RNA in the cytosol [246]. Translation of the SARS-CoV-2 genomic RNA from the open reading frame (ORF) 1ab generates two polypeptides (pp1a and pp1b), which are processed into 16 non-structural proteins (NSPs) [247]. The NSPs, including the RNA-dependent RNA polymerase (RdRP), drive viral transcription and replication. Approaches for therapeutic intervention of SARS-CoV-2 infection include blocking through neutralizing antibodies [248] and inhibition of viral replication or processing (or host factors) through small molecule drugs [249]. Another potential therapeutic target is the viral genomic RNA itself [250-252]. As discussed in the preceding chapters, RNA-guided degradation of target RNAs is feasible through several mechanisms, and AAV-based delivery can introduce the required components thereof into target cells. Hence, AAV-induced targeting of SARS-CoV-2 genomic RNA may enable an interference with or prevention of viral infection.

#### *1.2.3.2 Targeting endogenous RNAs: role of CD44 in metabolic steatohepatitis*

Metabolic dysfunction-associated fatty liver disease (MAFLD; formerly termed non-alcoholic fatty liver disease, NAFLD) is a highly prevalent chronic liver disorder that includes liver steatosis (fatty liver disease) and, potentially progressing thereof, metabolic dysfunction-associated steatohepatitis (MASH; formerly termed non-alcoholic steatohepatitis, NASH) [253,254]. MAFLD is a leading cause of liver failure with a global prevalence of ca. 25% and predicted increase in the near future [255-257]. While liver

steatosis describes an accumulation of fat in otherwise healthy liver tissue, MASH is distinguished by an accompanying inflammation leading to liver injury. This disorder can progress into liver fibrosis, cirrhosis and hepatocellular carcinoma (HCC) [254]. MAFLD and MASH are related to metabolic syndrome, unhealthy lifestyle, obesity and insulin resistance, and can increase risks of cardiovascular disease [258]. While conservative treatment options mostly focus on weight loss, dietary change and change in lifestyle, a targeted treatment for halting MAFLD progression is not available [259]. Recent findings suggest that the process of inflammation in MASH may be regulated by CD44 [260,261]. CD44 is a cell-surface protein interacting with the extracellular matrix and contributing to the recruitment and activation of macrophages in MASH [260]. Knockout of CD44 and treatment with anti-CD44 antibodies has demonstrated a decrease in MASH progression and steatosis in mouse models [260,262]. Although known to be expressed across different cell types including macrophages, the exact mechanism by which CD44 contributes to liver inflammation in MAFLD is not fully understood. Targeted downregulation of CD44 in liver cells may not only contribute to understanding its involvement in MASH progression, but may also enable a therapeutic intervention. Due to CD44's multifaceted physiological functions, however, it is imperative to limit this downregulation to disease-inferring cell types, which may include macrophages and stellate cells [260]. This directed targeting may be achieved through AAV vectors. However, the liver is a complex organ, built with an intricate structure and multiple cell types of different function [263]. Cell type-directed targeting is therefore difficult.

The liver consists of hexagonal lobules, which contain a central vein that is connected to radiating portal veins and hepatic arteries through the liver sinusoids (Figure 4). Liver sinusoids receive blood from both portal vein and hepatic artery, and are thus highly exposed to antigens. Liver-residing macrophages called Kupffer cells remain within the sinusoid lumen, making up 10% of all liver cells [263]. The capillary-like sinusoid is lined by a specialized fenestrated endothelium of liver sinusoidal endothelial cells (LSECs). LSECs express major histocompatibility complexes type I and II, and can induce immunotolerance by stimulating regulatory T cells [264]. Lacking a basal membrane, LSECs are separated from hepatocytes only by the perisinusoidal space of Disse. Here, hepatic stellate cells (HSCs) reside, a type of vitamin A-storing cells that contribute to fibrosis upon liver injury [263,265]. Apart from the non-parenchymal liver cells (NPCs; Kupffer cells, LSECs, HSCs and immune cells), hepatocytes make up 70% of all liver cells [266]. These are replicating cells that carry out multiple functions including production of bile,

detoxification and metabolism of lipids, proteins, and carbohydrates. Bile is secreted from hepatocytes into bile canaliculi, which have an antiparallel flow direction relative to the sinusoid. From the canaliculi, bile is collected in the bile ducts that are lined by cholangiocytes, *i.e.*, epithelial cells closely related to hepatocytes [267]. Taken together, bile duct, portal vein and hepatic artery make up the portal triad cornering the hepatic lobule. AAV-based transduction of liver cells is possible with many capsid variants and is clinically utilized for expression of Factor VIII and IX for treatment of hemophilia A and B, respectively [112,113]. This transduction is mostly focused on but not limited to hepatocytes. Selective transduction of specific liver cell types other than hepatocytes is difficult, however, and requires the creation of novel vector capsids.

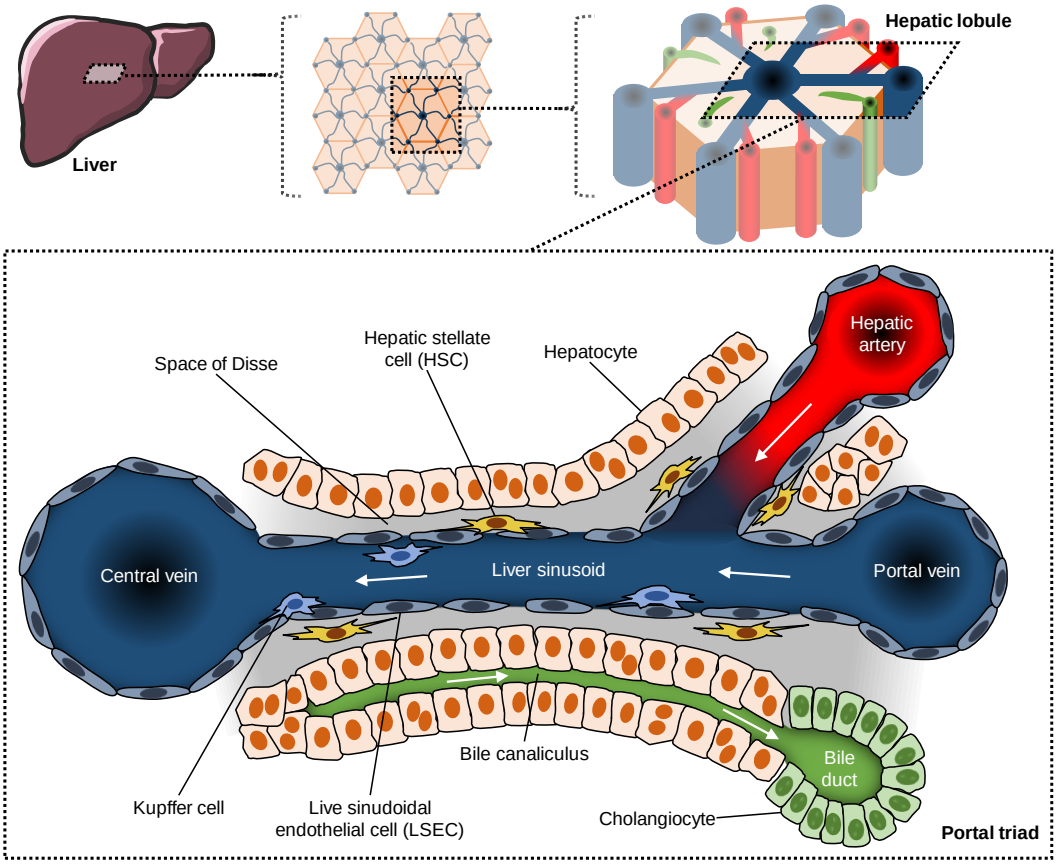



**Figure 4: Structural and cellular composition of the liver.** Liver tissue is structured in hexagonal lobules. Each lobule contains a central vein and cornering portal triads, which consist of a hepatic artery, portal vein and bile duct [263]. Blood from the portal vein and hepatic artery flows through the liver sinusoid into the central vein. Hepatocytes contribute to 70% of all liver cells [266]. Among other functions, these parenchymal cells produce bile which is secreted into bile canaliculi leading to bile ducts lined by cholangiocytes [267]. Non-parenchymal liver cells include liver sinusoidal endothelial cells (LSEC), hepatic stellate cells (HSC) and liver-residing macrophages called Kupffer cells (amongst other immune cell types). LSECs line the liver sinusoid between portal vein/hepatic artery and central vein, creating a fenestrated endothelium. The perisinusoidal Disse space between LSECs and hepatocytes contains HSCs, a fibrogenic cell type which is activated upon liver injury [265]. Kupffer cells reside in the sinusoid lumen and thus have direct access to antigens delivered through the portal vein or hepatic artery.

## 1.3 Aims of this thesis

In this thesis, the AAV toolbox is expanded with the aim of creating vectors that allow (i) cell type-specific transduction within the liver, (ii) optimized transgene expression, and (iii) efficient knockdown of endogenous and viral target RNAs.

In the first part of the study, I utilized reporter knockdown assays for comparative evaluation of shRNA and Cas13d effectors. Both systems offer RNA-guided sequence specific degradation of target RNAs, but employ either the endogenous RNAi machinery (shRNAs) or an exogenous system evolved in prokaryotes as an adaptive defense against viral infection (CasRx). After confirming efficient knockdown of reporter targets, I used both systems with the goal of silencing the expression of CD44, a potential driver of metabolic dysfunction-associated steatohepatitis (MASH). Targeted downregulation of CD44 may assist in understanding MASH disease progression and potentially offer therapeutic interventions in the future. I then applied the more effective of the two RNA knockdown strategies (shRNA-induced RNAi) for testing the suppression of cellular infections with an RNA virus, namely SARS-CoV-2. The focus here was on the investigation of viral susceptibility to single or multiplexed shRNA effectors. Viral suppression *via* RNAi triggers offers insights into SARS-CoV-2 escape mutagenesis after targeting of viral genomic RNA.

The second part of this study focused on screening of promoter and capsid variants for the creation of cell-type-specific AAV vectors. To improve combinatorial evaluation of various promoters, I validated the results of a preceding promoter library screen by individually testing four promoters *in vivo* following systemic delivery with AAV9. Next, I investigated the liver-directed expression of the GFAP promoter with the goal of dissecting its cell-type-specific transcription profile.

AAV capsid engineering enables the creation of synthetic vectors with enhanced or specialized transduction profiles, and heavily relies on the directed evolution of novel capsids. To improve selection conditions for capsid evolution, I employed Cas9-mediated negative selection aiming to deplete non-specific and unwanted variants from *cap* libraries. Furthermore, I implemented RNA-based functional screening in on- and off-target cells by CMV promoter-driven ubiquitous expression of the *cap* gene. Screening of *cap* variants on the RNA level was then used for the directed evolution of AAV6 peptide display capsids *in vivo* across different mouse liver cell types. Such vectors capable of cell-type-specific transduction within the liver are direly needed for studying and countering multiple pathogenic indications including MASH.

# 2 Materials and methods

## 2.1 Materials

### 2.1.1 Devices

**Table 2: Laboratory devices used in this thesis.**

| Application | Device | Provider |
|---|---|---|
| **Pipetting** | | |
| Pipetting | accu-jet pro | Brand GmbH Co Kg (Wertheim am Main, Germany) |
| Pipetting | E-1840, E-1841 | Eppendorf (Hamburg, Germany) |
| Pipetting | P2, P10, P20, P200, P1000 | Gilson (Middleton, USA) |
| Pipetting | Research plus multichannel (12-well; 10 and 100 µL) | Eppendorf (Hamburg, Germany) |
| **Centrifugation** | | |
| Centrifugation | Allegra X-12 | Beckman Coulter GmbH (Krefeld, Germany) |
| Centrifugation | Avanti J-26XP | Beckman Coulter GmbH (Krefeld, Germany) |
| Centrifugation | JA-10 rotor | Beckman Coulter GmbH (Krefeld, Germany) |
| Centrifugation | Microcentrifuge 5415R | Eppendorf (Hamburg, Germany) |
| Ultracentrifugation | Optima L-90K | Beckman Coulter GmbH (Krefeld, Germany) |
| Ultracentrifugation | Fixed angle rotor type 70 Ti / 70.1 Ti | Beckman Coulter GmbH (Krefeld, Germany) |
| **Cell culture** | | |
| Automated cell counter | Countess | Thermo Fisher Scientific (Waltham, USA) |
| Cell culture hood | Herasafe KS 12 | Thermo Fisher Scientific (Waltham, USA) |
| Incubator | Heracell 150 $CO_2$ incubator | Thermo Fisher Scientific (Waltham, USA) |
| **Gel Electrophoresis and blotting** | | |
| Agarose gel running system | EasyPhor | Biozym Scientific (Hessisch Oldendorf, Germany) |
| Agarose gel power device | E385 power supply | Consort (Turnhout, Belgium) |
| Agarose gel imaging | Gel Doc XR | Bio-Rad (Hercules, USA) |
| Dot blot | Bio-Dot Apparatus | Bio-Rad (Hercules, USA) |
| SDS-PAGE system | Mini-PROTEAN Tetra Cell | Bio-Rad (Hercules, USA) |
| SDS-PAGE power device | PowerPac HV | Bio-Rad (Hercules, USA) |
| Imaging for dot blot, silver stain, agarose gels | Azure 400 Visible Fluorescent Imager | Azure Biosystems (Dublin, USA) |
| **Droplet digital (dd)PCR** | | |
| Droplet generation | QX200 Droplet Generator | Bio-Rad (Hercules, USA) |
| Plate sealing | PX1 PCR Plate Sealer | Bio-Rad (Hercules, USA) |
| PCR | C1000 Touch Thermal Cycler | Bio-Rad (Hercules, USA) |
| Droplet analysis | QX200 Droplet Reader | Bio-Rad (Hercules, USA) |
| **Miscellaneous** | | |
| Bacteria incubator | Heraus Function Line Incubator | Thermo Fisher Scientific (Waltham, USA) |

| Application | Device | Provider |
|---|---|---|
| Bacteria incubator (shaking) | Multitron | INFORS HT (Basel, Switzerland) |
| DNA quality assessment | 2100 Bioanalyzer | Agilent Technologies (Santa Clara, USA) |
| DNA quantification | Qubit 2.0 Fluorometer | Thermo Fisher Scientific (Waltham, USA) |
| DNA/RNA quantification | NanoDrop 2000 Spectrophotometer | Thermo Fisher Scientific (Waltham, USA) |
| Electroporation | Gene Pulser Xcell | Bio-Rad (Hercules, USA) |
| Incubation and mixing | Mixing Block MB 102 | Bioer Technology (Hangzhou, China) |
| Indirect IF assay | Odyssey CLx | LI-COR Biosciences (Lincoln, USA) |
| Iodixanol density gradient tube sealing | Tube Sealer 342428 | Beckman Coulter GmbH (Krefeld, Germany) |
| Luminescence measurement | GloMax Navigator Microplate Luminometer | Promega (Madison, USA) |
| Mixing | Vortex Genie 2 | neoLab (Heidelberg, Germany) |
| PCR cycler | Mastercycler pro S | Eppendorf (Hamburg, Germany) |
| qRT-PCR cycler | Corbett Rotor-Gene 6000 | Qiagen (Hilden, Germany) |
| qRT-PCR cycler (96-well plates) | StepOne Plus | Applied Biosystems/Thermo Fisher Scientific (Waltham, USA) |
| Sterile hood | Captair Bio Smart PCR-Hood | Erlab (Val-de-Reuil, France) |
| Thawing and pre-warming | TW12 Water Bath | Julabo (Seelbach, Germany) |
| Tissue lysis | TissueLyser LT | Qiagen (Hilden, Germany) |

### 2.1.2 Software

**Table 3: Software employed for this thesis.**

| Software | Provider |
|---|---|
| ApE - A Plasmid Editor v2.0.61 | M. Wayne Davis [268] |
| ImageJ 1.53c | Wayne Rasband, NIH (Bethesda, USA) [269] |
| Microsoft Office 365 | Microsoft (Redmond, USA) |
| Python 3.6 | Python Software Foundation (Wilmington, USA) |
| QuantaSoft Software, Regulatory Edition | Bio-Rad (Hercules, USA) |
| Rotor Gene Q-Series Software | Qiagen (Hilden, Germany) |
| GraphPad Prism 8.0.1 | GraphPad Software (Boston, USA) |
| 2100 Expert | Agilent (Santa Clara, USA) |

### 2.1.3 Consumables

**Table 4: Consumables used for the conduction of experiments.**

| Consumable | Specification | Provider |
|---|---|---|
| 8-strip PCR tubes | 0.2 mL | Biozym Scientific (Hessisch Oldendorf, Germany) |
| Amicon Ultra-15 Centrifugal Filter Units | 100,000 MWCO | Merck (Darmstadt, Germany) |
| Blotting Membran CN | 0.45 µm | neoLab (Heidelberg, Germany) |
| Cell culture dishes | 150 x 20 mm | Sarstedt (Nümbrecht, Germany) |

| Consumable | Specification | Provider |
|---|---|---|
| Cell culture flasks | 75 $cm^2$; 175 $cm^2$ | Greiner Bio-One (Kremsmünster, Austria) |
| Cell lifter | 3008 | Corning (New York, USA) |
| Centrifuge tubes | 500 mL | Corning (New York, USA) |
| Countess cell counting chamber slides | | Thermo Fisher Scientific (Waltham, USA) |
| ddPCR plates | 96-well, semi-skirted | Bio-Rad (Hercules, USA) |
| DG8 Cartridges for Droplet Generator | | Bio-Rad (Hercules, USA) |
| DNA LoBind Tubes | 1.5 mL | Eppendorf (Hamburg, Germany) |
| Droplet generator DG8 gasket | | Bio-Rad (Hercules, USA) |
| Electroporation cuvette | 1 mm | Biozym Scientific (Hessisch Oldendorf, Germany) |
| Erlenmeyer flasks | | DWK Life Sciences (Wertheim, Germany) |
| Filtered pipet tips | | Biozym Scientific (Hessisch Oldendorf, Germany) |
| Glass bottles | | DWK Life Sciences (Wertheim, Germany) |
| Glass test tubes | 160 x 15 mm | DWK Life Sciences (Wertheim, Germany) |
| MF-Millipore Membrane Filter | 0.025 µm MCE | Merck (Darmstadt, Germany) |
| MicroAmp optical adhesive film | | Thermo Fisher Scientific (Waltham, USA) |
| Microcentrifuge tubes | | Sarstedt (Nümbrecht, Germany) |
| Microlance 3 needles | 21 g / 30 g | BD (Franklin Lakes, USA) |
| Microplate 96-well | F-bottom, white | Greiner Bio-One (Kremsmünster, Austria) |
| Mini-PROTEAN TGX Precast Protein Gels | 7.5%, 10-well, 50 µl | Bio-Rad (Hercules, USA) |
| N-well tissue culture plates | 6/12/24/48 wells | Greiner Bio-One (Kremsmünster, Austria) |
| Pasteur capillary pipettes | 230 mm | Corning (New York, USA) |
| Petri dishes | 94 x 16 mm | Greiner Bio-One (Kremsmünster, Austria) |
| Pierceable foil heat seal | | Bio-Rad (Hercules, USA) |
| qPCR plate | 96 well, semi-skirted | Biozym Scientific (Hessisch Oldendorf, Germany) |
| Quick-Seal centrifuge tubes | 25 × 89 mm | Beckman Coulter (Brea, USA) |
| Re-seal polyallomer centrifuge tubes | 16 × 76 mm | Seton Scientific (Petaluma, USA) |
| Scalpel blades | Carbon steel, figure 23 | Heniz Herenz (Hamburg, Germany) |
| Serological pipettes | | Greiner Bio-One (Kremsmünster, Austria) |
| Stainless steel beads | 3 mm | Qiagen (Hilden, Germany) |
| Syringes | Luer-Lok (3 mL) | BD (Franklin Lakes, USA) |
| Tubes (Falcon) | 15 mL / 50 mL | Corning (New York, USA) |

### 2.1.4 Kits

**Table 5: Commercial kits used in this thesis.**

| Application | Kit | Provider |
|---|---|---|
| Bead-based PCR amplicon purification | ProNex Size-Selective Purification System | Promega (Madison, USA) |
| Bioanalyzer 2100 | DNA 1000 Kit | Agilent Technologies (Santa Clara, USA) |

| Application | Kit | Provider |
|---|---|---|
| cDNA synthesis | High-Capacity cDNA Reverse Transcription Kit | Applied Biosystems/Thermo Fisher Scientific (Waltham, USA) |
| cDNA synthesis | SuperScript IV VILO Master Mix with ezDNase Enzyme | Invitrogen/Thermo Fisher Scientific (Waltham, USA) |
| cDNA synthesis | SuperScript IV First-Strand Synthesis System with ezDNase Enzyme | Invitrogen/Thermo Fisher Scientific (Waltham, USA) |
| cDNA synthesis | iScript Reverse Transcription Supermix for RT-qPCR | Bio-Rad (Hercules, USA) |
| ddPCR | ddPCR Supermix for Probes (No dUTP) | Bio-Rad (Hercules, USA) |
| DNA and RNA extraction (combined) | AllPrep Mini/Micro Kit | Qiagen (Hilden, Germany) |
| DNA extraction (cells/tissues) | DNeasy Blood & Tissue Kit | Qiagen (Hilden, Germany) |
| DNA extraction (from agarose gels) | QIAquick Gel Extraction Kit | Qiagen (Hilden, Germany) |
| DNA purification (amplicons *etc*.) | QIAquick PCR Purification Kit | Qiagen (Hilden, Germany) |
| DNA purification (amplicons *etc*.) | QIAquick Nucleotide Removal Kit | Qiagen (Hilden, Germany) |
| Dual luciferase assay | Dual-Luciferase Reporter Assay System | Promega (Madison, USA) |
| NGS indexing PCR | KAPA HiFi HotStart ReadyMix PCR Kit | Roche (Basel, Switzerland) |
| NGS library preparation *via* adapter ligation | Ovation Low Complexity Library Preparation Kit | NuGEN/Tecan (Männedorf, Switzerland) |
| NGS library preparation, PCR-based | Nextera XT Index Kit v2 | Illumina (San Diego, USA) |
| On-column digestion of genomic DNA for RNA extraction | RNase-Free DNase Set | Qiagen (Hilden, Germany) |
| Plasmid purification | QIAprep Spin Miniprep Kit | Qiagen (Hilden, Germany) |
| Plasmid purification | PureYield Plasmid Midiprep System | Promega (Madison, USA) |
| Plasmid purification | NucleoBond PC 500 Maxi Kit | Macherey-Nagel (Düren, Germany) |
| qPCR (probe-based) | Sensimix II Probe Kit | Bioline (London, UK) |
| qPCR (SYBR green-based) | iTaq Universal SYBR Green Supermix | Bio-Rad (Hercules, USA) |
| Qubit DNA quantification | Qubit dsDNA HS Assay Kit | Thermo Fisher Scientific (Waltham, USA) |
| RNA extraction (cells/tissues) | RNeasy Mini Kit | Qiagen (Hilden, Germany) |
| sgRNA clean-up | RNA Clean & Concentrator-5 Kit | Zymo Research (Irvine, USA) |
| sgRNA synthesis | AmpliScribe T7 Flash Transcription Kit | Biozym Scientific (Hessisch Oldendorf, Germany) |
| sgRNA synthesis from DNA oligos | EnGen sgRNA Synthesis Kit, S. pyogenes | New England Biolabs (NEB; Ipswich, USA) |
| Silver staining | SilverQuest Silver Staining Kit | Invitrogen/Thermo Fisher Scientific (Waltham, USA) |

### 2.1.5 Reagents

**Table 6: Reagents used for experimental procedures.**

| Reagent | Provider |
|---|---|
| 0.25% Trypsin / EDTA | Gibco/Thermo Fisher Scientific (Waltham, USA) |
| 1 kb Plus DNA ladder | Thermo Fisher Scientific (Waltham, USA) |
| 1× Dulbecco's phosphate buffered saline (PBS) | Gibco/Thermo Fisher Scientific (Waltham, USA) |
| 10× Tris/Glycine/SDS Electrophoresis Buffer | Bio-Rad (Hercules, USA) |
| 5× HF Buffer | Thermo Fisher Scientific (Waltham, USA) |
| A20 antibody | [270] |
| Acetic acid | VWR chemicals (Fenenay-sous-Bais, France) |
| Ampicillin | Roth (Karlsruhe, Germany) |
| α-mouse-HRP antibody (115-035-003) | Jackson ImmunoResearch Laboratories (West Grove, USA) |
| anti dsRNA mAb J2, 10010200 | SCICONS/Nordic-Mubio (Wolfskoul, Netherlands) |
| Aqua B. Braun | B. Braun Avitum Saxonia GmbH (Melsungen, Germany) |
| Bacto agar | BD (Franklin Lakes, USA) |
| Bacto tryptone | BD (Franklin Lakes, USA) |
| Bacto yeast extract | BD (Franklin Lakes, USA) |
| Biozym LE Agarose | Biozym Scientific (Hessisch Oldendorf, Germany) |
| CutSmart buffer | NEB (Ipswich, USA) |
| ddPCR GEX HEX Assay Rpp30, *Mmu* (#10031255) | Bio-Rad (Hercules, USA) |
| Deoxynucleotide (dNTP) Solution Mix (10 mM of each) | NEB (Ipswich, USA) |
| Diluent B | NEB (Ipswich, USA) |
| Dithiothreitol (DTT) | Thermo Fisher Scientific (Waltham, USA) |
| DMEM, high glucose, GlutaMAX (61965026) | Gibco/Thermo Fisher Scientific (Waltham, USA) |
| DRAQ5 (ab108410) | Abcam (Cambridge, UK) |
| Droplet Generation Oil for Probes | Bio-Rad (Hercules, USA) |
| EDTA | GRÜSSING GmbH (Filsum, Germany) |
| Ethanol absolute | Merck (Darmstadt, Germany) |
| Ethidium Bromide 1% | Roth (Karlsruhe, Germany) |
| Fetal bovine serum (FBS) | Capricorn Scientific (Ebsdorfergrund, Germany) |
| Gel Loading Dye, Purple (6×) | NEB (Ipswich, USA) |
| Gelred Nucleic Acid Gel Stain | Biotium (Fremont, USA) |
| GeneRuler DNA Ladder Mix | Thermo Fisher Scientific (Waltham, USA) |
| Glycerol | VWR chemicals (Fenenay-sous-Bais, France) |
| hB2M-PE (#316317) | Biolegend (San Diego, USA) |
| Hydrochloric acid (HCl), 1.0 M | VWR chemicals (Fenenay-sous-Bais, France) |
| IRDye 800CW Donkey anti-Mouse IgG Secondary Antibody | LI-COR Biosciences (Lincoln, USA) |
| Isopropanol | Merck (Darmstadt, Germany) |
| Laemmli Sample Buffer 4× | Bio-Rad (Hercules, USA) |
| Lipofectamine 2000 | Invitrogen/Thermo Fisher Scientific (Waltham, USA) |
| Magnesium chloride ($MgCl_2$) | Applichem (Darmstadt, Germany) |

| Reagent | Provider |
|---|---|
| MEM Non-Essential Amino Acids Solution (100×) | Gibco/Thermo Fisher Scientific (Waltham, USA) |
| mH-2Kd-BV711 (#742864) | Biolegend (San Diego, USA) |
| NEBuffer 1.1, 2.1, 3.1 | NEB (Ipswich, USA) |
| Nuclease-free $H_2O$ | Qiagen (Hilden, Germany) |
| O'RangeRuler 10 bp DNA Ladder | Thermo Fisher Scientific (Waltham, USA) |
| OptiPrep (Iodixanol) | Progen (Heidelberg, Germany) |
| Orange DNA Loading Dye (6X) | Thermo Fisher Scientific (Waltham, USA) |
| PageRuler Plus Prestained Protein Ladder | Thermo Fisher Scientific (Waltham, USA) |
| Paraformaldehyde | Merck (Darmstadt, Germany) |
| Penicillin / Streptomycin (P/S) | Gibco/Thermo Fisher Scientific (Waltham, USA) |
| Phenol red | Merck (Darmstadt, Germany) |
| PolR2A primer/probe mix (Mm00839502_m1) | Thermo Fisher Scientific (Waltham, USA) |
| Polyethyleneimine (PEI MAX) | Polysciences Europe GmbH (Eppelheim, Germany) |
| Potassium acetate (KAc) | Honeywell (Seelze, Germany) |
| Potassium chloride (KCl) | GRÜSSING GmbH (Filsum, Germany) |
| RNAlater | Thermo Fisher Scientific (Waltham, USA) |
| Sodium chloride (NaCl) | GRÜSSING GmbH (Filsum, Germany) |
| Sodium Dodecylsulfate (SDS) | Serva (Heidelberg, Germany) |
| Sodium hydroxide (NaOH) 2 M | Merck (Darmstadt, Germany) |
| T4 DNA Ligase Buffer | NEB (Ipswich, USA) |
| TE Buffer | Thermo Fisher Scientific (Waltham, USA) |
| TrickTrack DNA Loading dye (6x) | Thermo Fisher Scientific (Waltham, USA) |
| TRIS | Roth (Karlsruhe, Germany) |
| TRIS-HCl | Roth (Karlsruhe, Germany) |
| Triton X-100 | Merck (Darmstadt, Germany) |
| Trypan Blue Solution, 0.4% | Thermo Fisher Scientific (Waltham, USA) |
| Tween20 | Roth (Karlsruhe, Germany) |
| WesternBright Chemiluminescence Substrate Quantum | Biozym Scientific (Hessisch Oldendorf, Germany) |

### 2.1.6 Enzymes

**Table 7: Enzymes used for this study.**

| Enzyme | Provider | Reference number |
|---|---|---|
| Antarctic Phosphatase | NEB (Ipswich, USA) | M0289S |
| AscI | NEB (Ipswich, USA) | R0558S |
| BamHI-HF | NEB (Ipswich, USA) | R3136S |
| BbsI-HF | NEB (Ipswich, USA) | R3539S |
| Benzonase | Merck (Darmstadt, Germany) | 101695 |
| BglI | NEB (Ipswich, USA) | R0143S |
| BglII | NEB (Ipswich, USA) | R0144S |
| BsmBI-v2 | NEB (Ipswich, USA) | R0739S |
| BstBI | NEB (Ipswich, USA) | R0519S |
| ClaI | NEB (Ipswich, USA) | R0197S |
| DNA Polymerase I, Large (Klenow) Fragment | NEB (Ipswich, USA) | M0210S |

| Enzyme | Provider | Reference number |
|---|---|---|
| DNase I (RNase-free) | NEB (Ipswich, USA) | M0303S |
| ezDNase enzyme | Thermo Fisher Scientific (Waltham, USA) | 11766051 |
| HindIII-HF | NEB (Ipswich, USA) | R3104S |
| NotI-HF | NEB (Ipswich, USA) | R3189S |
| OneTaq Quick-Load 2× Master Mix with Standard Buffer | NEB (Ipswich, USA) | M0486S |
| PacI | NEB (Ipswich, USA) | R0547S |
| Phusion Flash High-Fidelity PCR Master Mix | Thermo Fisher Scientific (Waltham, USA) | F548S |
| Phusion HS II | Thermo Fisher Scientific (Waltham, USA) | F549S |
| Proteinase K | Qiagen (Hilden, Germany) | 19131 |
| Proteinase K | NEB (Ipswich, USA) | P8107S |
| Q5 High-Fidelity 2X Master Mix | NEB (Ipswich, USA) | M0492S |
| RNase A | Qiagen (Hilden, Germany) | 19101 |
| SalI-HF | NEB (Ipswich, USA) | R3138S |
| SfiI | NEB (Ipswich, USA) | R0123S |
| SpCas9 (Cas9 Nuclease, *S. pyogenes*) | NEB (Ipswich, USA) | M0386T |
| SpeI-HF | NEB (Ipswich, USA) | R3133S |
| T4 DNA Ligase | NEB (Ipswich, USA) | M0202S |
| T4 Polynucleotide Kinase | NEB (Ipswich, USA) | M0201S |
| XhoI | NEB (Ipswich, USA) | R0146S |
| XmaI | NEB (Ipswich, USA) | R0180S |

### 2.1.7 Buffers

**Table 8: Utilized buffers and their composition.**

| Buffer | Ingredients |
|---|---|
| 15% iodixanol | 75.00% PBS-MK-NaCl, 25.00% OptiPrep |
| 25% iodixanol | 58.19% PBS-MK, 41.56% OptiPrep, 0.25% Phenol red stock |
| 40% iodixanol | 66.67% OptiPrep, 33.33% PBS-MK |
| 60% iodixanol | 99.75% OptiPrep, 0.25% Phenol red stock |
| Benzonase buffer | 150 mM NaCl, 50 mM TRIS-HCl (pH 8.0), 2 mM MgCl2 |
| LB agar | 1.5% Bacto agar, 1.0% NaCl, 1% Bacto tryptone, 0.5% Bacto yeast extract |
| LB medium | 1.0% Bacto tryptone, 1.0% NaCl, 0.5% Bacto yeast extract |
| P1 resuspension buffer | 50 mM Tris-HCl (pH 8.0), 10 mM EDTA, 100 µg/mL RNase A |
| P2 lysis buffer | 200 mM NaOH, 1% SDS |
| P3 neutralization buffer | 2.8 M Kac (pH 5.1) |
| PBS-MK | PBS (1×), 2.5 mM KCl, 1 mM MgCl2 |
| PBS-MK-NaCl | 1 M NaCl in PBS-MK |
| Phenol red stock | Nuclease-free H2O, 0.5% Phenol red |
| TAE buffer | 2 M TRIS, 1 M acetic acid, 50 mM EDTA |
| TBS-T | 1.25 M NaCl, 250 mM Tris/HCl, pH 7.4; 0.05% ween20 |

### 2.1.8 Bacterial strains

**Table 9: Bacterial strains used for cloning procedures.**

| Strain | Provider |
|---|---|
| E. cloni 10G Supreme | Lucigen (Madison, USA) |
| One Shot ccdB Survival 2 T1 R Competent Cells | Thermo Fisher Scientific (Waltham, USA) |
| MAX Efficiency DH5α Competent Cells | Thermo Fisher Scientific (Waltham, USA) |

### 2.1.9 Eukaryotic cells

**Table 10: Eukaryotic cells (cell lines and primary cells) used in this study.**

| Eukaryotic Cells | Origin |
|---|---|
| **Cell lines** | |
| Caco2 | Human (colorectal adenocarcinoma) |
| Hek293T | Human (embryonic kidney) |
| Hepa1-6 | Mouse (hepatoma) |
| Huh7 | Human (hepatoma) |
| Vero E6 | African green monkey (kidney) |
| **Primary cells** | |
| Primary human hepatocytes | Human; kindly provided by DZIF PHH Core Facility |

### 2.1.10 Oligonucleotides

DNA oligonucleotides were purchased from Merck (Darmstadt, Germany) or Integrated DNA Technologies (IDT; San Jose, USA). Probes were purchased from IDT with the indicated fluorophores and quenchers.

**Table 11: DNA oligonucleotides used for this thesis project.**

| Name | Sequence (5'->3') | Application |
|---|---|---|
| Lseq_fw | GATCTGGTCAATGTGGATTTG | cap rescue primer |
| Lseq_rev | GACCGCAGCCTTTCGAATGTC | cap rescue primer |
| scaff_fw | [PHOS]GTTTAAGAGCTATGCTGG | SpCas9 scaffold amplification |
| scaff_rev | ACTGCAGGTCTTCGGATCCAAAAAAAGCACCGACTCGG | SpCas9 scaffold amplification |
| rep2_resc_fw | AGACGCGGAAGCTTCGATCAA | cap rescue primer |
| cap2_resc_rev | ACAGAGGCGCGCCTTACAGATTACGAGTCAGGTATC | cap2 rescue primer |
| cap6_resc_fw | GTTGCCGTACGTCCTCGGCTCTG | cap6 rescue primer |
| cap6_resc_rev | ACTGTACTAGTTTACAGGGGACGGGTGAGGT | cap6 rescue primer |
| cap2_RT | TTACAGATTACGAGTCAGGTATC | reverse transcription |
| SV40pA_RT1 | CAAATTTCACAAATAAAGCTAG | reverse transcription |
| SV40pA_RT2 | GCATCACAAATTTCAC | reverse transcription |
| hACE2_f | TCCATTGGTCTTCTGTCACCCG | RTqPCR |
| hACE2_r | AGACCATCCACCTCCACTTCTC | RTqPCR |
| mCD44_f | AGCCCCTCCTGAAGAAGACT | RTqPCR |
| mCD44_r | GCGAGTACCATCACGGTTGA | RTqPCR |
| hGAPDH_f | ACCCACTCCTCCACCTTTGAC | RTqPCR |

| Name | Sequence (5'->3') | Application |
|---|---|---|
| hGAPDH_r | TGTTGCTGTAGCCAAATTCGTT | RTqPCR |
| mGAPDH_f | TTGATGGCAACAATCTCCAC | RTqPCR |
| mGAPDH_r | CGTCCCGTAGACAAAATGGT | RTqPCR |
| eGFP_f | GAGCGCACCATCTTCTTCAAG | qPCR, RTqPCR, ddPCR |
| eGFP_r | TGTCGCCCTCGAACTTCAC | qPCR, RTqPCR, ddPCR |
| eGFP_probe | FAM-ACGACGGCAACTACA-BHQ1 | qPCR, RTqPCR, ddPCR |
| cap6_qPCR_f | ATGCTGAGAACGGGCAATAA | RTqPCR, ddPCR |
| cap6_qPCR_r | TACAGGTACTGGTCGATGAGAG | RTqPCR, ddPCR |
| cap6_qPCR_probe | FAM-TTCAGCTAC/ZEN/ACCTTCGAGGACGTG-IABKFQ | RTqPCR, ddPCR |
| HPRT_f | GAGGATTTGGAAAGGGTGTTTATTC | RTqPCR |
| HPRT_r | CTCCCATCTCCTTCATCACATCTC | RTqPCR |
| HPRT_probe | HEX-ACAGGACTGAACGTCTTGC-BHQ1 | RTqPCR |
| cap2_f | TGACATTCGGGACCAGTCTA | RTqPCR, ddPCR |
| cap2_r | TAGCTCCAGTCCACGAGTATT | RTqPCR, ddPCR |
| cap2_probe | FAM-TTACCGCCA/ZEN/GCAGCGAGTATCAAA-IABKFQ | RTqPCR, ddPCR |
| rep2_f | AAGTCCTCGGCCCAGATAGAC | ddPCR |
| rep2_r | CAATCACGGCGCACATGT | ddPCR |
| rep2_probe | FAM-TGATCGTCACCTCCAACA-BHQ1 | ddPCR |
| h_rpp30_f | AATTAGCCTCTCAAGATGACCAG | ddPCR |
| h_rpp30_r | AGATTTCAACCCAGAGTCACC | ddPCR |
| h_rpp30_probe | HEX-AGCTGGCACTTTCTGTTATGTTTGTGT-BHQ1 | ddPCR |
| hluc_f | CGCCCGCGACCCTATTTTCG | qPCR |
| hluc_r | CAGGTAGCCCAGGGTGGTGAAC | qPCR |
| hluc_probe | FAM-AACCAGATCATCCCCGACACCGCTATTCTGAGCGT-BHQ1 | qPCR |
| U6_f | AATGCTTTCGCGTCGCGCAG | qPCR |
| U6_r | TTGCCTGCGCGTCTTTCCAC | qPCR |
| U6_probe | FAM-TGAGTAAGAGCCCGCGTCTGAACCCTCC-BHQ1 | qPCR |
| ITR_f | GGAACCCCTAGTGATGGAGTT | ddPCR |
| ITR_r | CGGCCTCAGTGAGCGA | ddPCR |
| ITR_probe | HEX-CACTCCCTCTCTGCGCGCTCG-BHQ1 | ddPCR |
| Rx_pre-gRNA-BsmBI_fw | CACCGCAAGTAAACCCCTACCAACTGGTCGGGGTTTGAAACGGAGACGGACGTCTCTCAAGTAAACCCCTACCAACTGGTCGGGGTTTGAAACT | CasRx; cloning of gRNA acceptor site |
| Rx_pre-gRNA-BsmBI_rev | AAAAAGTTTCAAACCCCGACCAGTTGGTAGGGGTTTACTTGAGAGACGTCCGTCTCCGTTTCAAACCCCGACCAGTTGGTAGGGGTTTACTTGC | CasRx; cloning of gRNA acceptor site |
| BamHI_PacI_U6_for | CAGAGGGATCCCAGTTTAATTAACGAGTCCAACACCCGTGG | CasRx; cloning of gRNA acceptor site |
| HindIII_RxAcc_rev | TGATCAAGCTTCCGCAAAAAAAGTTTCAAACC | CasRx; cloning of gRNA acceptor site |
| BsmBI_CasRx_A_fw | AGGATCGTCTCCGGCCGCTGACTGGCGCGCCACCATGAGCCCCAAGAAG | CasRx; amplification of CasRx |
| CasRx_A_rev | TGCGTCGTCTCCCTTGCCATCCAGGAACATGGTCAGG | CasRx; amplification of CasRx |
| CasRx_B_fw | ACGATCGTCTCGCAAGGAGATCAACGACCTCC | CasRx; amplification of CasRx |
| ClaI_HA_CasRx_rev | GATCTATCGATTTAAGCGTAATCTGGAACATCGTA | CasRx; amplification of CasRx |
| CMV_SpeI_fw | GAGCAACTAGTCGTTACATAACTTACGGTAAATG | CasRx; amplification of CMV |
| CMV_AscI_rev | CATGGTGGCGCGCCTCTGACGGTTCACTAAACG | CasRx; amplification of CMV |
| minipA_SalI_for | TCGACTAGCAATAAAGGATCGTTTATTTTCATTGGAAGCGTGTGTTGGTTTTTTGATCAGGCGCGTTAAT | CasRx; mini-pA |

| Name | Sequence (5'->3') | Application |
|---|---|---|
| minipA_PacI_rev | TAACGCGCCTGATCAAAAAACCAACACACGCTTCCAATGAAAATAAACGATCCTTTATTGCTAG | CasRx; mini-pA |
| miR122_2xBS_fw | CGATACAAACACCATTGTCACACTCCAGCTAGGACAAACACCATTGTCACACTCCAG | CasRx; miR122-OFF |
| miR122_2xBS_rev | TCGACTGGAGTGTGACAATGGTGTTTGTCCTAGCTGGAGTGTGACAATGGTGTTTGTAT | CasRx; miR122-OFF |
| mCD44v6_XhoI_for | ATGACCTCGAGCTCCTAATAGTACAGCAGAAGC | psiCheck-2 reporter with mCD44v6 exon |
| mCD44v6_NotI_rev | TGCTAGCGGCCGCCAGTTGTCCCTTCTGTCAC | psiCheck-2 reporter with mCD44v6 exon |
| GFAP_fw | AGAGCAACTAGTCCCACCTCCCTCTCTGTGCTG | GFAP: cloning of deletion mutants |
| GFAPdel1_fw | AGAGCAACTAGTAACATATCCTGGTGTGGAGTAGG | GFAP: cloning of deletion mutants |
| GFAPdel2_fw | GACACAAATGGGTGAGGGGACTCTGGGAGAGAGGCACAGG | GFAP: cloning of deletion mutants |
| GFAPdel3_fw | GTGGAGCTGTCAAGGCCTGGTCTTGAGGGTACAGAACAGG | GFAP: cloning of deletion mutants |
| GFAPdel4_fw | CCTAGTAGGAAATGAGGTGGCCACAAGCATGAGCCACCCCAC | GFAP: cloning of deletion mutants |
| GFAPdel5_fw | CCTCCCAAAGTGCTGGGATTGAGCTCTCCCCATAGCTGGG | GFAP: cloning of deletion mutants |
| GFAP_rev | ACCGGTGCGGCCGCCGAGCAGCGGAGGTGAT | GFAP: cloning of deletion mutants |
| GFAPdel2_rev | CCTGTGCCTCTCTCCCAGAGTCCCCTCACCCATTTGTGTC | GFAP: cloning of deletion mutants |
| GFAPdel3_rev | CCTGTTCTGTACCCTCAAGACCAGGCCTTGACAGCTCCAC | GFAP: cloning of deletion mutants |
| GFAPdel4_rev | GTGGGGTGGCTCATGCTTGTGGCCACCTCATTTCCTACTAGG | GFAP: cloning of deletion mutants |
| GFAPdel5_rev | CCCAGCTATGGGGAGAGCTCAATCCCAGCACTTTGGGAGG | GFAP: cloning of deletion mutants |
| RS_BglII_pA | ACTAGAGATCTCTCCCCAGCATGCCTGCTATTG | RNA selection cloning |
| RS_BamHI_RSV | AGCATGGATCCCAATTCTCATGTTTGACAGC | RNA selection cloning |
| RS_BamHI_EFS | AGCATGGATCCCTGGCTCCGGTGCCCGTCAGTG | RNA selection cloning |
| delRSV_FW | GGCCGCCGACGTATGCATAGCGGAG | RNA selection cloning |
| delRSV_RV | GATCCTCCGCTATGCATACGTCGGC | RNA selection cloning |
| SV40pA_SpeI_for | GCAGCACTAGTCTAGCTTTATTTGTGAAATTTGTG | RNA selection cloning |
| SV40pA_BstBI_rev | AGCCTTTCGAATGTCCGCCGCTTTAAAAAACCTCCCACATCTCCCC | RNA selection cloning |
| ClaI_minipA_fw | CGATAGGCTAGCAATAAAGGATCGTTTATTTTCATTGGAAGCGTGTGTTGGTTTTTA | RNA selection cloning |
| BglII_minipA_rev | GATCTAAAAACCAACACACGCTTCCAATGAAAATAAACGATCCTTTATTGCTAGCCTAT | RNA selection cloning |
| minipA_SpeI_for | CTAGTTAGCAATAAAGGATCGTTTATTTTCATTGGAAGCGTGTGTTGGTTTTTTGATCAGGCGCGCGGACATT | RNA selection cloning |
| minipA_BstBI_rev | CGAATGTCCGCGCGCCTGATCAAAAAACCAACACACGCTTCCAATGAAAATAAACGATCCTTTATTGCTAA | RNA selection cloning |
| rep2i_for | ACTCCAGATCTGCTAGTGAGACGTAAGGCGGCCGCTGCGCAGCCATCGACGTC | RNA selection cloning |
| rep2i_rev | ACCTGTTAATTAACATTTATTGTTCAAAGATGC | RNA selection cloning |
| cap-rplc_for | TAAGATCTCACTCCAGGG | RNA selection cloning |
| cap-rplc_rev | CGCGCCCTGGAGTGAGATCTTAAT | RNA selection cloning |
| PacI_lacZ_2.2_fw | TACATTTAATTAATCCATTTCGCTGGTGGTCAG | RNA selection cloning |
| AscI_lacZ_2.2_rev | CCGGATGGCGCGCCCAATTGCAACGTCGTGACTG | RNA selection cloning |

| Name | Sequence (5'->3') | Application |
|---|---|---|
| LacZ-for | CCGGATGGATCCCAATTGCAACGTCGTGACTG | RNA selection cloning |
| NotI_lacZ_0.6 | TACATGCGGCCGCACGAGACGTCACGGAAAATGC | RNA selection cloning |
| BglII_lacZ_1.6 | TACATAGATCTCCCATCGCGTGGGCGTATTC | RNA selection cloning |
| P7P_oligo | CAGTCGGCCAGAGAGGCCCANNNNNNNNNNNNNNNNNNNNCCAGCCCAGGCGGCTGACGAG | AAV6 peptide library cloning |
| pept_amplif_rev | CTCGTCAGCCGCCTGG | AAV6 peptide library cloning |
| pept_amplif_for | CAGTCGGCCAGAGAGG | AAV6 peptide library cloning |
| cH001_f | AGGCCCATGCGAAGGTGATGATGTTTGCCCAGCCCAGG | AAV6 peptide variant; DEPOOL testing target plasmid |
| cH001_r | GGGCTGGGCAAACATCATCACCTTCGCATGGGCCTCTC | AAV6 peptide variant; DEPOOL testing target plasmid |
| HPRT1_f | CCTGGCGTCGTGATTAGTGAT | qPCR, SARS-CoV-2 project |
| HPRT1_r | AGACGTTCAGTCCTGTCCATAA | qPCR, SARS-CoV-2 project |
| SARS-CoV-2_f | GCCTCTTCTGTTCCTCATCAC | qPCR, SARS-CoV-2 project |
| SARS-CoV-2_r | AGACAGCATCACCGCCATTG | qPCR, SARS-CoV-2 project |
| C8C12_Rseq_fw | GCAAATTCTATGGTGGTTGGCA | SARS-CoV-2 cDNA amplification/sequencing |
| C8C12_Rseq_rev | CCGGCCCCTAGGATTCTTGA | SARS-CoV-2 cDNA amplification/sequencing |
| C3_Nseq_fw | AGAATGGAGAACGCAGTGGG | SARS-CoV-2 cDNA amplification/sequencing |
| C3_Nseq_rev | GCTTCTGGCCCAGTTCCTAG | SARS-CoV-2 cDNA amplification/sequencing |
| BC_AEZ_f | ACACTCTTTCCCTACACGACGCTCTTCCGATCTAGTCCGCCCTGAGCAAAGAC | NGS, YFP/BC region (AmpliconEZ) |
| BC_AEZ_r | GACTGGAGTTCAGACGTGTGCTCTTCCGATCTGGCTGGCAACTAGAAGGCAC | NGS, YFP/BC region (AmpliconEZ) |
| AAV6_AEZ_f | ACACTCTTTCCCTACACGACGCTCTTCCGATCTGCCGGAGCTTCAAACACTGC | NGS, AAV6 peptide region (AmpliconEZ) |
| AAV6_AEZ_r | GACTGGAGTTCAGACGTGTGCTCTTCCGATCTACGGGTGAAAGTGTCCATCCG | NGS, AAV6 peptide region (AmpliconEZ) |
| BC_NGS_f | ATCACTCTCGGCATGGACGAGC | NGS, YFP/BC region (NuGen) |
| BC_NGS_r | GGCTGGCAACTAGAAGGCACA | NGS, YFP/BC region (NuGen) |
| BC_NGS2_f | GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACCACTACCAGCAGAACAC | NGS, YFP/BC region (Nextera) |
| BC_NGS2_r | TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG(N)1-7CAACTAGAAGGCACAGTCG | NGS, YFP/BC region (Nextera) |
| s_Cap6_1703 | GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGGACGAAGAGGAAATCAAAGC | NGS, AAV6 peptide region (Nextera) |
| a_Cap6_1802 | TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG(N)1-7GCTCCCATAACATGCACATC | NGS, AAV6 peptide region (Nextera) |
| D2_T7_cassette_fw | CGCGCCTGAGGTCCGACGAAGACCGTGAGACGACAGTACGTCTCACCTATAGTGAGTCGTATTAGAATGATCGCATCCTTGATACGGCTCCGGGACAA | DEPOOL2 acceptor cloning |
| D2_T7_cassette_rev | CTAGTTGTCCCGGAGCCGTATCAAGGATGCGATCATTCTAATACGACTCACTATAGGTGAGACGTACTGTCGTCTCACGGTCTTCGTCGGACCTCAGG | DEPOOL2 acceptor cloning |
| D2_BC1_fw | TAGGACCGCGGGCGGCTGACGAG | DEPOOL2 BC cloning |
| D2_BC1_rev | ACCGCTCGTCAGCCGCCCGCGGT | DEPOOL2 BC cloning |
| D2_BC2_fw | TAGGGGACTCACACCGAGAAACT | DEPOOL2 BC cloning |
| D2_BC2_rev | ACCGAGTTTCTCGGTGTGAGTCC | DEPOOL2 BC cloning |

| Name | Sequence (5'->3') | Application |
|---|---|---|
| D2_BC3_fw | TAGGAGACAGTTGCGCGCTGGAA | DEPOOL2 BC cloning |
| D2_BC3_rev | ACCGTTCCAGCGCGCAACTGTCT | DEPOOL2 BC cloning |
| D2_BC_BbsI_fw | TGAGGTCCGACGAAGACCG | DEPOOL2 BC amplification |

### 2.1.11 Plasmids

**Table 12: Plasmids used and cloned for this study.**

| Plasmid ID | Name | Origin |
|---|---|---|
| #0067 | pBS-U6 | Dirk Grimm |
| #0102 | psiCheck-2 | Promega (Madison, USA) |
| #0139 | shRen2 | Nina Schürmann |
| #0140 | shRen3 | Nina Schürmann |
| #0141 | shRen4 | Nina Schürmann |
| #0193 | WH-Rep2-CapDJ | Dirk Grimm/ Eike Kienle |
| #0200 | WH-Rep2-Cap8IS | Eike Kienle |
| #0401 | pBs-sds-RSV-hcr (miR122) | This thesis |
| #0402 | pBs-sds-RSV-bic (miR155) | This thesis |
| #0545 | scAAV_GFP | Stefan Mockenhaupt |
| #0552 | pBSUF3rev-YFP-sds | Eike Kienle |
| #0714 | pSSV9_pSi | Dominik Niopek |
| #0778 | pSSV9_Pac_Asc | Stefanie Große |
| #0833 | WHC rh10 p 1 | Eike Kienle |
| #1514 | pBS-H1_F+E_RSV:GFP | Florian Schmidt |
| #1539 | pAAV-LK03 | Marc Kay |
| #1576 | pBSU6(long)ccdB_shRNA | Florian Schmidt |
| #1588 | #48_GGC_1+2_pBSU6(long)ccdB_shRNA | Florian Schmidt |
| #1591 | #48_GGC_2+3_pBSH1_ccdB_shRNA | Florian Schmidt |
| #1594 | #48_GGC_3+4_pBS7SK_ccdB_shRNA | Florian Schmidt |
| #1600 | AAV TRISPR 2.0 ccdB GGC 1+4_YFP Assembly Vector | Florian Schmidt |
| #1614 | WHc6'new' insertion site | Eike Kienle |
| #1736 | WH-Rep2-Cap2WT | Jonas Weinmann |
| #1750 | WH-Rep2-Cap4WT | Jonas Weinmann |
| #1771 | WH-Rep2-Cap7WT | Jonas Weinmann |
| #1778 | WH-Rep2-Cap8WT | Jonas Weinmann |
| #1785 | WH-Rep2-Cap9WT | Jonas Weinmann |
| #1831 | pBS_sds_pGL3 | Dirk Grimm |
| #1935 | pAAV-Rep2-cap2_L1 | Thorsten Lamla |
| #1937 | pAAV-Rep2-cap6.2 | Thorsten Lamla |
| #2056 ff. | pJWx-CMV-EYFP-BC#x | Jonas Weinmann |
| #2509 | pPL03--CMV Jonas_barcode15_bGH_LacZ 2.0 | Claire Domenger |
| #2510 | pPL04--CMV 1_barcode16_bGH_LacZ 2.0 | Claire Domenger |
| #2512 | pPL06--LP1_barcode1_bGH_LacZ 2.0 | Claire Domenger |
| #2517 | pPL11--CAGlab_barcode4_bGH_LacZ 0.8 | Claire Domenger |
| #2518 | pPL12--cd11b_barcode46_bGH_LacZ 0.8 | Claire Domenger |
| #2531 | pPL26--EFS_barcode9_bGH_LacZ 2.2 | Claire Domenger |

| Plasmid ID | Name | Origin |
|---|---|---|
| #2536 | pPL31--Spc5-12_barcode3_bGH_LacZ 2.2 | Claire Domenger |
| #2543 | pPL38--RSV_barcode36_bGH_LacZ 1.8 | Claire Domenger |
| #2553 | pPL48--gfaABC1D_barcode86_bGH_LacZ 1.8 | Claire Domenger |
| #2554 | pPL49--GFAP_barcode74_bGH_noLacZ | Claire Domenger |
| #2564 | pPL59--DeadLaZ_barcode76_bGH_noLacZ | Claire Domenger |
| #3033 | sh_CoV_C16 | Ali Ghanem |
| #3034 | sh_CoV_C17 | Ali Ghanem |
| #3037 | sh_CoV_C20 | Ali Ghanem |
| #3038 | sh_CoV_C21 | Ali Ghanem |
| #3039 | sh_CoV_C22 | Ali Ghanem |
| #jb01 (#3066) | pSSV9_EFS-CasRx | This thesis |
| #jb02 (#3120) | pSSV9_CMV-CasRx | This thesis |
| #jb03 (#3121) | pSSV9_CMV-CasRx_miR122-OFF | This thesis |
| #jb04 | EFS-CasRx_gxCtrl | This thesis |
| #jb05 | EFS-CasRx_Ren1 | This thesis |
| #jb06 | EFS-CasRx_Ren2 | This thesis |
| #jb07 | EFS-CasRx_Ren3 | This thesis |
| #jb08 | EFS-CasRx_Ren4 | This thesis |
| #jb09 | EFS-CasRx_Ren5 | This thesis |
| #jb10 | CMV-CasRx_gxCtrl (C) | This thesis |
| #jb11 | CMV-CasRx_Ren1 (X) | This thesis |
| #jb12 | CMV-CasRx_Ren2 | This thesis |
| #jb13 | CMV-CasRx_Ren3 | This thesis |
| #jb14 | CMV-CasRx_Ren4 | This thesis |
| #jb15 | CMV-CasRx_Ren5 | This thesis |
| #jb16 | CMV-CasRx_X-C | This thesis |
| #jb17 | CMV-CasRx_C-X | This thesis |
| #jb18 | CMV-CasRx_X-C-C | This thesis |
| #jb19 | CMV-CasRx_C-X-C | This thesis |
| #jb20 | CMV-CasRx_C-C-X | This thesis |
| #jb21 | CMV-CasRx_X-C-C-C | This thesis |
| #jb22 | CMV-CasRx_C-X-C-C | This thesis |
| #jb23 | CMV-CasRx_C-C-C-X | This thesis |
| #jb24 | CMV-CasRx_miR-OFF_C | This thesis |
| #jb25 | CMV-CasRx_miR-OFF_X | This thesis |
| #jb26 | CMV-CasRx_gx_v6_1 (vs. CD44v6) | This thesis |
| #jb27 | CMV-CasRx_gx_v6_2 (vs. CD44v6) | This thesis |
| #jb28 | CMV-CasRx_gx_v6_1+2 (vs. CD44v6) | This thesis |
| #jb29 | CMV-CasRx_gxCD44s1 (gx1) | This thesis |
| #jb30 | CMV-CasRx_gxCD44s2 (gx2) | This thesis |
| #jb31 | CMV-CasRx_gxCD44s3 (gx3) | This thesis |
| #jb32 | CMV-CasRx_gxCD44s4 (gx4) | This thesis |
| #jb33 | CMV-CasRx_gxCD44s5 (gx5) | This thesis |
| #jb34 | CMV-CasRx_gxCD44s6 (gx6) | This thesis |
| #jb35 | CMV-CasRx_gxCD44s7 (gx7) | This thesis |
| #jb36 | CMV-CasRx_gxCD44s1+2+7 (gx1-2-7) | This thesis |
| #jb37 | shCtrl | This thesis |
| #jb38 | sh_v6_1 | This thesis |

| Plasmid ID | Name | Origin |
|---|---|---|
| #jb39 | sh_v6_2 | This thesis |
| #jb40 | shmCD44s1 | This thesis |
| #jb41 | shmCD44s2 | This thesis |
| #jb42 | shmCD44s3 | This thesis |
| #jb43 | TRISPR_mCD44_A (3-2-1) | This thesis |
| #jb44 | TRISPR_mCD44_B (2-3-1) | This thesis |
| #jb45 (#3020) | sh_CoV_C3 | This thesis |
| #jb46 (#3022) | sh_CoV_C5 | This thesis |
| #jb47 (#3025) | sh_CoV_C8 | This thesis |
| #jb48 (#3026) | sh_CoV_C9 | This thesis |
| #jb49 (#3029) | sh_CoV_C12 | This thesis |
| #jb50 (#3040) | TRISPR A_C8-C12-C3 | This thesis |
| #jb51 (#3041) | TRISPR B_C3-C12-C8 | This thesis |
| #jb52 (#3042) | TRISPR C_ctrl-ctr-ctrl | This thesis |
| #jb53 (#3043) | TRISPR D_C3-ctr-ctrl | This thesis |
| #jb54 (#3044) | TRISPR E_C8-ctr-ctrl | This thesis |
| #jb55 (#3045) | TRISPR F_ctr-ctrl-C3 | This thesis |
| #jb56 (#3046) | TRISPR G_ctr-ctrl-C8 | This thesis |
| #jb57 (#3047) | TRISPR H_ctr-C12-ctrl | This thesis |
| #jb58 (#3048) | TRISPR I_C8-C12-ctrl | This thesis |
| #jb59 (#3049) | TRISPR J_C8-ctrl-C3 | This thesis |
| #jb60 (#3050) | TRISPR K_ctrl-C12-C3 | This thesis |
| #jb61 (#3051) | TRISPR L_C3-C12-ctrl | This thesis |
| #jb62 (#3052) | TRISPR M_C3-ctrl-C8 | This thesis |
| #jb63 (#3053) | TRISPR N_ctrl-C12-C8 | This thesis |
| #jb64 | psi_mCD44v6 | This thesis |
| #jb65 | psi_BS_8-12-3 | This thesis |
| #jb66 | psi_BS_8*-12*-3* | This thesis |
| #jb67 | psi_BS_8*-12-3 | This thesis |
| #jb68 | psi_BS_8-12*-3 | This thesis |
| #jb69 | psi_BS_8-12-3* | This thesis |
| #jb70 | psi_BS_8*-12*-3 | This thesis |
| #jb71 | psi_BS_8*-12-3* | This thesis |
| #jb72 | psi_BS_8-12*-3* | This thesis |
| #jb73 (#3157) | GFAPdel1_BCA203_lacZ0.8 | This thesis |
| #jb74 (#3158) | GFAPdel2_BCA205_lacZ0.8 | This thesis |
| #jb75 (#3159) | GFAPdel3_BCA207_lacZ0.8 | This thesis |
| #jb76 (#3160) | GFAPdel4_BCA208_lacZ0.8 | This thesis |
| #jb77 (#3161) | GFAPdel5_BCA209_lacZ0.8 | This thesis |
| #jb78 (#3098) | BYRPc2 | This thesis |
| #jb79 (#3093) | rep-c2 | This thesis |
| #jb80 (#3094) | BYRCc2 | This thesis |
| #jb81 (#3099) | RS_BYECc2 | This thesis |
| #jb82 (#3101) | RS_BYxCc2 | This thesis |
| #jb83 (#3104) | BYxCc2S | This thesis |
| #jb84 (#3107) | MYECc2S | This thesis |
| #jb85 (#3096) | BYRCPc2 | This thesis |
| #jb86 (#3119) | BYxCPc2 | This thesis |

| Plasmid ID | Name | Origin |
|---|---|---|
| #jb87 (#3110) | PCswitch_mini-pA | This thesis |
| #jb88 (#3112) | CAGxM | This thesis |
| #jb89 (#3113) | CAGc2M | This thesis |
| #jb90 (#3154) | rep-z | This thesis |
| #jb91 (#3155) | BYzPc2 | This thesis |
| #jb92 (#3156) | zPc2 | This thesis |
| #jb93 (#3108) | rep-c6_nis | This thesis |
| #jb94 (#3109) | MYECc6nisS | This thesis |
| #jb95 | CMV_cap6-cH001 | This thesis |
| #jb96 | D2_BC_acc | This thesis |
| #jb97 | rep-cap2-BC1 | This thesis |
| #jb98 | rep-cap2-BC2 | This thesis |
| #jb99 | rep-cap2-BC3 | This thesis |
| Addgene #109049 | pXR001: EF1a-CasRx-2A-EGFP | [210] |

## 2.2 Methods

### 2.2.1 Molecular biology methods

#### *2.2.1.1 Liquid bacterial cultures and plasmid preparation*

For plasmid preparations, I incubated bacteria in LB medium supplemented with either ampicillin (50 µg/mL) or chloramphenicol (30 µg/mL) according to the plasmid's resistance gene. Incubation was performed at 37 °C for 16 h while shaking at 180 rpm.

Depending on the required plasmid quantity, bacteria were grown in three different quantities and purified using commercially available kits following the manufacturer's instructions: (i) for small-scale plasmid preparation (Mini-Preps), bacteria were grown in a liquid volume of 2 mL of LB medium and plasmids were isolated using the QIAprep Spin Miniprep Kit (Qiagen); (ii) for mid-scale plasmid preparation (Midi-Preps) a liquid culture of 80 mL was prepared, and isolation was conducted using the PureYield Plasmid Midiprep System (Promega); for (iii) large-scale plasmid preparations (Maxi-Preps), a culture volume of 400 mL was prepared and plasmid isolation was performed with the NucleoBond PC 500 kit (Macherey-Nagel). DNA concentrations were measured using the NanoDrop 2000 Spectrophotometer (Thermo Fisher Scientific).

For small-scale plasmid preparations that were not intended for sensitive applications (such as Lipofectamine transfection), Mini-Prep isolations were conducted using self-made buffers. In this case, bacteria were pelleted from liquid cultures by centrifugation at 5,000 × g for 3 min, followed by resuspension of the bacterial pellet in 300 µL of P1 resuspension buffer. After incubation for 5 min at room temperature (rt), 300 µL of P2 lysis buffer were added. The suspension was mixed by gently inverting the tube and

incubated for 5 min at room temperature (rt). Next, 300 µL of P3 neutralization buffer were added, followed by gentle mixing, incubation for 5 min and subsequent centrifugation at 16,000 × g. After centrifugation, 800 µL of supernatant were collected and mixed to 600 µL isopropanol, followed by centrifugation at 16,000 × g for 10 min. Subsequently, supernatant was discarded, and pelleted DNA was washed by adding 500 µL of 70% ethanol and centrifuging at 16,000 × g for 5 min. Finally, after discarding the supernatant, the DNA pellet was dried at rt for 10 min and resuspended in 50 µL nuclease-free $H_2O$.

#### *2.2.1.2 Agarose gel electrophoresis*

Agarose gel electrophoresis was performed for standard analysis of DNA fragments or for size-based preparative separation thereof. Depending on DNA fragment size, agarose (Biozym) was dissolved in 1x TAE buffer with 0.5 - 2% w/v. Prior to gel casting, ethidium bromide was added to a final concentration of 1 µg per mL gel. Samples were mixed with 6x Purple Loading Dye (NEB) prior to loading. For size comparison of DNA fragments, the 1 kb Plus DNA ladder was loaded next to the sample wells (Thermo Fisher Scientific). Gels were run at 120 V for 30 min and imaged with the GelDoc XR system (Bio-Rad).
Purification of DNA fragments from agarose gels was performed using the QIAquick Gel Extraction Kit (Qiagen) following the manufacturer's instructions.

#### *2.2.1.3 Polymerase chain reaction (PCR)*

##### 2.2.1.3.1 Standard PCR amplification

Standard PCR amplification was conducted using either the Q5 High-Fidelity 2X Master Mix (New England Biolabs; NEB), the Phusion Flash High-Fidelity PCR Master Mix (Thermo Fisher Scientific), or the Phusion HS II polymerase with 1× HF buffer (Thermo Fisher Scientific) and 200 µM of deoxynucleotides (dNTPs). For both systems, 500 nM of forward and reverse primers were used. Templates were typically 1 ng of plasmid DNA, $10^7$ vector genomes (vg) of AAV samples, or up to 150 ng of genomic DNA. Reaction volumes were adjusted to 50 µL with nuclease-free $H_2O$. Each PCR reaction was also prepared for a no template control (NTC) to ensure absence of contamination. Cycling conditions are displayed in Table 13. Annealing temperatures were adapted to the respective primer sets if necessary, and elongation times were adjusted to template size (30 s/kb).

**Table 13: PCR cycling conditions.**

| Step | Temperature [°C] | Time [min:sec] | Cycles |
|---|---|---|---|
| Initial denaturation | 98 | 5:00 | |
| Denaturation | 98 | 0:15 | |
| Annealing | 60 | 0:30 | 35 |
| Elongation | 72 | 30 s/kb | |
| Final elongation | 72 | 5:00 | |
| Hold | 4 | ∞ | |

PCR products were analyzed *via* agarose gel electrophoresis using 2 µL of PCR reaction mixed with 2 µL of 6x loading dye (NEB) and 10 µL nuclease-free $H_2O$. Purification of amplicons was conducted either *via* agarose gel electrophoresis and subsequent gel extraction (in cases where multiple amplification products were visible), or *via* column purification using either the QIAquick PCR Purification Kit (Qiagen) for amplicons larger than 100 bp or the QIAquick Nucleotide Removal Kit (Qiagen) for amplicons smaller than 100 bp.

#### 2.2.1.3.2 Rescue PCR

PCR-based rescue of *cap* genes or fragments thereof from gDNA or cDNA input were tested and optimized for (i) *cap2* (section 3.2.3.1) and (ii) *cap6* (section 3.2.3.2) using PCR settings as described above. (i) Optimization of rescue PCR conditions from cDNA were conducted for *cap2* cDNA derived from Hek293T cells transduced with MYECc2S. Reverse transcription of RNA was performed with different settings as described in section 2.2.1.13. PCR amplification was conducted in 10 µL reaction volume containing 5 µL Phusion Flash High-Fidelity PCR Master Mix, 2 µL nuclease-free H2O, 2 µL cDNA template (+/-RT; $H_2O$ for NTC) and 0.5 µL of each 10 µM forward (cap2_fw, Lseq_fw or rep2_resc_fw) and reverse (cap2_resc_rev) primers. PCR was performed as described above with 35 amplification cycles, annealing at 60°C, and elongation for 30 sec (cap2_fw) or 60 sec (Lseq_fw or rep2_resc_fw).

(ii) Rescue of *cap6* was tested in 25 µL reaction volume (adjusted with nuclease-free $H_2O$) containing 12.5 µL Phusion Flash High-Fidelity PCR Master Mix, 1.25 µL of each 10 µM forward (cap6_resc_fw, Lseq_fw or rep2_resc_fw) and reverse (cap6_resc_rev) primers, and templates from gDNA or cDNA extracted from mouse hepatocytes or NPCs, using 75 ng gDNA input or cDNA from the equivalent of 150 ng RNA. PCR amplification

was conducted with 40 cycles, using 60 °C annealing temperature and elongation for 30 sec (cap6_resc_fw) or 60 sec (Lseq_fw or rep2_resc_fw).

2.2.1.3.3 Colony PCR

For investigating cloning results, colony PCRs were conducted using single colonies from bacterial transformations as templates. Colony PCR was performed using OneTaq Quick-Load 2X Master Mix (NEB) with 200 nM of forward and reverse primers in a reaction volume of 10 µL (adjusted with nuclease-free $H_2O$). As opposed to the standard cycling conditions displayed in Table 13, the temperature for denaturation steps was here set to 94 °C and elongation was performed at 68 °C. PCR products were again analyzed *via* agarose gel electrophoresis.

#### *2.2.1.4 Restriction digest*

Enzymatic restriction of plasmids or amplicons was performed using restriction endonucleases listed in Table 7. Typically, 1-10 µg of plasmid DNA or PCR amplicons were digested in a reaction volume of 50 µL containing 1-2 µL of the respective restriction enzyme and its corresponding buffer. Incubation was performed at the appropriate temperature for 4-16 h. Digested DNA fragments were purified using agarose gel electrophoresis and subsequent gel extraction as described above.

For ITR-bearing AAV plasmids, control restriction digests were performed for confirming the presence of ITR sequences. This was achieved by restriction with XmaI (NEB) for single-stranded AAV constructs (AAV2 ITR), and additionally with BsaI (NEB) for self-complementary AAV vectors (AAV2 and AAV4 ITRs). Presence of correctly sized DNA fragments after restriction digest were used as indicators for presence of ITRs and correct construct sizes.

#### *2.2.1.5 Dephosphorylation of plasmids*

Prior to ligation of DNA inserts into plasmids, purified pre-digested plasmids were dephosphorylated using the Antarctic Phosphatase (NEB) in a reaction volume of 20 µL, containing 1 µL Antarctic Phosphatase, 1x Antarctic Phosphatase buffer and up to 1 µg of plasmid DNA. Reactions were incubated for 30 min at 37 °C and inactivated at 80 °C for 2 min.

#### *2.2.1.6 Annealing and phosphorylation of DNA oligonucleotides for ligation*

For preparing inserts for ligation based on synthetic DNA oligonucleotides, 2.5 µL of each matching forward and reverse oligonucleotide stocks (100 µM) were mixed to 5 µL of NEBuffer 2.1 (NEB) and 40 µL of nuclease-free H2O. Annealing was performed in a thermal cycler by heating to 98 °C for 7 min, followed by cooling to 25 °C with a 5% ramp. Annealed oligonucleotides (now at 5 µM) were then phosphorylated using the T4 polynucleotide kinase (PNK) where 4 µL of annealed oligonucleotides were mixed to 2 µL of T4 PNK buffer, 1 µL PNK, and 14 µL nuclease-free $H_2O$. The reaction was incubated at 37 °C for 30 min and inactivated at 65 °C for 20 min. Annealed and phosphorylated oligonucleotides (now at 1 µM) were then diluted in $H_2O$ prior to DNA ligation.

#### *2.2.1.7 DNA ligation*

DNA inserts (annealed and phosphorylated oligonucleotides, pre-digested amplicons or fragments excised from other plasmids) were ligated into plasmid backbones typically in a ratio of 3:1 (insert : backbone) by mixing 100 ng of plasmid backbone to the appropriate amount of insert, 2 µL T4 ligase buffer and 1 µL of T4 DNA ligase (NEB). Reaction volumes were adjusted to 20 µL with nuclease-free $H_2O$. Reactions were incubated at rt for 30 min or at 16 °C for 16 h. Ligation controls were performed in the absence of insert DNA to assess backbone re-ligation.

#### *2.2.1.8 Golden Gate Assembly*

Golden gate assembly [271] was performed for cloning of annealed oligonucleotides into plasmid acceptors containing two inverted binding sites for Type IIS restriction enzymes (*e.g.*, BbsI, BsmBI). This was achieved by mixing 50 ng acceptor plasmid with a 10-fold excess of annealed Oligo, 1 µL T4 ligase buffer, 0.5 µL T4 ligase (NEB) and 0.5 µL of the respective restriction enzyme (BbsI-HF or BsmBI-v2; NEB) in a total volume of 10 µL. The mixture was incubated for 25 cycles between enzymatic restriction at 37 °C (BbsI-HF) or 55 °C (BsmBI-v2) for 2 min and ligation at 16 °C for 3 min, followed by inactivation at 80 °C for 10 min. Ligation control was performed with absence of oligonucleotide insert.

#### *2.2.1.9 Bacterial transformation – chemically*

Transformation of chemically competent MAX Efficiency DH5α bacteria or One Shot ccdB Survival 2 T1 Competent Cells (Thermo Fisher Scientific) was conducted by adding 1 µL of dissolved plasmid (re-transformation) or 5 µL of ligation mix (section 2.2.1.7) to one

aliquot (50 µL) of bacteria. The bacterial suspension was mixed and incubated on ice for 20 min. Heat shock was then performed by heating to 42 °C for 45 sec on a mixing block, followed by incubation on ice for 2 min. For plasmid backbones with ampicillin resistance, the bacterial suspension was then plated onto LB-ampicillin agar plates and incubated at 37 °C overnight. Colonies were quantified and picked for downstream applications (colony PCR and/or liquid cultures).

For plasmids with chloramphenicol resistance, bacteria were recovered prior to plating by mixing them to 1 mL of antibiotic-free LB medium and incubating at 37 °C for 1 h while shaking at 600 rpm. Next, bacteria were pelleted by centrifugation at 1,000 × g for 3 min and resuspended in 100 µL LB medium before plating onto LB-chloramphenicol agar plates.

#### *2.2.1.10 Bacterial transformation – electroporation*

For preparation of plasmid libraries, bacterial transformation was performed *via* electroporation in multiple replicates (4-8) to achieve sufficient transformation efficiency. Prior to electroporation, ligation mixes were first desalted by incubation on MF-Millipore Membrane Filters (Merck) placed in nuclease-free $H_2O$. After 20 min, desalted ligation mixes were collected. Transformations were performed by mixing 3 µL of desalted ligation mix to 25 µL of E. cloni 10G Supreme bacteria (Lucigen). Each bacterial suspension was added to an ice-cooled electroporation cuvette and pulsed at 1,800 V using a Gene Pulser Xcell. Directly after pulsing, bacteria were mixed with Lucigen recovery medium and incubated at 37 °C for 60 min while shaking at 600 rpm. After recovery, replicate reactions were pooled and a serial dilution was prepared by mixing 10 µL of suspension to 90 µL of medium (1:10 to $1:10^4$). Each dilution was plated onto LB-ampicillin agar plates and incubated at 37 °C for 16 h. The remaining bacterial solution was used for inoculation of Midi- or MaxiPrep liquid cultures.

Library diversity was calculated as follows: (i) bacterial colonies were counted for the different dilution steps. (ii) The colony number (*e.g.*, 100 colonies) was multiplied with the dilution factor of the respective dilution step (*e.g.*, counted for the $1:10^3$ dilution step), multiplying by 10 (to calculate per 1 mL bacterial suspension) and multiplying by the total volume of bacterial suspension (*e.g.*, for 5 electroporations: 5 mL of bacterial suspension present after recovery and mixing). In this example, the library would be present with a theoretical diversity of $100\times10^3\times10\times5 = 5\times10^6$.

#### *2.2.1.11 Sanger sequencing*

Plasmids or PCR amplicons were submitted for sequencing with the Eurofins Genomics Sanger sequencing service. For each sequencing run, a 10 µL reaction was prepared containing 2.5 µM of an appropriate sequencing primer and 100-500 ng DNA template (volume adjusted with nuclease-free $H_2O$). Sequencing results were analyzed using the ApE software [268].

#### *2.2.1.12 Extraction of DNA and/or RNA from cells and tissues*

Extractions of DNA and/or RNA from cells or tissues were performed using column-based purification kits from Qiagen following the manufacturer's instructions. Extractions from genomic DNA were performed using the DNeasy Blood & Tissue Kit, while RNA was extracted using the RNeasy Mini Kit. Parallel extraction of DNA and RNA was conducted with the AllPrep Mini Kit for samples with $>10^5$ cells or the AllPrep Micro Kit for samples with $<10^5$ cells. For all extractions of RNA (both RNeasy and AllPrep kits), an on-column digest was performed using the RNase-free DNase set during the step of column-wash with RW1 buffer. All extracts were eluted with nuclease-free $H_2O$. DNA and RNA concentrations were quantified using the NanoDrop 2000 system.

For extractions from murine tissues, small tissue samples (~ 25 mg) were sliced and transferred to a 2 mL tube containing a stainless-steel bead and 180 µL ATL buffer (DNeasy kit) or 600 µL RLTplus buffer (AllPrep kit) buffer. Tissues were dissociated using the TissueLyser LT (Qiagen), followed by addition of 20 µL Proteinase K (Qiagen) and incubation at 56 °C for 30 min. After tissue dissociation, supernatants were subjected to downstream processing using the DNeasy or AllPrep kits.

#### *2.2.1.13 cDNA synthesis*

Complementary (c)DNA was synthesized from RNA using either the (i) High-Capacity cDNA Reverse Transcription kit, (ii) the SuperScript IV VILO Master Mix with ezDNase Enzyme (section 3.2.3) or (iii) the SuperScript IV First-Strand Synthesis System with ezDNase Enzyme (section 3.2.3 samples from *in vivo* study) (Thermo Fisher Scientific) following the manufacturer's instructions.

For the High-Capacity kit (i), 424 ng RNA were diluted with nuclease-free $H_2O$ to a total volume of 35 µL. Remaining genomic (g)DNA was removed by incubation with 1 µL DNase I and 4 µL RDD buffer (from the RNase-free DNase set) for 30 min at rt, followed by inactivation at 75 °C for 10 min. Synthesis of cDNA was then performed by mixing 28.4

µL of DNase-digested RNA (equals 300 ng) to 4 µl 10X RT Buffer, 1,6 µl 25X dNTP Mix, 4 µl 10X RT Random Primers and 2 µl MultiScribe Reverse Transcriptase (components of High-Capacity cDNA Reverse Transcription kit). The mixture was then incubated as described in Table 14.

**Table 14: Conditions for cDNA synthesis with the High-Capacity kit.**

| Step | Temperature [°C] | Time [min] |
|---|---|---|
| Primer annealing | 25 | 10 |
| Incubation | 37 | 120 |
| Inactivation | 85 | 5 |
| Hold | 4 | ∞ |

For the SuperScript kits (ii and iii), remaining gDNA was removed by combining 150 ng RNA diluted in up to 12 µL with 1.5 µL ezDNase and 1.5 µL ezDNase buffer. DNase reactions were incubated at 37 °C for 5 min. From this reaction, 5 µL were used for no-reverse transcriptase (no-RT) control reactions, whereas the other 10 µL were subjected to reverse transcription. For the SuperScript IV VILO kit (ii), cDNA synthesis reactions contained 10 µL ezDNase-digested RNA, 4 µL RT master mix and 6 µL nuclease-free $H_2O$. Incubation was performed as described in Table 15. No-RT control reactions were performed in 50% volume using the supplied no-RT master mix.

For the SuperScript IV First-Strand Synthesis System (iii), cDNA synthesis was performed by firstly combining 1.5 µL RT primer (supplied random hexamers at 50 ng/µL or oligo-dT primers at 50 µM, or 2 µM gene-specific reverse transcription primers cap2_RT, SV40pA_RT1 or SV40pA_RT2) with 1.5 µL dNTP mix, 1.5 µL $H_2O$ and 15 µL ezDNase-treated RNA. For initial primer annealing, this mixture was incubated at 65 °C for 5 min, followed by incubation on ice for 1 min. The RT reaction was performed in a total volume of 20 µL by combining 13 µL of the pre-annealed primer/RNA mix, 4 µL 4x SuperScript buffer, 1 µL 100 mM DTT, 1 µL Ribonuclease inhibitor and 1 µL SuperScript IV. Synthesis of cDNA was then performed by incubation as described in Table 15. The no-RT reaction was performed in 50% volume (*i.e.*, 10 µL) while omitting the addition of reverse transcriptase. After cDNA synthesis, RNA was removed by adding 1 µL RNase H and incubating at 37 °C for 20 min.

**Table 15: Conditions for cDNA synthesis with the SuperScript IV VILO kit.**

| Step | Temperature [°C] | Time [min] |
|---|---|---|
| Primer annealing | 25 | 10 |
| Incubation | 50 | 10 |
| Inactivation | 85 | 5 |
| Hold | 4 | ∞ |

#### *2.2.1.14 Quantification of gene expression by RTqPCR*

Quantification of gene expression from cDNA was performed using reverse-transcription quantitative PCR (RTqPCR). This was performed using either (i) SYBR-green based quantification or (ii) probe-based quantification. For (i) SYBR-green based quantification, each reaction was performed with a total volume of 10 µL containing 5 µL iTaq Universal SYBR Green Supermix, 2 µL nuclease-free $H_2O$, 2 µL of 1:10 diluted cDNA and 0.5 µL of each 10 µM forward and reverse primer (final concentration: 500 nM). Target genes detected with this method were human *(h)ACE2* (using hACE2_f and hACE2_r primers), murine *(m)CD44* (mCD44_f and mCD44_r), *eYFP* (eGFP_f and eGFP_r) and *cap2* (cap2_f and cap2_r), while housekeeping genes were *GAPDH* for either human (hGAPDH_f and hGAPDH_r) or murine cells (mGAPDH_f and mGAPDH_r). RTqPCR reactions were pipetted in 96-well plates (Biozym), sealed with optical adhesive film (Thermo Fisher Scientific) and run with the StepOne Plus system (Applied Biosystems) using the cycling conditions described in Table 16.

The (ii) probe-based quantification was utilized for sections 3.2.1 and 3.2.3.2/3.2.3.3 for quantifying the expression of *eYFP* or *cap6* transgenes and human or mouse housekeeping reference genes. *eYFP* was detected using eGFP_f, eGFP_r and eGFP_probe, while *cap6* was detected using cap6_f, cap6_r and cap6_probe. For housekeeping reference genes, *HPRT* was measured for human cells using HPRT_f, HPRT_r and HPRT_probe, while *PolR2A* was measured for murine cells using the PolR2A primer/probe mix (Mm00839502_m1; Thermo Fisher Scientific). Each reaction was set up containing 12.5 µL Sensimix II probe mix, 0.5 µL ROX (final conc.: 500 nM), 0.1 µL of each forward and reverse primer (100 µM stock; final conc.: 400 nM), 0.025 µL of probe (100 µM stock; final conc.: 100 nM) and 2 µL of 1:10 diluted cDNA, and adjusted to a final volume of 25 µL with nuclease-free $H_2O$. RTqPCR was run with the StepOne Plus system using cycling conditions described in

Analysis of RTqPCR data was performed by extracting the Ct values of target gene ($Ct_{target}$) and housekeeper ($Ct_{ref}$). Relative expression ($2^{-\Delta Ct}$) was calculated as follows:

$$2^{-\Delta Ct} = 2^{-(Ct_{target} - Ct_{ref})}$$

For knockdown analysis, $2^{-\Delta\Delta Ct}$ was calculated by comparing the ΔCt of each condition i (ΔCt(i)) to the average ΔCt(ref) of the reference condition (non-targeting control shRNA or gRNA):

$$2^{-\Delta\Delta Ct} = 2^{-(\Delta Ct(i) - \Delta Ct(ref))}$$

For evaluating promoter strength (3.2.1) or functional transduction (3.2.3) across different tissues with varying vector presence, the relative expression Cβ measured by RTqPCR for cDNA from a sample β was normalized to the vector presence Gβ in the same sample as measured by ddPCR from the gDNA. This value is herein referred to as normalized expression Qβ.

**Table 16: SYBR-green based RTqPCR cycling conditions.**

| Step | Temperature [°C] | Time [min:sec] | |
|---|---|---|---|
| Initial denaturation | 95 | 10:00 | |
| Denaturation | 95 | 00:15 | 40 cycles |
| Annealing/elongation | 60 | 01:00 | |
| Denaturation | 95 | 00:15 | Melt curve (slope: 0.3°C/sec) |
| Anneal | 60 | 01:00 | |
| Melt | 95 | 00:15 | |

**Table 17: Probe-green based RTqPCR cycling conditions.**

| Step | Temperature [°C] | Time [min:sec] | |
|---|---|---|---|
| Initial denaturation | 95 | 10:00 | |
| Denaturation | 95 | 00:20 | 40 cycles |
| Annealing/elongation | 60 | 01:00 | |

#### *2.2.1.15 Quantification of vector presence in gDNA by ddPCR*

For quantification of the number of vector genomes present per nucleus after AAV transductions (vector genome per diploid host genome; vg/dg), droplet digital (dd)PCR was performed for parallel measurement of the vector transgene and a host reference gene. Quantified transgenes were *rep2* (using rep2_f, rep2_r and rep2_probe), *eYFP*

(eGFP_f, eGFP_r and eGFP_probe) or *cap6* (cap6_f, cap6_r and cap6_probe), where each probe was labeled with a FAM fluorophore. As reference host gene, *rpp30* was measured for either human (h_rpp30_f, h_rpp30_r and h_rpp30_probe) or murine cells (using the ddPCR GEX HEX Assay Rpp30, Mmu (#10031255) mix, Bio-Rad) using HEX-labeled probes. Reactions for ddPCR were prepared in 22 µL volume containing 11 µL ddPCR Supermix for Probes (No dUTP; Bio-Rad), 1.1 µL of each 20x transgene and housekeeper mix (final concentration: 900 nM for primers, 250 nM for probes), 1.1 µL of 1:4 diluted HindIII-HF (NEB, diluted in Diluent B), 2.2 µL $H_2O$ and 5.5 µL diluted gDNA (concentration: 5 ng/µL). Reactions were incubated at rt for 15 min for digestion of gDNA with HindIII-HF, after which droplets were generated using Droplet Generation Oil for Probes with the QX200 Droplet Generator (Bio-Rad). Droplets were transferred to 96-well plates, which were sealed with pierceable foil using the PX1 PCR Plate Sealer (Bio-Rad). ddPCR reactions were run with the C1000 Touch Thermal Cycler using cycling conditions described in described in Table 18.

**Table 18: Cycling conditions for ddPCR.**

| Step | Temperature [°C] | Time [min:sec] | |
|---|---|---|---|
| Initial denaturation | 94 | 10:00 | |
| Denaturation | 94 | 00:30 | 40 cycles |
| Annealing/elongation | 58 | 01:00 | |
| Final elongation | 98 | 10:00 | |
| Hold | 12 | ∞ | |

Analysis of droplets was performed using the QX200 Droplet Reader, measuring transgene (FAM-labeled probe) on channel 1 and reference gene (HEX-labeled probe) on channel 2. The number of vector genomes per host diploid genome Gβ [vg/dg] was then calculated from the output of each reaction ($N_{transgene}$ = copy number of transgene templates per reaction, $N_{ref}$ = copy number of reference templates per reaction) as follows:

$$G\beta\left[\frac{vg}{dg}\right] = \frac{N_{\text{transgene}}}{2 \times N_{ref}}$$

### 2.2.2 Specific cloning procedures

#### *2.2.2.1 CasRx constructs and gRNAs*

2.2.2.1.1 EFS-CasRx

For expressing *Ruminococcus flavefaciens* Cas13d (CasRx) and respective gRNAs from an ssAAV construct, the EFS_CasRx vector was cloned in three steps: (i) the gRNA acceptor site was generated and amplified, (ii) the CasRx transgene was amplified, (iii) both were cloned into an ssAAV backbone. For the generation of the gRNA acceptor site (i), the shRNA-acceptor site in pBSU6(long)ccdB_shRNA (#1576) was replaced by a gRNA acceptor site containing two CasRx direct repeats separated by inverted BsmBI sites. This was achieved by digesting #1576 with BsmBI-v2, followed by ligation of annealed oligonucleotides Rx_pre-gRNA-BsmBI_fw and Rx_pre-gRNA-BsmBI_rev. The resulting gRNA acceptor cassette containing U6-promoter, direct repeats separated by inverted BsmBI sites, and a poly-T termination sequence was then PCR-amplified using the BamHI_PacI_U6_for and HindIII_RxAcc_rev primers. Amplicon and #2531 plasmid backbone were digested with BamHI-HF and HindIII-HF (NEB) and ligated, producing an intermediate gRNA-acceptor plasmid. (ii) The CasRx-2xNLS sequence was amplified from the Addgene #109049 plasmid [210]. Since conventional PCR did not produce an amplicon, the amplicon was split into two parts A (5' region) and B (3' region), using the BsmBI_CasRx_A_fw and CasRx_A_rev primers for part A and the CasRx_B_fw and ClaI_HA_CasRx_rev primers for part B. (iii) Restriction digest was performed with NotI-HF and ClaI (NEB) for the intermediate gRNA-acceptor plasmid, with BsmBI-v2 for CasRx amplicon A, and with BsmBI-v2 and ClaI for CasRx amplicon B. Ligation of all three yielded the final EFS-CasRx plasmid (#3066). Correct cloning was confirmed *via* Sanger sequencing.

2.2.2.1.2 CMV-CasRx

The CMV-CasRx plasmid (#3120) was created by amplifying the CMV promoter from #2509 using the CMV_SpeI_fw and CMV_AscI_rev primers, followed by digestion of amplicon and EFS_CasRx plasmid with SpeI and AscI (NEB). Ligation thereof yielded a CMV-CasRx-bgh intermediate. Next, the bgh-polyA from the intermediate was removed by restriction with SalI-HF and PacI (NEB), followed by ligation of a mini-pA signal [161] *via* oligonucleotide annealing of minipA_SalI_for with minipA_PacI_rev.

#### 2.2.2.1.3 CMV-CasRx_miR122-OFF

CMV-CasRx_miR122-OFF (#3121) was cloned by digestion of CMV-CasRx with ClaI and SalI-HF and ligation of an annealed oligonucleotide consisting of miR122_2xBS_fw and miR122_2xBS_rev.

#### 2.2.2.1.4 Design and cloning of CasRx gRNAs

CasRx gRNAs targeting *Renilla* luciferase *Rluc* (psiCheck-2 target; gxRen1-5) and murine *CD44* (gx1-7; NC_000068.8) were designed using the cas13design web tool (cas13design.nygenome.org) [272,273]. The non-targeting control gRNA (gxCtrl) was derived from Konermann *et al*. [210]. Cloning was performed by golden gate assembly of annealed oligonucleotides into the EFS-CasRx, CMV-CasRx or CMV-CasRx_miR122-OFF plasmids. Single gRNAs in EFS-CasRx were cloned jointly with Emma Gerstmann, whereas gRNAs in CMV-CasRx_miR122-OFF were cloned jointly with Jun Kai Ong. Oligonucleotides for single spacers had the following design:

| | | | |
|---|---|---|---|
| Forward: | 5’- AAAC | $N_{23}$ | -3’ |
| Reverse: | 3’- | $N_{23}$ GTTC | -5’ |

Here, $N_{23}$ is the 23-nucleotide long target-specific spacer (antisense in forward). Spacers sequences are listed below in Table 19.

**Table 19: Cas13d spacer sequences (antisense binding sites).**

| gRNA name | Spacer sequence (antisense) |
|---|---|
| gxCtrl (“C”) | TCACCAGAAGCGTACCATACTCA |
| gxRen1 (“X”) | ACGTTCATTTGCTTGCAGCGAGC |
| gxRen2 | ACCATGCAGAAAAATCACGGCGT |
| gxRen3 | TCCTCCTCGATGTCAGGCCACTC |
| gxRen4 | CGAAGAAGTTATTCTCAAGCACC |
| gxRen5 | GATGAACATCTTAGGCAGATCGT |
| gX_mCD44v6_1 | CCATCCGTTCTGAAACCACGTCT |
| gX_mCD44v6_2 | CAGTTGTCCCTTCTGTCACATGG |
| gxCD44s1 | CCACATGGAATACACCTGCGTAG |
| gxCD44s2 | AATAGTTATGGTAACCGGTCCAT |
| gxCD44s3 | CATCTATAATGTTTGAAGCATCG |
| gxCD44s4 | TAATTCGGATCCATGAGTCACAG |
| gxCD44s5 | AGATGCCAAGATGATGAGCCATT |
| gxCD44s6 | GCAAGAATCAGAGCCAGTGCCAG |
| gxCD44s7 | CCCCAATCTTCATGTCCACACTC |

2.2.2.1.5 Multiplexing of CasRx gRNAs

CasRx gRNAs were multiplexed into 2-, 3- or 4-spacer constructs as depicted in Figure 7. Two-fold multiplexing was achieved using a single annealed forward and reverse oligonucleotide, while triplex and quadruplex spacer constructs were cloned with one annealed oligonucleotide per spacer position choosing unique overhangs per position. Individual spacer sequences are the same as above. Cloning for multiplexed constructs was performed with ligation of annealed oligonucleotide mixes (equimolar ratio for each position) into CMV-CasRx pre-digested with BsmBI-v2 instead of using golden gate assembly. Multiplexed gRNA constructs targeting *Rluc* were cloned jointly with Jun Kai Ong.

### *2.2.2.2 shRNA constructs*

2.2.2.2.1 Design of shRNAs

Short hairpin (sh)RNAs were designed using the siRNAWizard tool (invivogen.com/sirna-wizard) targeting SARS-CoV-2 *RdRP* (C8, C9 and C12 designed by Dirk Grimm and cloned by me; NCBI reference: NC_045512.2) and *N* genes (C3, C5 designed by Dirk Grimm and cloned by me; C16 and C17 designed and cloned by Ali Ghanem), human (h)*ACE2* (C20, C21 and C22 designed and cloned by Ali Ghanem; NCBI reference: NM_021804.3), *Renilla* luciferase transgene *Rluc* (shRen2, shRen3 and shRen4 designed and cloned by Nina Schürmann) or the murine (m)*CD44* (sh_v6_1, sh_v6_2 targeting variant exon 6 were designed and cloned by me; sh1, sh2, sh3 targeting standard exons [shmCD44s1-3] were designed by me and cloned jointly with Emma Gerstmann; NCBI reference: NC_000068.8). The sequence for a non-targeting scramble shRNA (shCtrl) was received from Philippe Gual and cloned by me. Individual shRNA binding sequences are listed below in Table 20.

Oligonucleotides for cloning of shRNAs have the following design (overhang – sense ($N_{18-23}$) – loop – antisense ($X_{18-23}$) – overhang):

Forward: 5’- CACC $N_{18-23}$ TCAAGAG $X_{18-23}$ -3’

Reverse: 3’- $N_{18-23}$ AGTTCTC $X_{18-23}$ AAAA -5’

**Table 20: Sense binding sites of utilized shRNAs.**

| shRNA name | Sense binding sequence |
|---|---|
| shCtrl | CCTAAGGTTAAGTCGCCCTCG |
| shRen2 | GCTGGACTCCTTCATCAAC |
| shRen3 | GGCCTTTCACTACTCCTACGA |

| shRNA name | Sense binding sequence |
|---|---|
| shRen4 | GCCTGACATCGAGGAGGATAT |
| C3 | GACAAGGCGTTCCAATTAACA |
| C5 | GCAGACGTGGTCCAGAACAAA |
| C8 | GTGATAGAGCCATGCCTAACA |
| C9 | GCTTGTCACACCGTTTCTATA |
| C12 | GACTGAGACTGACCTTACTAA |
| C16 | GGCCGCAAATTGCACAAT |
| C17 | GTCTGGTAAAGGCCAACAACAA |
| C20 | GAAGTGGAGGTGGATGGTCTT |
| C21 | GGAGATGAAGCGAGAGATAGT |
| C22 | GTACCTGTTCCGATCATCTGT |
| shRNA_v6_1 | GCAGGAGACGTGGTTTCAGAA |
| shRNA_v6_2 | GAAGACTCCCATGTGACAGAA |
| shmCD44s1 | GGCTTTCAACAGTACCTTACC |
| shmCD44s2 | GATTCATCCCAACGCTATCTG |
| shmCD44s3 | GGACCGGTTACCATAACTATT |

#### 2.2.2.2.2 Cloning of single-shRNA scAAV constructs

Single shRNA constructs were cloned into the #1576 backbone (scAAV construct) by golden gate assembly of annealed oligonucleotides as described in 2.2.1.8.

#### 2.2.2.2.3 Cloning of triple-shRNA (TRISPR) scAAV constructs

Triple shRNA constructs were cloned following the TRISPR scheme designed by Florian Schmidt [274]. Briefly, each shRNA is expressed from a unique polymerase III promoter (U6, H1 and 7SK for positions 1, 2 and 3, respectively) placed in a scAAV construct that also contains a stuffer sequence with a non-functional GFP transgene (used for vector titration). For each position, the respective shRNA oligonucleotide (aligned from forward and reverse) is cloned into one of three pre-TRISPR plasmids #1588 (U6), #1591 (H1) and #1594 (7SK) *via* golden gate assembly using BsmBI-v2 restriction enzyme. All three resulting intermediate plasmids are then combined for a secondary golden gate assembly with the #1600 plasmid as acceptor using the BbsI-HF restriction enzyme. Since #1600 contains a resistance to ampicillin while #1588, #1591 and #1594 carry chloramphenicol resistance, the secondary assembly is aided by antibiotic selection with ampicillin after bacterial transformation.

### *2.2.2.3 psiCheck-2 reporters*

The original psiCheck-2 reporter (#102; Promega) as well as an ssAAV vector with the *Rluc* and *Fluc* cassettes from psiCheck-2 (#0714) were used for dual luciferase

knockdown assays with targeting of *Rluc*. For targeting of *CD44* variant exon 6 (*CD44v6*), the psi_mCD44v6 reporter was cloned by amplifying the murine *CD44* variant exon 6 from Hepa1-6 cDNA using the mCD44v6_XhoI_for and mCD44v6_NotI_rev primers. Amplicon and #102 acceptor plasmid were digested with XhoI and NotI-HF (NEB), followed by ligation and transformation.

To test knockdown of anti-SARS-CoV-2 shRNA constructs, psiCheck-2 reporters were created containing the C8-C12-C3 shRNA binding sites – either as found in the SARS-CoV-2 WT genome (psi_BS_8-12-3) or with point-mutations (psi_BS_8*-12*-3* and permutations with [*] or without point mutation). This was achieved by cloning of annealed oligonucleotides with matching overhangs (BS_8-12-3_fw/_rev and respective permutations with point mutation) into the multiple cloning site of #102 using the XhoI and NotI-HF restriction enzymes for digestion of the plasmid backbone. These oligonucleotides were designed as follows:

Forward: 5' - TCGAG [C8] CTCAC [C12] CTAGA [C3] GC - 3'

Reverse: 3' - C [C8] GAGTG [C12] GATCT [C3] CGCCGG - 5'

The individual binding sites (C8/C12/C3) and mutated variants thereof (C8*/C12*/C3*) are listed in Table 21.

**Table 21: Binding sites for C8, C12 and C3 shRNAs cloned into psiCheck-2 reporters.**

| Name | Binding seq. |
|---|---|
| C8 | GTGATAGAGCCATGCCTAACA |
| C12 | GACTGAGACTGACCTTACTAA |
| C3 | GACAAGGCGTTCCAATTAACA |

| Name | Binding seq. (point mutated) |
|---|---|
| C8* | GTGATAG**G**GCCATGCCTAACA |
| C12* | GACTGAGA**T**TGACCTTACTAA |
| C3* | GACAAGGCGT**C**CCAATTAACA |

### *2.2.2.4 GFAP deletion promoter variants*

Barcoded GFAP promoter variants with 5 different deletions were created by replacing the cd11b promoter in pPL12 (#2518; ssAAV construct) with amplicons of the respective mutant version of the GFAP promoter. These were amplified from the pPL49 plasmid (#2554). GFAP_del1 was amplified using the GFAPdel1_fw and GFAP_rev primers. GFAP_del2 – del5 were generated through overlap extension PCR: the 5' region of each variant was amplified with the general GFAP_fw primer and a variant specific reverse primer (GFAPdel2_rev, GFAPdel3_rev, GFAPdel4_rev, GFAPdel5_rev), while the 3' region was amplified using a variant-specific forward primer (GFAPdel2_fw, GFAPdel3_fw, GFAPdel4_fw, GFAPdel5_fw) and the general GFAP_rev primer (in all cases: Q5 High-Fidelity 2X Master Mix, 60 °C anneal and 45 sec elongation; 35 cycles).

Respective 5' and 3' amplicons were purified *via* agarose gel electrophoresis and extraction, followed by a secondary PCR in two steps: firstly, the secondary PCR was performed without primers for 15 cycles; next, GFAP_fw and GFAP_rev primers were added (500 nM each) and 30 additional PCR cycles were conducted. The amplicons were purified through agarose gel electrophoresis and extraction, followed by cloning into pPL12 *via* digestion with SpeI-HF/NotI-HF and subsequent ligation.

#### *2.2.2.5 AAV8 peptide variants for the DEPOOL validation library*

For validating hit candidates of the initial DEPOOL screen, I assembled a barcoded library as described in section 2.2.4.6. Within this library, 39 variants were derived from the initial DEPOOL screen in lung (Lu) and pancreas (Pa), using either depleted (D) or non-depleted (ND) selection parameters. All of these represent AAV8-peptide variants with an insertion of a GQ(R/S)GNXXRXXXAQAA peptide into the VR VIII region (replacing VP1 position 588-593). Peptide sequences are displayed in the supplementary information (Supplementary table 1). For cloning of these peptide variants into *cap8*, forward (5'- [T/A] GGC AAC NNN NNN CGT NNN NNN NNN GCC CAG G) and matching reverse (5'- G GGC NNN NNN NNN ACG NNN NNN GTT GCC [A/T]CT C) oligonucleotides were designed using the codons in Table 22 for defining the respective variant amino acid positions. These codons are derived from the Ella Biotech TRIMER codon set which was employed for synthesis of the original *cap8* NXXR peptide display library utilized in the initial DEPOOL screen (described in my Master thesis [275]). Annealed oligonucleotides were ligated into the WH-Rep2-Cap8IS (#200) plasmid after digestion thereof with SfiI (NEB).

**Table 22: Trimer codon table used for design of peptide oligonucleotides.**

| Amino acid (one-letter code) | Amino acid (three-letter code) | DNA codon |
|---|---|---|
| **I** | Ile | ATC |
| **M** | Met | ATG |
| **T** | Thr | ACC |
| **N** | Asn | AAC |
| **K** | Lys | AAA |
| **Q** | Gln | CAG |
| **H** | His | CAT |
| **L** | Leu | CTG |
| **P** | Pro | CCA |
| **R** | Arg | CGT |
| **V** | Val | GTT |

| Amino acid (one-letter code) | Amino acid (three-letter code) | DNA codon |
|---|---|---|
| **A** | Ala | GCA |
| **D** | Asp | GAT |
| **E** | Glu | GAA |
| **G** | Gly | GGT |
| **S** | Ser | TCT |
| **F** | Phe | TTC |
| **Y** | Tyr | TAC |
| **C** | Cys | TGC |
| **W** | Trp | TGG |

#### *2.2.2.6 RNA-selection constructs for in vitro testing*

I cloned constructs for testing of p40- or CMV-promoter-driven *cap2* expression *in vitro* (section 3.2.3.1) over multiple iterations as described below.

##### 2.2.2.6.1 BYRPc2

Initially, the RSV-eYFP-bgh-pA cassette from plasmid #2543 was amplified with the RS_BglII_pA and RS_BamHI_RSV primer set. The amplicon was digested with BglII and BamHI-HF (NEB) for cloning into the #0778 acceptor plasmid (pSSV9 acceptor containing *rep2* and a cloning site for *cap* genes). This intermediate plasmid was then digested with BamHI-HF/AscI, followed by ligation of a gene block sequence (ordered from GENEWIZ/Azenta Life Sciences) containing the p40 promoter with the 3' end of *rep*, a PacI restriction site, the full-length *cap2* gene and an AscI restriction site. This yielded the BYRPc2 construct ("B" = bgh-pA; "Y" = eYFP; "R" = RSV; "P" = p40; "c2" = *cap2*).

##### 2.2.2.6.2 rep-c2

The rep-c2 (wild-type AAV2) construct was cloned by transferring the *cap2* from BYRPc2 to the #0778 plasmid *via* PacI/AscI restriction and ligation.

##### 2.2.2.6.3 BYECc2

For creating CMV promoter-driven *cap2* expression constructs, I digested the BYRPc2 plasmid with BamHI-HF/PacI for removing p40 and *cap* intron, followed by ligation of a gene block (GENEWIZ) containing the *cap2* splice donor and intron sequence (starting at the FspI site) with matching overhangs. This yielded the oversized BYRCc2 construct (5.2 kb including ITRs; "C" = CMV). To adapt for improved packaging (size limit: 4.9 kb including ITRs), the RSV-eYFP-bgh-pA cassette was replaced with an EFS-eYFP-bgh-pA cassette amplified from #2531 using RS_BamHI_EFS and RS_BglII_pA primers. Ligation yielded the BYECc2 construct ("E" = EFS).

#### 2.2.2.6.4 BYxCc2S

The BYxCc2S construct was cloned by firstly removing the RSV promoter from BYRPc2 using restriction with NotI-HF/BamHI-HF, followed by ligation of an oligonucleotide annealed from delRSV_FW and delRSV_RV. This yielded the BYxCc2 intermediate. From this, the endogenous AAV pA was replaced with the SV40 late pA (SV40pA) [161] amplified from the #1831 plasmid using the SV40pA_SpeI_for and SV40pA_BstBI_rev primers. Both amplicon and BYxCc2 were digested with SpeI-HF and BstBI (NEB), followed by ligation to generate the BYxCc2S plasmid ("x" = deletion; "S" = SV40pA).

#### 2.2.2.6.5 MYECc2S

The MYECc2S construct ("M" = mini-pA) was cloned by exchanging the non-functional *eYFP* cassette in BYxCc2S with an EFS-eYFP-minipA cassette. This was achieved by digesting BYECc2 with BglII and ClaI to remove the bgh-pA, and instead inserting a minimal poly-A (minipA) [276] *via* ligation of an oligonucleotide annealed from ClaI_minipA_fw and BglII_minipA_rev. Next, the endogenous AAV poly-A was replaced with SV40-pA as described above for BYxCc2S.

#### 2.2.2.6.6 BYxCPc2

To create a construct with *cap2* expression driven by both CMV and p40 promoters, the BYRPc2 plasmid was digested with BamHI-HF/PacI and used for ligation of a geneblock (GENEWIZ) containing the p40 promoter and *cap2* intron sequence with matching overhangs. This yielded the oversized BYRCPc2 intermediate (5.3 kb), from which the RSV promoter was removed as described above for BYxCc2S, yielding BYxCPc2.

#### 2.2.2.6.7 CAGc2M

For creating CAGc2M, I conducted cloning over five steps: (i) a mini-pA was added to BYxCc2 by digesting the plasmid with SpeI-HF and BstBI, and ligating an oligonucleotide annealed from minipA_SpeI_for and minipA_BstBI_rev to generate the BYxCc2M intermediate. (ii) To construct an intermediate plasmid for easier promoter switching, the bgh-pA, *eYFP* and CMV were removed by digesting with BglII/PacI and inserting an amplicon containing a multiple cloning site and the *cap2* intron (amplified from BYxCc2M using rep2i_for and rep2i_rev primers). (iii) Next, *cap2* was removed by digesting with PacI/AscI and ligation of an oligonucleotide annealed from cap-rplc_for and cap-rplc_rev, yielding the PCswitch_mini-pA intermediate. (iv) The CAG promoter (CMV enhancer with chicken β-actin promoter and synthetic intron [178]) was excised from #2517 using SpeI-HF/NotI-HF and ligated into the BsmBI-v2/NotI-HF-digested PCswitch_mini-pA (CAGxM).

(v) Finally, *cap2* was excised from BYxCc2 using PacI/AscI restriction and ligated into PacI/AscI-digested CAGxM to yield CAGc2M.

2.2.2.6.8 rep-z

The *cap2* from rep-c2 was excised using PacI/AscI restriction and replaced by an inert *lacZ* stuffer amplified from #2564 using the PacI_lacZ_2.2_fw and AscI_lacZ_2.2_rev primers. This yielded the rep-z construct ("z" = *lacZ* stuffer).

2.2.2.6.9 BYzPc2

BYzPc2 was cloned by replacing the RSV promoter from BYRPc2 with an inert *lacZ* stuffer of the same size. This stuffer was amplified from #2564 using the *LacZ*-for and NotI_lacZ_0.6 primers. BYRPc2 backbone and insert were digested with BamHI-HF and NotI-HF prior to ligation.

2.2.2.6.10 zPc2

Analogous to BYzPc2, zPc2 was cloned by replacing the RSV-eYFP-bgh-pA cassette from BYRPc2 with an inert *lacZ* stuffer amplified from #2564 using the LacZ-for and BglII_lacZ_1.6 primers. BYRPc2 backbone and *lacZ* insert were digested with BamHI-HF and BglII prior to ligation.

### *2.2.2.7 AAV6 peptide display: CMV-cap6-P7P library*

The *cap6* with peptide insertion site in VR VIII (cap6_nis) was excised from plasmid #1614 using AscI/PacI and used for replacing *cap2* in the rep-c2 and MYECc2S plasmids. This yielded the rep-c6_nis and MYECc6nisS plasmids. Peptide insertion was performed *via* ligation of partially randomized oligonucleotides into the insertion site of cap6_nis. The insertion replaces *cap6* VP1 amino acid positions 586-591 with the peptide sequence GQRGPXXXXXXXPAQAA, *i.e.*, a randomized heptamer peptide with adjacent flanking sequences. I achieved this by (i) digesting rep-c6_nis and MYECc6nisS with SfiI. (ii) Next, I prepared the oligonucleotide insert by performing second-strand synthesis of the partially randomized P7P_oligo (5'- `CA GTC GGC CAG AGA GGC CCA NNN NNN NNN NNN NNN NNN NNN CCA GCC CAG GCG GCT GAC GAG`; ordered from Ella Biotech by Joanna Szumska; `NNN` following the codons in Table 22 with equal contributions of each codon). For second strand synthesis, a Phusion HS II polymerase reaction was set up using the P7P_oligo and the pept_amplif_rev reverse amplification primer. PCR reaction was conducted for 3 cycles, using 60 °C annealing and amplification for 10 sec. (iii) The now double-stranded oligonucleotide was digested with BglI (NEB) and purified with the QIAquick Nucleotide Removal kit. (iv) Finally, ligation of digested and purified plasmids

with the oligonucleotide insert was performed using a molar ratio of 1:10 (plasmid : oligonucleotide). Bacterial transformation of plasmid libraries was performed by electroporation.

#### *2.2.2.8 AAV6 peptide display: second round variants and libraries*

After the initial selection *in vivo*, I selected 536 AAV6 peptide candidates for secondary validation. This secondary library (cap6-536) was generated by pooled oligonucleotide synthesis: the 536 oligonucleotide variants were designed according to the P7P_oligo described above, each coding for one of the respective validation peptides (using codons in Table 22). These oligonucleotides were synthesized in one pool using the Twist Bioscience Oligo Pool service. Second-strand synthesis and amplification were conducted as recommended by the manufacturer, using the KAPA HiFi HotStart ReadyMix PCR Kit (Roche) in a 50 µL reaction volume with 10 ng single stranded oligonucleotide pool as template and 500 nM of pept_amplif_for and pept_amplif_rev amplification primers. Cycling was conducted as depicted in Table 23. Amplicons were purified with the QIAquick Nucleotide Removal kit, followed by BglI digestion and ligation into SfiI-digested rep-c6_nis (rep-cap6-536 library) and MYECc6nisS (CMV-cap6-536 library) plasmids as described above.

I used the individual variant CMV_cap6-cH001 as a target plasmid for DEPOOL *in vitro* testing. This individual plasmid was cloned by ligation of annealed cH001_f and cH001_r into SfiI-digested MYECc6nisS.

**Table 23: Cycling conditions for AAV6 peptide pool oligonucleotides.**

| Step | Temperature [°C] | Time [min:sec] | Cycles |
|---|---|---|---|
| Initial denaturation | 95 | 3:00 | |
| Denaturation | 95 | 0:20 | 10 |
| Annealing | 58 | 0:15 | |
| Elongation | 72 | 0:15 | |
| Final elongation | 72 | 1:00 | |
| Hold | 4 | ∞ | |

### 2.2.3 Cell culture

#### *2.2.3.1 Maintenance of eukaryotic cells*

Cultured cells were maintained in 75 $cm^2$ or 175 $cm^2$ flasks (Greiner Bio-One) in complete medium consisting of DMEM GlutaMAX (Thermo Fisher Scientific) supplemented with 10% Fetal bovine serum (FBS; Capricorn Scientific) and 1% Penicillin / Streptomycin (Thermo Fisher Scientific). For Huh7 cells, medium was additionally supplemented with 1% MEM Non-Essential Amino Acids Solution (Thermo Fisher Scientific). Cells were incubated at 37 °C with 5% $CO_2$.

Depending on confluency and incubation time, I split cells at 1:3 to 1:10 every 2-4 days by removing supernatant, washing with PBS, and trypsinization using 2 mL 0.25% Trypsin / EDTA (Thermo Fisher Scientific) per flask. After incubation for 2-5 min at 37 °C, trypsinization was stopped by addition of 5× volume (10 mL) of complete medium and resuspension. Cell numbers were quantified using the Countess automated cell counter after staining 1:1 with Trypan Blue Solution (Thermo Fisher Scientific).

#### *2.2.3.2 Lipofectamine transfection*

For Lipofectamine transfections, I seeded Hek293T cells in 96-well plates at a density of $1.5\times10^4$ cells in a volume of 100 µL per well and incubated for 24 h. Next, DNA mixes were prepared by diluting up to 200 ng of plasmid DNA in 25 µL of supplement-free DMEM per well. Secondly, a Lipofectamine mix was prepared containing 25 µL of supplement-free DMEM and 0.4 µL of Lipofectamine 2000 (Thermo Fisher Scientific) per well. Next, DNA and Lipofectamine mixes were combined, vortexed, and incubated at rt for 20 min, before adding the entire mix dropwise to each a well.

#### *2.2.3.3 Dual luciferase knockdown assay*

I performed dual luciferase assays for quantifying shRNA- or CasRx-based knockdown by targeting either the *Renilla*-luciferase (*Rluc*) of psiCheck-2 directly or target sequences cloned into the 3'-UTR of *Rluc*. For each assay, *Rluc* expression was modified by different effectors (shRNAs or CasRx/gRNA constructs), while the Firefly luciferase (*Fluc*) expression remained unaffected and was quantified as an internal reference. Expression of psiCheck-2 reporter (or derivatives thereof) and effectors was achieved by either (i) Lipofectamine transfection of Hek293T cells or (ii) by transduction of Hek293T or Huh7 cells. Each assay was performed in triplicates within a 96-well plate, using the Dual-Luciferase Reporter Assay System for cell lysis and luciferase quantification (Promega).

For (i) Lipofectamine transfection, each well was transfected with 100 ng effector (shRNA or CasRx gRNA; plus 25 ng miRNA effector for section 3.1.1.2) and 10 ng reporter (psiCheck-2 or derivates thereof). For (ii) transduction based knockdown assays, cells were co-transduced with two AAV2 vectors: a reporter vector packaging psiCheck-2 as the transgene (AAV transgene plasmid: #0714) and an effector vector packaging a CasRx-gRNA construct. Reporter vectors were applied with multiplicity of infection (MOI) of $1\times10^4$ vg per cell, whereas effectors were applied with an MOI of $1\times10^5$.

Cells were lysed on day 3 post transfection or transduction with 30 µL 1x passive lysis buffer (from the Dual-Luciferase Reporter Assay System kit) per well. For signal detection, 10 µL lysate were transferred to each well within white 96-well plates. Luciferase signal was measured with the GloMax Navigator Microplate Luminometer with dual injectors/pumps (Promega) using the following program: (1) injection of 35 µL LAR II reagent, (2) 2 sec delay, (3) detection of Firefly luminescence signal with 10 sec integration, (4) 2 sec delay, (5) injection of 35 µL Stop & Glo Reagent, (6) 2 sec delay and (7) detection of *Renilla* luminescence signal with 10 sec integration. Analysis was performed by normalizing the Rluc signal to the Fluc signal from each well, giving relative luminescence units (RLU) as output. RLU values were then normalized to the respective control conditions (*e.g.*, non-targeting control).

### 2.2.4 Virological methods

#### *2.2.4.1 AAV vector production by triple transfection*

Recombinant (r)AAV vectors were produced by triple-transfection with polyethyleneimine (PEI) using the three plasmid types (i) adenoviral helper plasmid (AdH), (ii) AAV-helper plasmid and (iii) transgene-plasmid. Of these, (i) AdH expresses the adenoviral helper genes E2A, E4 and VA RNA, whereas (ii) the AAV-helper expresses *rep2* and a *cap* gene of choice defining the capsid of the assembled vectors. (iii) The transgene plasmid contains an ITR-flanked packaging cassette for either single-stranded (ss)AAV vectors (ITRs of AAV2 flanking transgene cassettes of 3 – 4.9 kb size) or self-complementary (sc)AAV vectors (ITR2 and mutated ITR4 flanking a transgene cassette of 1.6 – 2.4 kb size) and thus defines the vector cargo.

For virus production by triple transfection, Hek293T cells were seeded on 150 mm cell culture dishes (Sarstedt; hereafter referred to as plates) with a density of $4\times10^6$ cells in 22 mL complete medium per plate (cell harvesting and culture conditions as described in section 2.2.3.1). Depending on the required quantity of vector product, productions were

performed with 3-10 plates per small iodixanol gradient purification (see section iodixanol) and 10-40 plates per large gradient. At day two after seeding, cells were transfected with 2 mL transfection mix per plate containing a total of 44 µg DNA with a molar ratio of 1:1:1 between AdH, AAV-helper and transgene plasmid. Each transfection mix contained a final NaCl concentration of 300 mM and PEI MAX (Polysciences Europe GmbH) at an N/P ratio (amines in PEI vs phosphates in DNA) of 30. DNA, nuclease-free $H_2O$ (Aqua B. Braun) and NaCl were pre-mixed, after which PEI MAX was added. The mixture was then vortexed, incubated for 10 min at rt, and added dropwise to the plate of pre-seeded cells. On day 3 post transfection, cells were harvested by gentle lifting with a cell lifter (Corning) and collected by resuspension in their supernatant. Cell suspensions were then centrifuged at 800 × g for 15 min, after which the supernatant was discarded, and the cell pellet was resuspended in Benzonase buffer. For each small iodixanol gradient purification, cell pellets from 3-10 plates were resuspended in 5 mL Benzonase buffer, while pellets from 10-40 plates were resuspended in 20 mL buffer per large gradient. Next, cells were lysed by five iterative cycles of (i) incubation in liquid nitrogen for 5 min and (ii) 37 °C water bath for 15 min. After freeze/thaw lysis, non-packaged nucleic acids were removed by digestion with 50 U/ml Benzonase (Merck) at 37 °C for 60 min while inverting the tube at 10 min intervals. Lysates were then centrifuged three times at 4,000 × g for 15 min, each time collecting the supernatant (lysate containing vector particles) and discarding the pelleted cell debris. The cleared lysate was then subjected to iodixanol gradient density centrifugation for purification of AAV vectors.

#### *2.2.4.2 Production of AAVs for rep-, p40- and CMV-cap constructs*

*Cap*-library-mimicking *cap2* constructs for *in vitro* testing of RNA-selection settings and *cap6*-libraries (section 3.2.3) were produced by PEI transfection with a two-plasmid setting: one plasmid being the ITR-bearing *cap* plasmid and the other being a helper plasmid. For *rep-cap* constructs (rep-c2 in section 3.2.3.1, rep-cap6-536 peptide library in section 3.2.3.3), AdH was used as a helper plasmid as before, while the *rep*- and AdH-expressing pDGΔVP plasmid was used as helper for *cap*-packaging constructs missing *rep* (BYRPc2, BYxCPc2, BYxCc2S, BYECc2, MYECc2S, CAGc2M, BYzPc2, zPc2 in section 3.2.3.1, or CMV-cap6 peptide libraries in sections 3.2.3.2/3.2.3.3). To prevent excessive cross-packaging and to ensure genotype-phenotype linkage, the number of *cap*-library plasmids was limited to 5,000 copies per cell (counting with a doubling time of 24 h for Hek293T cells, *i.e.*, $1.6\times10^7$ cells and $8\times10^{10}$ copies of cap plasmid per plate).

For the respective helper, a 20-fold molar excess was used as compared to the ITR-*cap* plasmid. To produce the second round CMV-cap6 peptide library (section 3.2.3.3), I tested the addition of a second helper plasmid expressing AAV2 *AAP* as well (CMV-AAP plasmid). Here, a molar ratio of 1:20:10 was used (ITR-*cap* : pDGΔVP : CMV-AAP). Transfection, harvesting, and processing were otherwise performed as described above.

#### *2.2.4.3 AAV purification by iodixanol gradient density centrifugation*

I purified AAV vectors using iodixanol gradient density centrifugation with either small or large gradient settings. For small gradients (lysate from 3-10 plates), cleared cell lysates were transferred to re-seal polyallomer centrifuge tubes (16 × 76 mm; Seton Scientific) and consecutively sub-layered with 2 mL of each of the 15%, 25%, 40% and finally 60% iodixanol phases. Tubes were sealed using the Tube Sealer and balanced pairwise. Ultracentrifugation was performed with the 70.1 TI rotor in the OptimaTM L-90K ultracentrifuge (Beckman Coulter) at 50,000 rpm for 2 h at 4 °C. After centrifugation, tubes were punctured at the top (for pressure release) and the lowest part of the tube bottom (for virus collecting) using syringe needles. By controlling the pressure release on the top puncture, the lower phase was released dropwise from the tube bottom. Here, the lowest 1.5 mL were discarded while the following 0.8 mL were collected containing the majority of assembled full capsids.

For large gradients (lysate from 10-40 plates), up to 20 mL lysate were applied to Quick-Seal centrifuge tubes (25 × 89 mm; Beckman Coulter) and consecutively sub-layered with 7 mL of 15%, 5 mL of 25%, 4 mL of 40% and 4 mL of 60% iodixanol phases. After sealing, centrifugation was conducted in a 70 TI rotor with the OptimaTM L-90K ultracentrifuge at 50,000 rpm for 2.5 h at 4 °C. Virus collection was performed as above, but instead discarding the bottom-most 3 mL and collecting the following 1.8 mL.

For virus samples used in transducing mice or primary human hepatocytes, I performed a buffer exchange to remove iodixanol from purified virus products. This was achieved by mixing the purified virus (1.8 mL for large gradients) to 13 mL PBS and transferring the mixture to an Amicon Ultra-15 Centrifugal Filter Unit (MWCO 100,000; Merck). Centrifugal filters were then centrifuged iteratively at 1,000 × g for 2 min until only 2 mL remained in the top part. This supernatant was again filled to 15 mL with PBS and centrifuged as before, until reaching the desired volume of 300 – 600 µL of final product.

#### *2.2.4.4 AAV titration by qPCR*

For vectors produced in projects 3.1 and 3.2.2, I conducted vector titration using probe-based quantitative (q)PCR by establishing a standard curve with reference plasmids of a known copy number. Vector samples were prepared by alkaline lysis, mixing 1 µL virus, 19 µL TE buffer and 20 µL of 2 M NaOH (Merck). Lysis was performed by incubation at 56 °C for 30 min, after which the solution was neutralized by addition of 38 mL 1 M HCl (VWR Chemicals). Next, 922 µL of nuclease-free $H_2O$ were added, resulting in a 1:1,000 dilution of the initial virus sample. Reactions for qPCR were prepared in triplicates, where 3.5-fold triplex master mixes were prepared containing 17.5 µL Sensimix II Probe mix, 9.35 µL nuclease-free $H_2O$, 1.4 µL of each 10 µM forward and reverse primer (final concentration: 400 nM), 0.35 µL of 10 µM probe (final concentration: 100 nM) and 5 µL of sample (virus after alkaline lysis or standard plasmid with $3.5\times10^{10}$ - $3.5\times10^{4}$ copies in 1:10 dilution steps). The triplex mix was then separated into three reactions with 10 µL reaction volume and run with the Corbett Rotor-Gene 6000 system (Qiagen) using the conditions described in Table 24. Primers/probe sets and corresponding reference plasmids used for quantification were the GFP set (eGFP_f/eGFP_r/eGFP_probe and eGFP standard plasmid #0552) for shRNA and GFP vectors, the hluc set (hluc_f/hluc_r /hluc_probe and hluc standard plasmid #0714) for AAV2/psiCheck-2, the rep2 set (rep2_f/rep2_r/rep2_probe and rep2 standard plasmid #0778) for rep-cap-BC vectors and the U6 set (U6_f/U6_r/U6_probe and U6 standard plasmid #0067) for CasRx vectors.
Vector concentration per 10 µL reaction was calculated by linear regression from the standard curve (using the Rotor Gene Q-Series Software, Qiagen) and translated to vector copy number per mL by correcting *via* multiplying with 7 (5 µL sample per 35 µL triplex reaction mix) × 100 (from 10 µL reaction to concentration per mL) × 1,000 (dilution factor after alkaline lysis) × 2 (only for single-stranded vectors).

**Table 24: Cycling conditions for AAV titration by qPCR.**

| Step | Temperature [°C] | Time [min:sec] | |
|---|---|---|---|
| Initial denaturation | 95 | 10:00 | |
| Denaturation | 95 | 00:10 | 40 cycles |
| Annealing/elongation | 60 | 00:20 | |

#### 2.2.4.5 AAV titration by ddPCR

For vectors produced in projects 3.2.1 and 3.2.3, AAV titration was performed using ddPCR. Herre, vector samples were diluted at 1:$10^5$ to 1:$10^7$, of which 5 µL were used as template for ddPCR reactions containing additionally 10 µL ddPCR Supermix for Probes, 1 µL of 20x target primer/probe mix for transgene, 1 µL of 20x target primer/probe mix for ITR (final concentration: 900 nM for primers, 250 nM for probes) and 3 µL $H_2O$. Droplet generation and ddPCR were otherwise performed as described in section 2.2.1.15. Transgenes were measured with FAM-labeled probes using the GFP set (eGFP_f, eGFP_r and eGFP_probe) for promoter-eYFP constructs in section 3.2.1, the cap2 set (cap2_f, cap2_r and cap2_probe) for *cap2*-vectors in section 3.2.3.1, the rep2 set (rep2_f, rep2_r and rep2_probe) for construct rep-z in 3.2.3.1, and the cap6 set (cap6_f, cap6_r and cap6_probe) for library constructs in sections 3.2.3.2/3.2.3.3. In the same reaction, ITRs were quantified on the HEX channel using the ITR set (ITR_f, ITR_r and ITR_probe). Vector quantity was determined by firstly calculating a correction factor CF that quantifies the ratio of transgene- and ITR-positive droplets amongst the set of all transgene-positive droplets (ranging from 0.9-1.0 for most vector productions). The number of vector genomes per µL (c) was then calculated from the measured transgene copy number per reaction (N) by applying the correction factor CF, the dilution factor DF (*e.g.*, $10^6$ for measuring a 1:$10^6$ diluted vector sample) and finally by dividing by 5 to account for 5 µL template volume per reaction:

$$c = \frac{N \times CF \times DF}{5}$$

#### 2.2.4.6 Pooling of barcoded vector libraries

For vector libraries packaging *eYFP* transgenes with variant-specific barcodes, each variant was produced separately as described in section 2.2.4.1. I generated two such libraries for this thesis: (i) the DEPOOL validation library (section 3.2.2.1), where each variant represents a different capsid packaging a CMV-eYFP-BC transgene in a scAAV setting, and (ii) the promoter mini-libraries, where promoter variants were packaged in either AAV9 or AAV-DJ capsids. In each case, the DNA barcode was present in the 3'-UTR of *eYFP* for enabling readout from both DNA and RNA as demonstrated in Weinmann *et al*. [142].

(i) The DEPOOL validation library contains benchmark capsid variants AAV2_L1, AAV4, AAV6.2, AAV7, AAV8, AAV9 and AAVrh10_P1, as well as 39 different AAV8-peptide variants resulting from the initial DEPOOL screen conducted during my Master thesis project (cloning: see section 2.2.2.5). Of these 39

variants, 10 were selected from each non-depleted (ND) or depleted (D) settings with primary target being lung (Lu) or pancreas (Pa). One variant (LuD9PaD2) was discovered in both depleted screens. Each variant packages an scAAV-eYFP-BC construct expressing *eYFP* from a CMV promoter, with a variant-specific barcode in the 3'-UTR of *eYFP* (cloned by Jonas Weinmann, [142]).

(ii) For the promoter mini-libraries, 10 promoter constructs were selected: CMV (pPL04), LP1 (pPL06), SPc5-12 (pPL31), gfaABC1D (pPL48) and GFAP (pPL49) were all cloned by Claire Domenger. The GFAP deletion mutants GFAP_del1-5 were cloned as described in section 2.2.2.4.

For these libraries, titrated vectors were mixed with equal variant contributions and re-concentrated using Amicon filter units as described in section 2.2.4.3. Titers from re-concentrated libraries were quantified as before, and library compositions were evaluated based on Illumina sequencing of the packaged barcodes using the NGS methods described in section 2.2.6.

#### *2.2.4.7 Silver staining*

To evaluate the presence and ratios of VP1:VP2:VP3 capsid proteins within assembled and purified AAV vectors (section 3.2.3.1), I conducted SDS-PAGE and subsequent silver staining. Purified and titrated AAV vectors were thus prepared by diluting $5\times10^9$ vg per sample with PBS to a total volume of 20 µL. Next, 7 µL of 4x Laemmli Sample Buffer (Bio-Rad) supplemented with 10% β-mercaptoethanol were added to each sample, followed by incubation at 95 °C for 5 min and 4 °C for 10 min. A reference protein ladder sample was prepared by mixing 2 µL PageRuler Plus (Thermo Fisher Scientific) to 18 µL PBS and 7 µL 4x Laemmli buffer. SDS-PAGE was assembled by placing the Mini-PROTEAN TGX Precast Protein gel in the Mini-PROTEAN Tetra Cell chamber system filled with 1x Tris/Glycine/SDS Electrophoresis Buffer (Bio-Rad). Samples were loaded and run at 90 V for 10 min followed by 110 V for 80 min. The gel was then removed and washed with $H_2O$ (Aqua B. Braun), before performing fixing, sensitizing, staining, and developing with the SilverQuest Silver Staining Kit (Invitrogen) according to the manufacturer's instructions. Imaging was performed with the Azure 400 imaging system.

#### *2.2.4.8 Dot blot*

Assembled AAV2 capsids (section 3.2.3.1) were visualized *via* dot blot using the Bio-Dot Apparatus (Bio-Rad). Samples were prepared by mixing $1\times10^{10}$, $5\times10^{9}$ or $1\times10^{9}$ vg of vector samples with PBS to a total volume of 100 µL. The Bio-Dot Apparatus was assembled with a nitrocellulose membrane (Blotting Membran CN, neoLab) pre-soaked in PBS. Samples were loaded to each well, after which vacuum was applied. After samples had passed through the membrane, each well position was washed twice with PBS. The membrane was then removed and blocked by incubating in 5% milk (w/v) in TBS-T buffer for 1 h at rt. Next, the membrane was incubated for 2 h at rt with the primary A20 antibody (detecting assembled AAV2 capsids, [270]) diluted 1:10 in 5% milk. After primary incubation, the membrane was washed thrice with TBS-T and subsequently incubated at rt for 1 h with the secondary α-mouse-HRP antibody (115-035-003; Jackson ImmunoResearch Laboratories) diluted 1:10,000 in 5% milk. The membrane was stained using the WesternBright Chemiluminescence Substrate Quantum (Biozym) and imaged with the Azure 400 system.

#### *2.2.4.9 Assessment of packaged genomes via native agarose gels*

To visualize vector genomes from packaged into AAV vectors, vector DNA was extracted and separated with native agarose gel electrophoresis: For scAAV vectors in section 3.1.2.2, $5\times10^{11}$ vg per sample were diluted with PBS to a final volume of 600 µL and incubated with 60 µL Proteinase K (Qiagen) for 30 min at 56 °C. Next, DNA extraction was performed with the DNeasy Blood and Tissue kit (Qiagen) following the manufacturer's instructions. DNA was eluted with 30 µL nuclease-free $H_2O$ and incubated at 95 °C for 2 min, followed by incubation on ice for 10 min. Each sample was then mixed with 6 µL purple gel loading dye (NEB) and loaded next to the 1 kb Plus DNA ladder to 1% agarose gels stained with ethidium bromide. Running and imaging was performed as described above (section 2.2.1.2). Relative band density was determined using the ImageJ 1.53c software (Wayne Rasband, NIH) [269].

For ssAAV vectors in section 3.2.3.1, the extraction was performed for $2\times10^{11}$ vg per sample. Here, DNA staining was performed with Gelred (Biotium) instead of ethidium bromide to improve visualization of single stranded DNA fragments. Thus, extracted DNA (30 µL) was mixed to 6 µL TrickTrack loading dye (Thermo Fisher Scientific) supplemented with 10% Gelred prior to loading. Here, the GeneRuler DNA Ladder Mix (Thermo Fisher Scientific; supplemented with Gelred) was used for size comparison.

#### *2.2.4.10 AAV transduction in vitro*

I performed AAV *in vitro* transductions directly after seeding of cells in 96-, 48- or 24-well plates (respective seeding density: $1.5\times10^4$ / $3\times10^4$ / $1\times10^5$ cells per well). Therefore, AAV vectors were diluted in complete medium to a volume of 10 µL (96-well) or 50 µL (48- or 24-well plate) for transductions with MOIs between $10^3$ and $10^5$ vg per cell (as indicated respectively). Unless stated otherwise, cells were harvested on day 3 post transduction. For dual luciferase assays, cell harvesting and analysis were performed as described in section 2.2.3.3. For DNA/RNA extractions, cells were harvested by (i) removing supernatant and washing with PBS, (ii) incubation with 0.25% Trypsin / EDTA for 5 min at 37 °C (100 µL per well for 48- and 24-well plates), (iii) resuspending in complete medium (400 µL per well for 48- and 24-well plates), and (iv) centrifugation at 1,000 × g for 3 min. Cell pellets were washed with PBS and stored at -80 °C prior to processing.
For transduction of primary human hepatocytes (kindly provided by the DZIF PHH Core Facility), cells were received plated on 48-well plates. I subjected these cells to transduction with AAV-DJ vectors using an MOI of $10^5$ vg per cell and harvested them at day 3 post transduction as described above.

#### *2.2.4.11 Infection of Vero E6 cells with SARS-CoV-2 and infection analysis*

For SARS-CoV-2 infection assays, I seeded Vero E6 cells in 48- or 24-well plates (as described above) and transduced them with AAV-LK03 vectors using an MOI of $10^5$ vg per cell. Subsequent steps of these experiments (infection with SARS-CoV-2, infection quantification *via* qPCR and indirect immunofluorescence (IF) assay) were performed by Megan Stanifer (Boulant lab). Briefly, at three days post AAV transduction, cells were infected with SARS-CoV-2 (BavPat1 strain, passage 3). For the passaging experiment (Figure 15), 50 µL of supernatant was collected at day 1 post infection and added to the next passage of cells (transduced 3 days earlier). SARS-CoV-2 infection was analyzed by indirect immunofluorescence (IF) assay (by Megan Stanifer) to quantify the percentage of infected cells. This was achieved by fixation of cells with 4% paraformaldehyde for 20 min at rt, washing with PBS and permeabilization with 0.5% Triton X-100 (15 min incubation at rt). Cells were stained by incubation for 1 h with a primary antibody binding double-stranded (ds)RNA (anti dsRNA mAb J2, 10010200; SCICONS/Nordic-MUbio; diluted 1:1,000 in PBS), followed by washing with PBS and incubation for 45 min at rt with a fluorescent secondary antibody (IRDye 800CW Donkey anti-Mouse IgG Secondary Antibody, 926-32212; LI-COR Biosciences; diluted 1:10,000 in PBS) and DRAQ5 DNA

dye (Abcam; diluted 1:1,000). After washing with PBS again, cells were imaged with the Odyssey CLx imager (LI-COR Biosciences).

Quantification of SARS-CoV-2 genomes was performed by Megan Stanifer using RTqPCR. RNA was extracted from infected cells with the RNeasy Mini kit (Qiagen), followed by cDNA synthesis with the iScript Reverse Transcription kit (Bio-Rad) with 250 ng RNA input per sample. RTqPCR was performed using the iTaq Universal SYBR Green Supermix system (Bio-Rad) for detecting HPRT housekeeper (HPRT1_f and HPRT1_r) and SARS-CoV-2 cDNA (SARS-CoV-2_f and SARS-CoV-2_r ).

For Sanger sequencing of SARS-CoV-2 passaged in pre-transduced cells, supernatant from passages 1, 3, 5, and 7 were inactivated in lysis buffer by Megan Stanifer. From these supernatants, I then performed RNA extraction, cDNA synthesis and Sanger sequencing in order to determine the genetic sequence of the shRNA-binding sites within the SARS-CoV-2 genome of each sample. SARS-CoV-2 RNA was extracted from supernatants using the RNeasy Mini Kit, eluting with 30 µL of nuclease-free $H_2O$ per sample. Next, cDNA synthesis was performed from the entire eluate as described in section 2.2.1.13 using the High Capacity cDNA Synthesis kit, followed by PCR amplification of two regions of interest: Region 1 is a 748 nt region in the SARS-CoV-2 RdRP gene that contains binding sites for shRNAs C8 and C12 (amplified *via* C8C12_Rseq_fw and C8C12_Rseq_rev), whereas region 2 is a 274 nt region in the N gene that contains the C3 binding site (amplified *via* C3_Nseq_fw and C3_Nseq_rev). PCR was performed using the Phusion HS II polymerase as described in section 2.2.1.2 (annealing at 60°C, elongation for 30 s). Successful amplification was confirmed *via* agarose gel electrophoresis (section 2.2.1.2). Sanger sequencing was performed as described in section 2.2.1.11 using 1 µL PCR product as template and respective forward or reverse amplification primers.

### 2.2.5 Animal experiments

#### *2.2.5.1 Animal experiments conducted in Heidelberg*

For the promoter evaluation and DEPOOL projects (section 3.2.1.1 and 3.2.2.1, respectively), animal studies were conducted by me at the IBF Heidelberg following certification with FELASA-A/D and approval by German authorities (35-9185.81/G-89/16). For the single vector experiment in the promoter project (section 3.2.1.1), female 6-week-old C57BL/6 mice were acquired from Janvier Labs (Le Genest-Saint-Isle, France) and injected at an age of 8 weeks with $10^{12}$ vg/mouse using the CMV-, LP1-,

SPc5-12- or GFAP-eYFP vectors, each packaged into the AAV9 capsid. Mice were sacrificed at two weeks post injection, followed by extraction of 16 major tissue types (brain, diaphragm, eye, (white) fat, gut, heart, kidney, liver, lung, (inguinal) lymph nodes, muscle (quadriceps femoris), ovaries, pancreas, skin, spleen, and stomach) which were stored in RNAlater until processed.

For the DEPOOL project (section 3.2.2.1), female 6-week-old NMRI mice were acquired from Charles River Laboratories (Wilmington, USA) and injected at 8-weeks old with $10^{12}$ vg/mouse of the barcoded DEPOOL validation library (containing 47 capsid variants) *via* tail-vein injection. Mice were sacrificed two weeks post injection. Liver, lung, pancreas, spleen, muscle (quadriceps femoris) and kidney were collected and stored in RNAlater (Thermo Fisher Scientific) prior to DNA/RNA extractions (described in section 2.2.1.12).

#### *2.2.5.2 Mouse experiments conducted at the Willenbring laboratory*

All other mouse experiments presented here were conducted by Pervinder Choksi and Holger Willenbring (University of California, San Francisco), including systemic mouse injections (retroorbital), sacrifice, liver perfusion and FACS-based isolation of the different cell types. Mouse procedures conducted in the Willenbring lab were approved by the Institutional Animal Care and Use Committee at UCSF. Isolated cells were shipped to Heidelberg for further processing by me (including DNA/RNA extractions, ddPCR, RTqPCR, NGS). These collaborative experiments with mouse work done in the Willenbring lab include (i) the AAV9/GFAP study in different cell types of the mouse liver (section 3.2.1), (ii) the application of the AAV9 promoter mini-library in a humanized liver mouse (section 3.2.1), and (iii) the application of AAV6-peptide display libraries for *in vivo* RNA-based capsid screening (section 3.2.3.2).

For (i) the AAV9-GFAP study in different liver cells, female *Lrat-Cre*$^{+/-}$*;R26-RFP*$^{+/+}$ mice [277] were injected with $5\times10^{11}$ vg each followed by liver cell isolation by *in situ* digestion perfusion as described by Mederacke *et al*. [278]. Viable hepatocytes were collected by centrifugation at 50 × g for 2 min and pelleted cells were subjected to Percoll density gradient centrifugation. A custom FACS scheme was used to isolate the different liver cell types, including cholangiocytes (EPCAM+), hepatic stellate cells (HSCs; LRAT-RFP+), endothelial cells (CD31+) and macrophages (F4/80+), by Pervinder Choksi, Jaejun Kim and Simone Kurial. A manuscript describing liver cell isolation in detail is in preparation (Choksi *et al*.). To study promoters in human and mouse hepatocytes *in vivo* (ii), an FRGN (Fah$^{-/-}$;Rag2$^{-/-}$;Il2rg$^{-/-}$;Sirpa$^{NOD/NOD}$) mouse [279] transplanted with primary adult human

hepatocytes was injected with $1\times10^{12}$ vg of the AAV9 promoter mini-library (containing barcoded *eYFP* constructs with CMV1, LP1, SPc5-12, GFAP, gfaABC1D and GFAP_del1-5 promoters). Two weeks after injection, viable hepatocytes were collected using Percoll density centrifugation. Pelleted hepatocytes were stained with antibodies for NPC markers (CD45-BV421, CD31-PECy7, PDGFRB-PECy7) to exclude contaminating NPCs and human-specific antibodies to separate human hepatocytes from mouse hepatocytes (mH-2Kd-BV711, hB2M-PE). Stained cells were analyzed on FACSAria II. The study of a CMV promoter-driven AAV6 peptide display library (iii) was performed by systemic injection of $3\times10^{11}$ vg of the primary AAV6 peptide library (see section 2.2.2.7) into a 10-week-old *Lrat-Cre*$^{+/-}$*;R26-RFP*$^{+/+}$ mouse. Liver perfusion was performed as above one week after injection. Digested liver cells were centrifuged at 50 × g for 2 mins to collect hepatocytes. Percoll centrifugation was used to enrich viable cells in the hepatocyte pellet. The supernatant was centrifuged at 580 × g for 10 mins to collect bulk NPCs. For screening of the secondary AAV6 peptide library (CMV-cap6-536), $3\times10^{11}$ vg of the library were injected into a 11-week-old C57BL/6 mouse. After one week, liver digestion was performed as described above, followed by hepatocyte isolation *via* Percoll density centrifugation. Other cell types (cholangiocytes, hepatic stellate cells, endothelial cells, and macrophages) were isolated using the custom FACS scheme described above.

### 2.2.6 High-throughput sequencing

High-throughput/Next generation sequencing (NGS) was performed for readout of either barcode sequences in barcoded *eYFP* transgenes of capsid or promoter libraries, or for the readout of peptide insertion sites in the AAV6 peptide display setting. In all cases, Illumina sequencing was applied for amplicons spanning the respective region of interest (*i.e*., the barcode or peptide insertion regions).

#### *2.2.6.1 AmpliconEZ platform (MiSeq)*

I used the AmpliconEZ sequencing service from Azenta Life Sciences/GENEWIZ for barcode sequencing of promoter mini-libraries (section 3.2.1.2) and for peptide sequencing of the cap6-536 libraries (sections 3.2.2.2 and 3.2.3.3). For barcode sequencing, the barcode region was amplified from either virus library input, gDNA or cDNA using the BC_AEZ_f and BC_AEZ_r primers. For sequencing of AAV6 peptide display insert regions in projects 3.2.2.2 and 3.2.3.3 (2$^{nd}$ round libraries rep- and CMV-

cap6-536), the AAV6_AEZ_f and AAV6_AEZ_r primers were used. A detailed description of the DEPOOL application with AmpliconEZ sequencing is given in section 2.2.7.1.3.
PCR amplification was performed with the Q5 High-Fidelity 2X Master Mix (NEB) as described in section 2.2.1.2, using as template $10^7$ copies of virus or plasmid libraries, 50 ng gDNA, or cDNA from equivalents of 150 ng RNA input (reverse transcription as described in section 2.2.1.13). Primer annealing was performed at 60 °C with 15 sec elongation and 35 cycles of amplification (20 cycles for plasmid and virus libraries). For virus libraries, pre-denaturation was extended to 5 min. No-template (NTC) and no-RT control PCR reactions were performed to confirm absence of contamination. From each 50 µL PCR reaction, 2 µL of product were used for testing by agarose gel electrophoresis. Amplicons were purified using the PCR purification kit (Qiagen) and quantified with the Qubit 2.0 Fluorometer and dsDNA HS Assay kit (Thermo Fisher Scientific) prior to sample submission and sequencing with GENEWIZ AmpliconEZ platform.

#### *2.2.6.2 NextSeq: NuGEN platform library preparation*

I prepared sequencing samples for of the DEPOOL validation library (section 2.2.4.6) by PCR amplification of barcode regions from virus, gDNA and cDNA samples with the Phusion HS II as described in section 2.2.1.2 using the BC_NGS_f and BC_NGS_r primers (60 °C annealing, 10 sec elongation, 40 cycles). Subsequent Illumina library preparation was conducted with the Ovation Low Complexity Library Preparation Kit (NuGEN/Tecan) following the manufacturer's instructions for ligating Illumina adaptors and indices to the respective amplicons. Intermediate and final purifications were performed *via* the QIAquick PCR purification kit (Qiagen). Correct amplification and adaptor ligation were confirmed with the 2100 Bioanalyzer (Agilent Technologies). Adaptor-ligated amplicons were amplified with 12 cycles as recommended. Product DNA concentrations were then measured with the Qubit dsDNA HS Assay kit. Up to 32 indexed samples (each with specific adaptor indices) were pooled equimolarly, followed again by quality and quantity assessment with the Bioanalyzer and Qubit systems. NGS libraries were then submitted to the EMBL Genomics Core facility (Heidelberg) for NextSeq 500 sequencing (Illumina), using the option of PhiX spike-in for low complexity samples.

#### *2.2.6.3 NextSeq: Nextera platform library preparation*

For NGS analysis of the AAV6 peptide display library round 1 selection data (section 3.2.3.2), library preparation was performed with a two-step PCR pipeline instead of

adaptor ligation: in the first PCR, the region of interest was amplified using a mix of off-set primers containing a degenerate region with 1-7 random nucleotides in 5' of the primer binding site (only for antisense primers) as well as overhangs for Illumina adaptors. The primers s_Cap6_1703 and a_Cap6_1802 (designed by Teng Wei Koay) were used for amplification of the peptide insertion region in the AAV6 peptide library. This PCR was performed using the Q5 High-Fidelity 2X Master Mix (NEB) as described in section 2.2.1.3.1, using as template $10^7$ copies of virus or plasmid libraries, 25-150 ng gDNA or cDNA from equivalents of 30-150 ng RNA input (depending on yield from DNA/RNA extraction). For primer annealing, a temperature of 62 °C was selected, while amplification was performed for 15 sec. Plasmid and virus libraries were amplified for 20 cycles, while gDNA and cDNA samples were amplified for 35-40 cycles. Specific amplification was confirmed with agarose gel electrophoresis, followed by bead-based amplicon purification using the ProNex Size-Selective Purification System (Promega) according to the manufacturer's instructions. Illumina indices and adaptors were added in a secondary PCR using the Nextera XT Index Kit v2 (Illumina). Here, KAPA HiFi HotStart ReadyMix PCR Kit (Roche) was used for amplification with sample-specific indices over 8 cycles of amplification according to the Nextera kit instructions. Next, amplicons were purified using the ProNex bead purification before performing quality and quantity assessment *via* Bioanalyzer and Qubit, respectively. Samples were pooled equimolarly before submission to the EMBL Genomics Core facility for NextSeq 2000 sequencing (single-end, with PhiX spike-in).

### 2.2.7 DEPOOL

The depletion of off-targeting AAVs from on-target libraries (DEPOOL) strategy was developed and applied during my Master thesis under the supervision of Dirk Grimm and Julia Fakhiri [275]. In the current study, I validated the results of the initial screen and extended its application as described below.

#### *2.2.7.1 DEPOOL: Targeting of peptide insertion sites in cap*

##### 2.2.7.1.1 EnGen sgRNA synthesis for targeting of peptide insertion sites

I performed *in vitro* depletion of peptide insertion targets using sgRNAs generated *via* the EnGen sgRNA Synthesis kit (NEB). Spacers were designed to match peptide insertion sequences (complementary to the sense strand) within the *cap6* peptide display library (cap6-536; see section 2.2.2.7 and 2.2.2.8). Spacer-specific oligonucleotides were

designed following the EnGen guidelines: Each oligonucleotide contains a T7 promoter sequence, followed by 1×G, 21 nt of antisense peptide target sequence ($N_{21}$) and an overhang matching the SpCas9 scaffold sequence (5'- TTCTAATACGACTCACTATA G $N_{21}$ GTTTTAGAGCTAGA -3'). These were designed for nine target peptide insertion variants within cap6-536 and are listed in Table 25.

Synthesis of sgRNAs *via* annealing of spacer-specific oligonucleotides to a scaffold-encoding reverse oligonucleotide, second strand synthesis, T7-based sgRNA synthesis from the resulting dsDNA and subsequent DNA removal was conducted following the manufacturer's instructions of the EnGen kit. This was performed using either cH001 or cH002 oligonucleotides separately, or a mix of all nine spacer oligonucleotides ("9×mix"). Synthesized sgRNAs were purified using the RNA Clean & Concentrator-5 kit (Zymo Research) and quantified using the Nanodrop 2000 system.

**Table 25: DNA oligonucleotides used for EnGen synthesis of sgRNAs targeting variants within the cap-536 library.**

| sgRNA name | DNA oligonucleotide sequence |
|---|---|
| sgRNA_cH001 | TTCTAATACGACTCACTATA G GCAAACATCATCACCTTCGCA GTTTTAGAGCTAGA |
| sgRNA_cH002 | TTCTAATACGACTCACTATA G ATCGGTAGATTCATCTTTACC GTTTTAGAGCTAGA |
| sgRNA_cH003 | TTCTAATACGACTCACTATA G TTCTGCATCCTGTTTACCAGA GTTTTAGAGCTAGA |
| sgRNA_gN001 | TTCTAATACGACTCACTATA G TTTATCAGAACCACCTTCTTC GTTTTAGAGCTAGA |
| sgRNA_gN002 | TTCTAATACGACTCACTATA G ACCACCTTCACCATCATCGTT GTTTTAGAGCTAGA |
| sgRNA_gN003 | TTCTAATACGACTCACTATA G ATCGTTACCCTGACCTGGTTC GTTTTAGAGCTAGA |
| sgRNA_cN004 | TTCTAATACGACTCACTATA G ATCACCATCTGCATCTTTACC GTTTTAGAGCTAGA |
| sgRNA_cN006 | TTCTAATACGACTCACTATA G AACATCTTTACCGTTTTCTTC GTTTTAGAGCTAGA |
| sgRNA_cN008 | TTCTAATACGACTCACTATA G AGAATCATCACCTTCTGCGGT GTTTTAGAGCTAGA |

#### 2.2.7.1.2 *In vitro* cleavage of single plasmid target

The functionality of sgRNAs synthesized with the EnGen kit was tested for the HindIII-linearized cH001 plasmid target (CMV_cap6-cH001) using *in vitro* cleavage assay with a reaction volume of 30 µL, containing 30 nM SpCas9 (NEB), 30 nM sgRNA (calculated with a total sgRNA length of 102 nt) and 3 nM plasmid target DNA in NEB 3.1 buffer (volume adjusted with nuclease-free $H_2O$). Prior to addition of the DNA target (30 nM in 3 µL), the SpCas9/sgRNA mix (27 µL) was pre-incubated at rt for 25 min. After addition of plasmid DNA, the SpCas9 cleavage was performed at 37 °C for 2 h. Next, 1 µL of Proteinase K (Qiagen) was added to remove Cas9 protein, incubating at 50 °C for 30 min. The digestion mix was then combined with Purple Loading Dye (NEB) and analyzed *via* agarose gel electrophoresis.

2.2.7.1.3 *In vitro* depletion with AmpliconEZ readout

I adapted the protocol for Cas9-based depletion during PCR amplification from Hardigan *et al*. [280]. This protocol allows removal of variants from a library of DNA templates *via* Cas9-based cleavage with variant-specific sgRNAs. I used the CMV-cap6-536 plasmid library (section 2.2.2.8) as template pool with the EnGen-synthesized sgRNAs from above. For this protocol, $10^7$ copies of the plasmid library were amplified with 5 PCR cycles (Q5 High Fidelity 2x mix; 60 °C annealing, 15 sec elongation) using the AAV6_AEZ_f and AAV6_AEZ_r primers. Next, a pre-depletion mix was prepared containing 7.5 µg sgRNA (228 pmol; either cH001-sgRNA or 9x sgRNA mix), 5 µg SpCas9 (32 pmol; NEB), 2 µL NEB 3.1 buffer and filled to 10 µL with nuclease-free $H_2O$. A non-depletion control was performed containing no sgRNAs. The pre-depletion mix was incubated for 20 min at 37°C and then combined 1:1 (10 µL + 10 µL) with the product of the primary PCR. The resulting depletion mix was then incubated at 37 °C for 30 min. Next, 1 µL of 4 µg/µL RNase A (Qiagen) was added for removing sgRNAs. After incubation for 15 min at 37 °C, 1 µL of Proteinase K (NEB) was added for removal of enzymes, again incubating at 37 °C for 15 min. Enzymes were inactivated by incubation at 95 °C for 15 min. From this depletion mix, 10 µL were used as template for a second amplification PCR (50 µL total volume) with the AAV6_AEZ_f and AAV6_AEZ_r primers, with annealing at 60 °C and elongation for 15 sec. For this secondary PCR, 35 amplification cycles were performed. A positive control was conducted using 5 µL product of the primary PCR as template, as well as a non-template control. Of the resulting PCR products, 2 µL were used for testing with agarose gel electrophoresis, whereas the rest was purified (PCR purification kit) and submitted to AmpliconEZ sequencing (only for the no-sgRNA, cH001-sgRNA and 9x-sgRNA samples) as described in section 2.2.6.1.

### *2.2.7.2 DEPOOL2: Targeting of barcode sequences in the 3'-UTR of cap*

2.2.7.2.1 Cloning of rep-cap2-BC constructs

For the barcode-directed application of DEPOOL (section 3.2.2.3), the #0778 plasmid was modified by adding a barcode-acceptor site with a T7 promoter next to the cap-acceptor site. This was achieved by AscI/SpeI-HF restriction of #0778, followed by ligation of an oligonucleotide annealed from D2_T7_cassette_fw and D2_T7_cassette_rev primers to generate the D2_BC_acc plasmid. Three test barcode sequences (BC1, BC2, BC3) were then cloned by golden gate assembly using the D2_BC_acc plasmid, BsmBI-v2 enzyme (NEB) and annealed oligonucleotides (D2_BC1-3_fw/_rev). For each of the

resulting intermediate plasmids, *cap2* was inserted as described in section 2.2.2.6.2. Insertion of *cap2* yielded the rep-cap2-BC1/-BC2/-BC3 plasmids.

#### 2.2.7.2.2 Cloning and synthesis of barcode-directed sgRNAs

Templates for sgRNA synthesis were derived from the rep-cap-BC constructs *via* a multi-step cloning procedure depicted in Figure 5. Cloning and subsequent sgRNA synthesis were conducted jointly with Svenja Beenders (internship student). Expression cassettes of sgRNAs require the following components: T7 RNA polymerase promoter, spacer (20 nt sequence matching the target barcode) and SpCas9 scaffold. These were cloned from PAM-BC-T7 templates within rep-cap-BC constructs over six steps (i-vi): (i) the respective barcode regions of rep-cap2-BC1 /-BC2 /-BC3 were amplified from 10 ng plasmid template using D2_BC_fw and Lseq_rev primers (PCR with Phusion HS II as described in section 2.2.1.2; 60 °C annealing, 15 sec elongation, 40 cycles). PCR products were purified using the QIAquick PCR Purification kit. (ii) Next, 1 µg of purified amplicon from each barcode was digested with BbsI-HF in 30 µL reaction volume with CutSmart buffer by incubation at 37 °C for 16 h, followed by inactivation at 65 °C for 20 min. Overhangs of the BbsI-HF-digested DNA fragment were then blunted by adding 1 µL Klenow enzyme (DNA Polymerase I, Large (Klenow) Fragment; NEB) as well as 0.4 µL dNTP mix (Deoxynucleotide (dNTP) Solution Mix; NEB) and filling to 40 µL with nuclease-free $H_2O$. Klenow incubation was performed at 25 °C for 15 min, followed by inactivation by adding EDTA to 10 mM and incubating at 75 °C for 20 min. After again performing purification *via* the QIAquick PCR Purification kit, the DNA fragment in the eluate was treated with Antarctic Phosphatase (NEB; 30 min at 37 °C, inactivated at 80 °C for 2 min) for dephosphorylation. In step (iii), the SpCas9 sgRNA scaffold was amplified from #1514 plasmid using the scaff_fw (phosphorylated) and scaff_rev (not phosphorylated) primers (PCR conditions as in (i)), followed by PCR purification. (iv) The blunted barcode fragment (from (ii)) was ligated to the scaffold amplicon (from (iii)) in a 1:1 molar ratio, using 2 µL T4 ligase (NEB) in a reaction volume of 20 µL and incubating at 16 °C overnight. (v) After ligation, the correct product was enriched by PCR using the Lseq_fw and scaff_rev primers with 10 µL ligation mix as template (Phusion HS II, 60 °C annealing, 15 sec elongation, 35 cycles). The resulting amplicons were purified using the QIAquick PCR Purification kit. (vi) Synthesis of sgRNAs was performed using 1 µg of the purified amplicon from (v) with the AmpliScribe T7 Flash Transcription Kit (Biozym Scientific)

according to the manufacturer's instructions followed by sgRNA purification with the RNA Clean & Concentrator-5 kit.

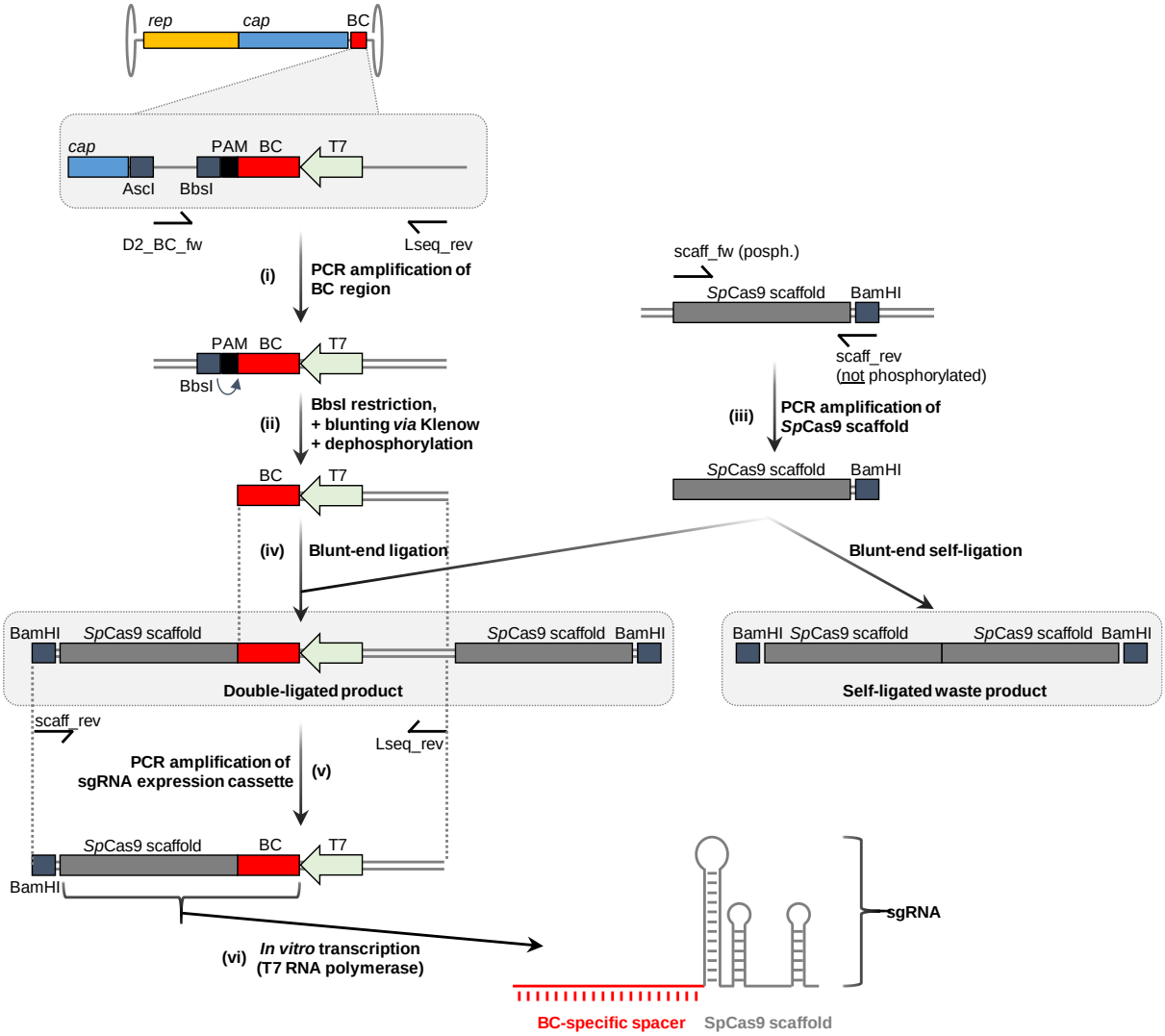



**Figure 5: Cloning procedure for sgRNA templates for the barcode-directed DEPOOL application.**

### 2.2.7.2.3 *In vitro* depletion of barcode targets

I performed *in vitro* depletion of individual barcode variants on rep-cap2-BC targets using the BC1-, BC2- or BC-3 sgRNAs from above. The depletion protocol described in section 2.2.7.1.3 was applied, using depletion combined with PCR amplification. For barcode-directed applications, the applied PCR was designed to imitate a *cap* rescue PCR instead of an NGS PCR. This rescue PCR was performed from gDNA extracted from Hek293T cells that had been transduced with either single barcode variants (rep-cap2-BC1/-BC2/-BC3) or pooled variants (Mini-Lib L1 and L2). Transduction was performed using an MOI

of $10^4$, harvesting at day 3 post transduction for gDNA extraction with the DNeasy Blood & Tissue kit. As in section 2.2.7.1.3, a primary PCR was performed with 5 cycles (using 100 ng gDNA as template and the Lseq_fw/ Lseq_rev primer set; 60 °C annealing, 1:20 min elongation), followed by mixing of PCR product to a pre-incubated depletion mix (this time using BC1-, BC2- or BC-3 sgRNAs from section 2.2.7.2.2), incubation for Cas9 cleavage, inactivation *via* RNase A and proteinase K, and a secondary PCR with 30 amplification cycles. Agarose gel electrophoresis was performed to assess PCR amplification/depletion. For the library samples L1 and L2, depleted and non-depleted (no sgRNA) amplicons were digested with PacI/SpeI-HF and cloned into the #0778 backbone. Individual colonies from each sample were picked, amplified with cap2_f and Lseq_rev primers and submitted for Sanger sequencing (cap2_f primer) for barcode readout.

### 2.2.8 Bioinformatics and statistical analyses

#### *2.2.8.1 NGS analysis and data normalization*

NGS results received for individual samples from GENEWIZ (AmpliconEZ) and de-multiplexed fastq files received from the EMBL Genomics Core facility (NextSeq) were analyzed with Python 3.6. I performed NGS data analysis with two custom scripts (written and/or modified by Josefine Sippel, Jonas Weinmann, Sabrina Weis, Olena Maiakovska, Mischa Schwendy, Kleopatra Rapti, and myself) as described in detail in Rapti *et al*. [281]. For libraries containing pre-defined variants (*i.e*., barcode libraries and the cap6-536 peptide libraries), the first script (“barcode detection”) searches for defined flanking sequences upstream and downstream of the region of interest, extracts the respective barcode or peptide sequence and assigns it to its respective capsid or promoter variant name as defined by a look-up table. For each sample, this generates a list of variant frequencies. Script 2 (“barcode analysis”) then allows the normalization of variant frequencies to the input library composition (*i.e.*, output of script 1 for the input virus library; not performed for promoter libraries) as well as an additional normalization to the vector abundance (Gβ [vg/dg] as measured by ddPCR) or relative expression (Cβ [$2^{-\Delta Ct}$] as measured by RTqPCR) in each sample.

For analysis of the initial AAV6 peptide library (CMV-cap6-P7P), the first script searches for defined flanking sequences upstream and downstream of the peptide insertion site. Next, correct peptide sequences (*i.e*., correct length and flanking sequences) are extracted and counted, giving the frequency of each variant. For each sample, script 2

then translates the sequences from nucleotides to amino acids and ranks them according to their frequency. I then matched and compared peptide variant frequencies across different samples using the Python Pandas 1.1.3 package. Plotting of these frequencies was performed using the matplotlib 3.3.2 and seaborn 0.11.0 packages.

#### *2.2.8.2 Promoter library datasets: Efficiency and Specificity*

The primary promoter library dataset was generated and analyzed by Claire Domenger. For this library, 53 barcoded promoter-eYFP constructs $\alpha_i$ had been packaged into the AAV9 capsid and screened in six female C57BL/6 mice, analyzing the following parameters across 16 major tissue types $\beta_i$: Cβ – bulk expression as measured *via* RTqPCR [$2^{-\Delta Ct}$]; Gβ – bulk distribution of vector genomes as measured *via* ddPCR [vg/dg]; Pαβ(gDNA) – relative frequency of each variant α in gDNA of each sample β quantified by barcode NGS; Pαβ(cDNA) – relative frequency of each variant α in cDNA of each sample β as also quantified by barcode NGS. The analysis pipeline for this dataset included NGS analysis as described above, followed by normalization of the cDNA read frequency Pαβ(cDNA) to the gDNA frequency Pαβ(gDNA) of the same variant in the same sample, yielding a normalized read frequency Rαβ for each promoter variant α in each tissue β. Next, analysis of normalized read frequencies was split into efficiency and specificity scores: the efficiency score E gives the percentual normalized expression in each tissue and thus allows the comparison of all promoters within a single tissue. It is calculated for each tissue β by normalizing individual Rα to the sum over all variants $\Sigma_\alpha R\alpha$ within this tissue.

$$E\alpha(\beta) = \frac{100 \times R\alpha(\beta)}{\sum_\alpha R\alpha\,(\beta)}$$

The specificity score S, on the other hand, allows the comparison of expression of one promoter variant α across all examined tissues. It is calculated for each variant α by normalizing individual Rβ values to the sum over all tissues $\Sigma_\beta R\beta$:

$$S\beta(\alpha) = \frac{100 \times R\beta(\alpha)}{\sum_\beta R\beta\,(\alpha)}$$

I contributed to the final promoter library analysis by assessing different normalization strategies as depicted in Figure 19: for comparisons between different tissues (*i.e.*, specificity scores), weighting of individual Rαβ values by bulk measurements in each tissue can account for differential transduction across these tissues. Thus, I tested different weighting factors ω for scaling of the Rαβ: for these factors ω, the relative expression (Cβ), vector distribution (Gβ) or normalized expression (Qβ = Cβ/Gβ) were

tested. After calculating Rαβ×ω, Eαβ and Sαβ were extracted for the subset of promoter variants that were also included in the validation experiment: CMV, LP1, SPc5-12 and GFAP. To be able to directly compare these values to the validation dataset (same subset of promoters tested individually as described in section 2.2.5), the Qαβ values calculated from the validation dataset were treated equally as Rαβ×ω from the library dataset, thus using them for establishing Eαβ and Sαβ as well. Specificity scores from library and validation datasets were then compared using mean values across replicate animals.
I utilized the same analysis pipeline for evaluating promoters within the promoter mini-libraries (section 3.2.1.2).

#### *2.2.8.3 Analysis of transcription factor binding sites within the GFAP promoter*

Binding sites of transcription factors active in or specific for hepatocytes[282-287] were predicted for the GFAP promoter sequence. Position frequency matrices (PFMs) for selected transcription factors from human and/or mouse origin (depending on availability) were obtained in MEME format from the JASPAR 2022 database [288]. Using the FIMO tool (version 5.5.3; meme-suite.org/meme/tools/fimo), the GFAP promoter sequence was scanned for matches with these PFMs (standard parameters; $p < 0.0001$).

#### *2.2.8.4 Statistical analyses*

Statistical analyses and plotting were conducted using the GraphPad Prism 8.0.1 software (GraphPad). Generally, one-way ANOVA was used for comparison of different groups of the same category, while two-way ANOVA was used for comparing groups of different categories. Multiple comparisons were corrected with Sidak's, Tukey's or Dunnett's multiple comparison tests as indicated. Pearson correlation coefficients between datasets were computed with either Prism or the Python NumPy 1.19.2 package (only for AAV6 peptide library *in vivo* analysis). Statistical significance is indicated with "ns", non significant; " * ", $p < 0.05$; " ** ", $p < 0.01$; " *** ", $p < 0.001$; and " **** ", $p < 0.0001$.

# 3 Results

## 3.1 Exploration and application of AAV knockdown tools

AAV vectors can be employed for the transfer of different effector modalities to induce targeted mRNA knockdown *in vitro* and *in vivo*. This allows the downregulation of endogenous or pathogen-derived target genes. To compare knockdown efficiency for two previously known effector types, I tested AAV-mediated silencing *via* the RNAi/shRNA and the CRISPR/CasRx systems. Both approaches were employed for targeted mRNA degradation of reporter and endogenous target genes. Additionally, shRNAs were investigated for their ability to counter or block infections with SARS-CoV-2 by AAV-induced targeting of viral genomic RNA.

### 3.1.1 Targeted mRNA knockdown using AAV-CasRx tools

As a class 2, type VI-D CRISPR system, Cas13d presents a programmable effector for targeted downregulation of RNAs. The *Ruminococcus flavefaciens* Cas13d ortholog (CasRx) is a highly efficient member of the Cas13d family [210]. The small size of CasRx protein (967 amino acids) and cDNA (2,901 bp) enables the generation of all-in-one AAV vectors packaging a CasRx expression cassette as well as a secondary cassette for expression of a guide (g)RNA or a multiplexed gRNA array.

#### *3.1.1.1 Reporter knockdown with EFS- and CMV-CasRx constructs*

To explore CasRx all-in-one AAV knockdown effectors, I cloned the nuclear localization signal (NLS)-flanked CasRx cDNA [210] into a single-stranded (ss)AAV vector along with a U6 promoter-driven gRNA cloning site flanked by two CasRx direct repeat (DR) sequences (Figure 6A). I designed CasRx expression cassettes in two configurations to ensure maximum expression while adhering to the AAV packaging limit of 4.9 kb. The cassette was created to either (i) use the EFS promoter (238 bp) and a bovine growth hormone polyadenylation signal (bgh-pA; 232 bp), or (ii) the CMV promoter (509 bp) and a minimal polyadenylation signal (mini-pA; 60 bp). To compare EFS- and CMV-CasRx effectors, I used dual luciferase knockdown assays with the psiCheck-2 reporter. This reporter plasmid expresses a *Renilla* luciferase (*Rluc*) targeted by CasRx gRNAs and a non-targeted Firefly luciferase (*Fluc*) for normalization. Delivery of reporter and effector constructs into Hek293T cells was achieved *via* Lipofectamine-based plasmid transfection or *via* AAV2-based transduction.

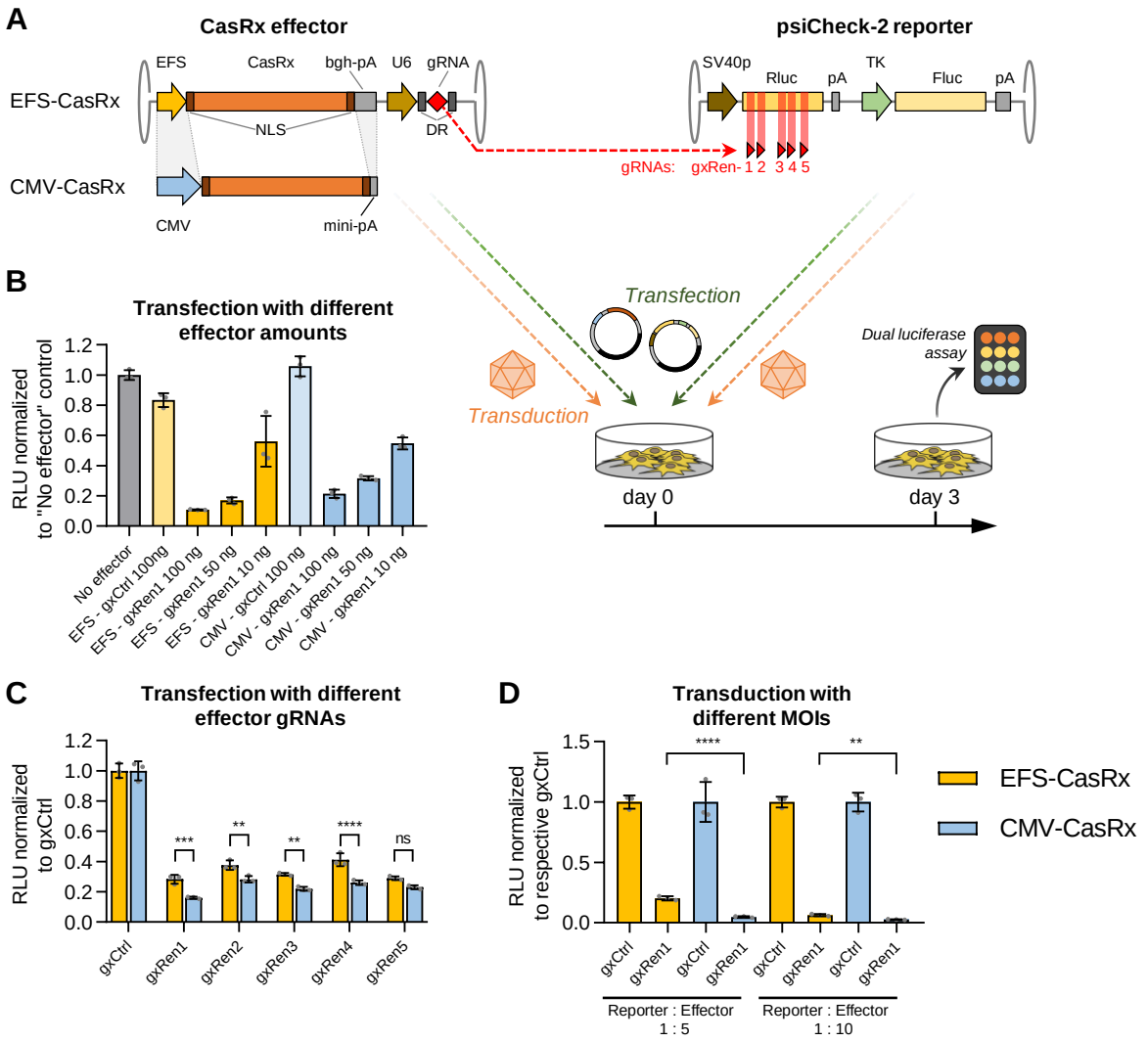



**Figure 6: EFS and CMV promoter-driven expression of AAV/CasRx effectors for psiCheck-2 reporter knockdown. (A)** CasRx-expressing single-stranded (ss)AAV constructs "EFS-CasRx" and "CMV-CasRx" were constructed using EFS or CMV promoter-driven expression of CasRx with flanking nuclear localization signals (NLS), and bovine growth hormone (bgh-pA) or minimal polyadenylation signals (mini-pA). Guide (g)RNAs were expressed from a U6-promoter-driven expression cassette in the same vector. Expression of *Renilla* luciferase (*Rluc*) from the psiCheck-2 reporter was targeted by CasRx gRNAs gxRen1-5. Effector and reporter constructs were co-transfected into Hek293T cells *via* Lipofectamine transfection or co-transduced after packaging into AAV2 vectors. Each condition was prepared in triplicates. Cells were lysed for dual luciferase assays on day 3 post-treatment. **(B)** Transfection of Hek293T cells with 10 ng psiCheck-2 reporter and different amounts of EFS- or CMV-CasRx with non-targeting gxCtrl or *Rluc*-targeting gxRen1 gRNA. Relative luminescence units (RLU) were normalized to the "No effector" control. **(C)** Transfection with 10 ng psiCheck-2 and 100 ng EFS- or CMV-CasRx constructs with gxCtrl or gxRen1-5. RLUs were normalized to gxCtrl of each effector type. **(D)** Co-transduction of the AAV/psiCheck-2 reporter (MOI of $10^4$) with EFS- or CMV-CasRx vectors expressing gxCtrl or gxRen1 using reporter:effector ratios of 1:5 or 1:10. Values are represented as mean ± standard deviation (SD). Statistical analyses of knockdown efficiency between EFS- and CMV-CasRx-gxRen1 were conducted using one-way ANOVA with Sidak's multiple comparison test.

Lipofectamine transfection of Hek293T cells with 10 ng psiCheck-2 plasmid and 10, 50 or 100 ng of effector plasmid demonstrated a dose-dependent knockdown for both EFS- and CMV-CasRx with the *Rluc*-targeting gxRen1 gRNA compared to a non-targeting gxCtrl (Figure 6B). Transfection of EFS- or CMV-CasRx effectors with gRNAs gxRen1-5 showed a significantly stronger knockdown efficiency for all gRNAs except gxRen5 when using the CMV-CasRx expression cassette (Figure 6C). CMV-CasRx also yielded a superior knockdown efficiency after co-transduction with the psiCheck-2 reporter using AAV2 vectors. For two different reporter:effector ratios (1:5 and 1:10; with a multiplicity of infection (MOI) of $10^4$ for reporter), CMV-CasRx-gxRen1 induced a significantly stronger knockdown as compared to EFS-CasRx (Figure 6D). Therefore, I selected the CMV-CasRx setting for further experimental evaluation.

Since CasRx is capable of processing a CRISPR array into functional individual gRNAs [210], multiplexing of gRNAs is feasible within a single RNA polymerase III-dependent expression cassette. I tested cloning of such arrays into the U6 cassette of CMV-CasRx for up to four gRNAs using ligation of multiple oligonucleotides into the BsmBI-digested gRNA acceptor (Figure 7A). To assess the knockdown from different multiplexing settings and array positions, a single targeting spacer (gxRen1 = “X”) was combined with one or more non-targeting spacers (gxCtrl = “C”) in different configurations. Lipofectamine transfection (Figure 7B) and AAV2 transduction (Figure 7C) of Hek293T cells demonstrated significant knockdown compared to a non-targeting control construct for gxRen1 (X) expressed within 1- to 3-fold multiplexed arrays. Multiplexing of X in a 4-fold array setting did not reduce the expression of the Rluc reporter for all three tested configurations (X-C-C-C, C-X-C-C and C-C-C-X) in either a transfection or a transduction setting. Thus, multiplexing with a maximum of three gRNAs per array was demonstrated to be feasible within the all-in-one vector setting of CMV-CasRx.

While both transfection and transduction settings demonstrated significant knockdown for either single or multiplexed gRNAs, transduction assays repeatedly showed stronger knockdown rates as compared to Lipofectamine transfection (Figure 6 and Figure 7). This possibly indicates a higher level of CasRx construct present within each cell for the transduction setting.

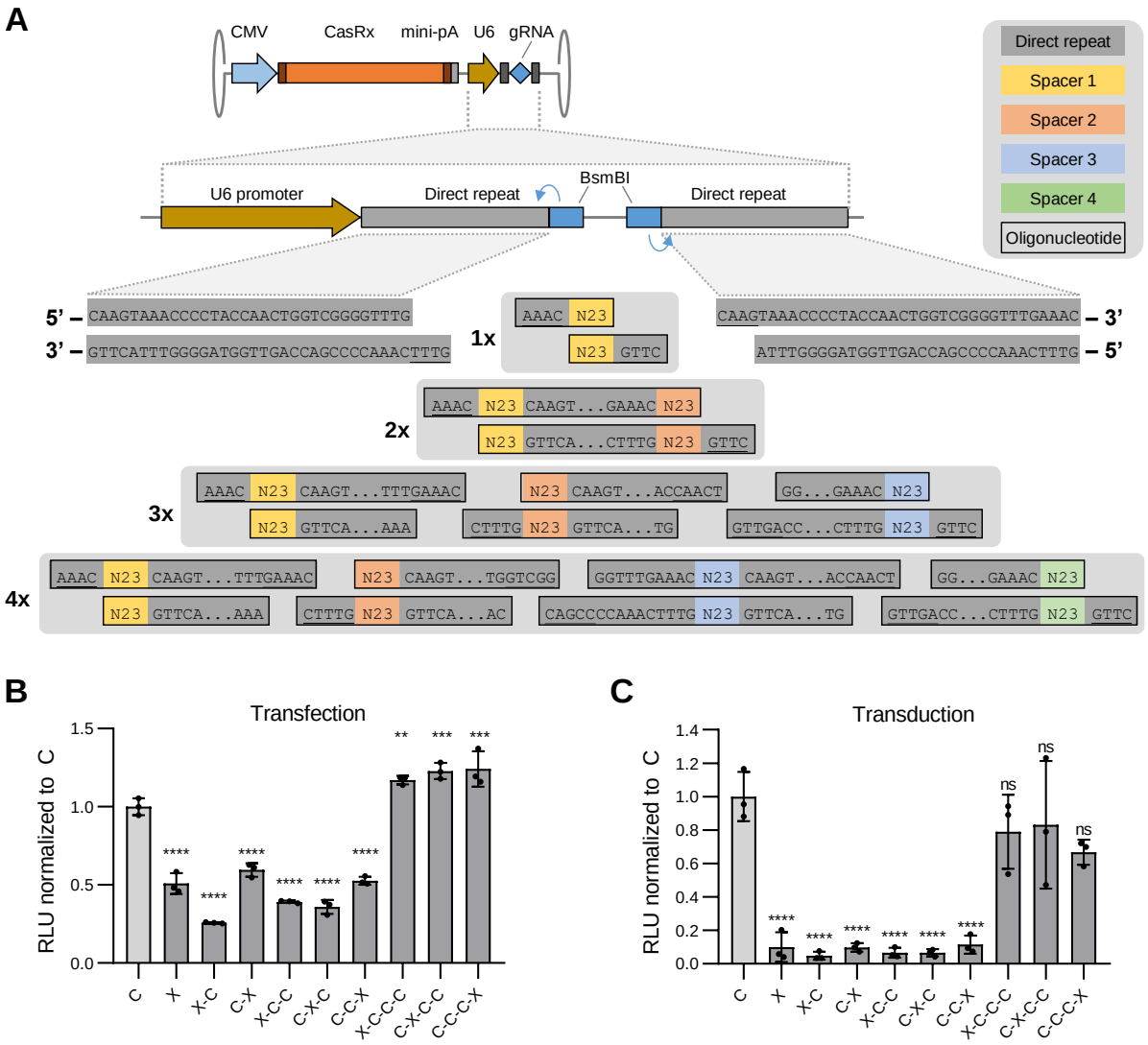



**Figure 7: CasRx gRNAs can be functionally multiplexed into AAV vectors for up to three spacers per array. (A)** CasRx gRNAs were multiplexed within the U6 cassette present in the CMV-CasRx vector construct. Multiplexing is achieved by ligation of sets of one (for 1x or 2x spacers), three (3x spacers) or four (4x spacers) double-stranded oligonucleotides into the BsmBI-digested gRNA acceptor site. Oligonucleotides are designed to have distinct overhangs within the direct repeat sequence in order to define the position of the respective spacer. **(B-C)** Knockdown of *Rluc* in the psiCheck-2 reporter was tested with CMV-CasRx containing arrays of one to four gRNAs. C is non-targeting gxCtrl, X is Rluc-targeting gxRen1. Reporter and effector constructs were delivered into Hek293T cells using Lipofectamine plasmid transfection (**B**) or transduction *via* AAV2 vectors (**C**). Dual luciferase assays were conducted on day 3 post-treatment. Individual values of triplicate measurements are presented with mean ± SD. Statistical analysis was performed using one-way ANOVA with Dunnett's multiple comparison test.

### *3.1.1.2 miR-122 binding sites in 3'-UTR allow off-switching of CasRx knockdown*

Binding sites for micro (mi)RNAs can be incorporated into the 3'-UTR of vector expression cassettes to control and regulate transgene expression. Introduction of binding sites for miR-122, for instance, enables an inactivation of transgene expression in miR-122-

expressing hepatocytes [185]. To introduce such cell-type-dependent regulation of CasRx knockdown and thus increase the selectivity of the knockdown vector, I cloned a miR-122-dependent off-switch. This was achieved by inserting two miR-122 binding sites into the 3'-UTR of the CasRx expression cassette in the CMV-CasRx construct (Figure 8A), thereby generating the CMV-CasRx_miR122-OFF construct.

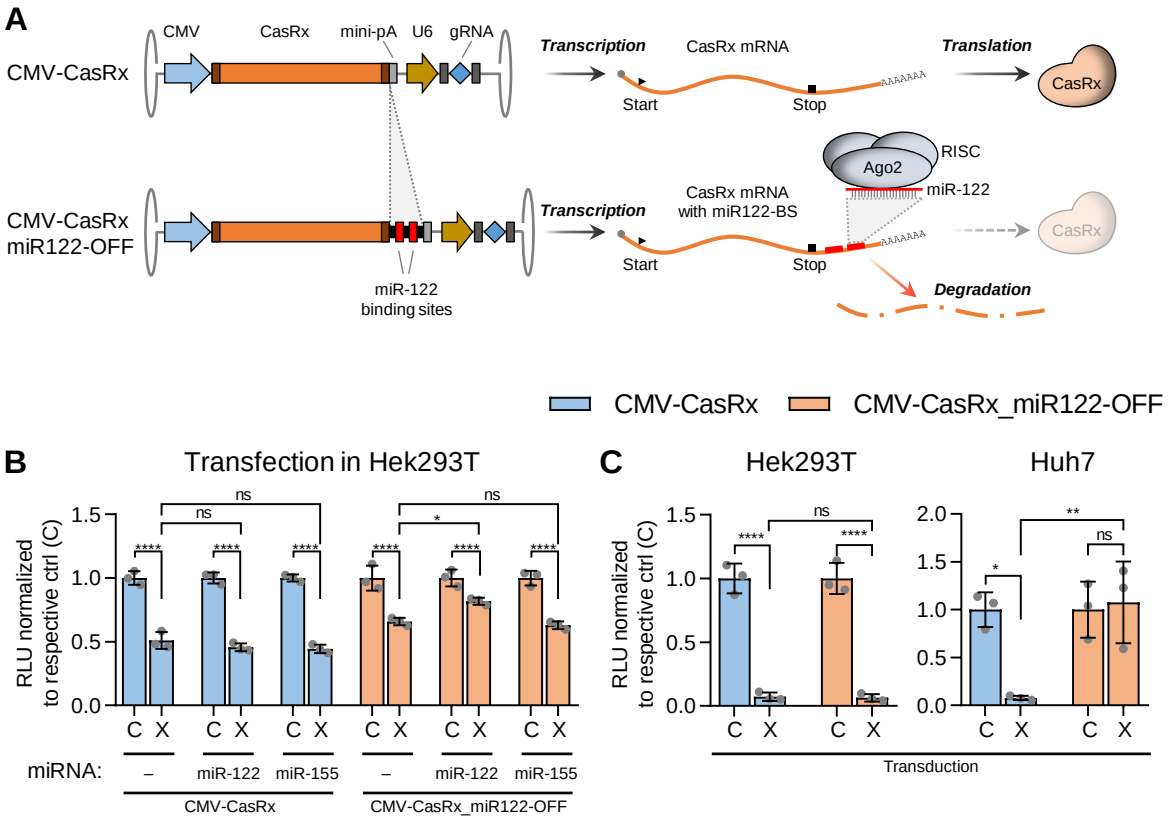



**Figure 8: Micro-RNA based off-switch for CasRx knockdown vectors. (A)** CMV-CasRx was modified by inserting miR-122 binding sites into the 3'-UTR of the CasRx expression cassette, yielding the CMV-CasRx_miR122-OFF construct. In cells expressing miR-122, the CasRx mRNA with miR-122 binding sites (miR122-BS) is degraded by the miR-122-RISC complex, leading to reduced knockdown efficiency due to lower levels of CasRx protein. **(B)** Hek293T cells were triple-transfected with psiCheck-2 reporter, CasRx effector (with or without miR-122 off-switch) and a miRNA expression plasmid (miR-122 or miR-155 control). **(C)** Hek293T cells (not expressing miR-122) and Huh7 cells (expressing miR-122) were transduced with AAV/psiCheck-2 reporter (MOI of $10^4$) and CasRx effector constructs (MOI of $10^5$). Dual luciferase assays were conducted three days post-treatment. Values are presented as mean ± SD. Statistical analysis was performed using two-way ANOVA with Tukey's multiple comparison test. C, non-targeting gxCtrl gRNA; X, Rluc-targeting gxRen1.

I compared knockdown through CMV-CasRx and CMV-CasRx_miR122-OFF constructs by Lipofectamine transfection and dual luciferase assays in Hek293T cells (Figure 8B). Non-targeting control gRNA "C" and *Rluc*-targeting gRNA "X" were cloned into both constructs. I co-transfected effectors with psiCheck-2 reporter plasmid as before and

expanded this setting by additionally supplying 25 ng of a miR-122-expressing plasmid to test miR-122-based downregulation. As a control setting, I tested transient expression of miR-155 instead of miR-122. This showed significant knockdown of gRNA “X” compared to the respective non-targeting gRNA “C” for all tested conditions. The addition of either miRNA expression plasmid did not significantly alter knockdown efficiency for the CMV-CasRx construct. As expected, addition of miR-122 but not miR-155 significantly reduced knockdown for CMV-CasRx-miR122-OFF with normalized RLU values of 0.66 for the condition with no miRNA expression (*i.e.*, 34% knockdown), 0.82 for miR-122 expression and 0.63 for miR-155 expression (18 and 37% knockdown, respectively).

To assess whether this miR-122-based reduction in knockdown efficiency would also be observed for the transduction setting, I again packaged CasRx effectors and psiCheck-2 reporter into AAV2 vectors (Figure 8C). Next, I performed transduction and dual luciferase assays in two different cell lines: Hek293T cells, which show very weak miR-122 expression, and Huh7 cells that exhibit strong miR-122 expression [289]. For Hek293T cells, both CMV-CasRx and CMV-CasRx_miR122-OFF produced strong knockdown with no significant difference between the two effector types. Strikingly, for Huh7 cells, only CMV-CasRx showed a reduction in RLU between the control C and *Rluc*-targeting X gRNAs while no knockdown was observed for CMV-CasRx_miR122-OFF, representing a significant difference between the two effector types in Huh7 cells. This demonstrates a functional miR-122-based regulation of knockdown efficiency for the CMV-CasRx_miR122-OFF construct and showcases the ability to control CasRx activity in a cell-type-specific manner.

#### *3.1.1.3 Knockdown of CD44 using CasRx and shRNA effectors*

In the next step, I compared the knockdown capacity of CasRx vectors to shRNA-based knockdown tools for targeting of either the *Rluc* reporter or an endogenous cellular target RNA. I tested this for the isoform-specific or general downregulation of CD44, a membrane glycoprotein involved in the progression of non-alcoholic (metabolic) steatohepatitis (Figure 9A) [260]. The murine *CD44* gene contains nine constant and ten variant exons. The standard isoform “*CD44s*” is expressed with all constant and no variant exons, while the *CD44v6* isoform contains the variant exon v6 (Figure 9B).

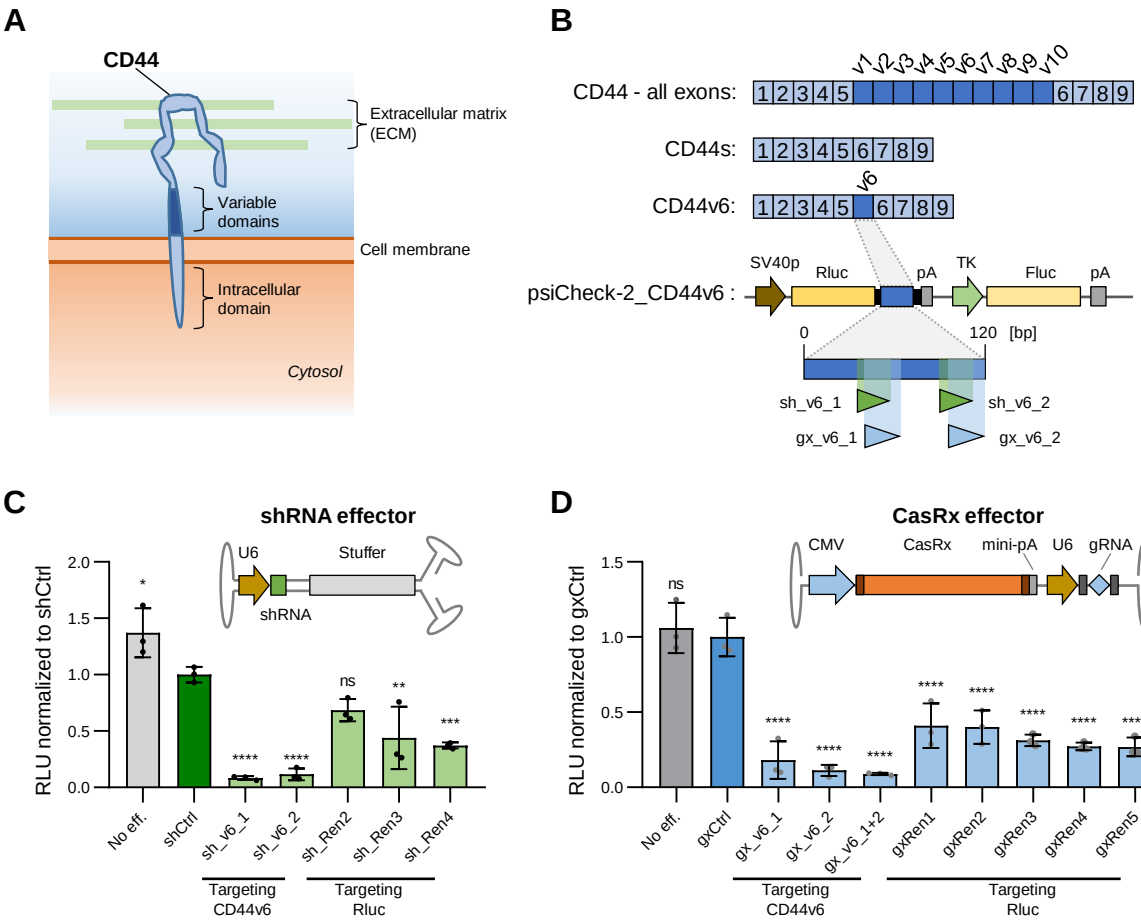



**Figure 9: Knockdown of *CD44v6* reporter with shRNAs or CasRx. (A)** CD44 is a glycoprotein anchored in the cell membrane and integrated into the extracellular matrix. **(B)** The cDNA of murine *CD44* contains nine constant exons (1-9) and ten variant exons (v1-v10). Standard isoform *CD44s* contains only exons 1-9, while the *CD44v6* isoform contains variant exon v6. The *v6* exon was cloned into the psiCheck-2 multiple cloning site to yield the psiCheck-2_CD44v6 reporter. This reporter allows testing of *CD44v6* knockdown in dual luciferase assays using shRNA effectors sh_v6_1 and sh_v6_2 or CasRx gRNA effectors gx_v6_1 and gx_v6_2. **(C)** Dual luciferase assay in Hek293T cells after plasmid co-transfection with psiCheck-2_CD44v6 reporter and either no effector ("No eff."), or shRNA effectors with non-targeting "shCtrl", *CD44v6*-targeting shRNAs sh_v6_1 and sh_v6_2, and *Renilla* luciferase (*Rluc*)-targeting shRNAs sh_Ren2, -3 and -4. RLU were normalized to the shCtrl condition. **(D)** Transfection and luciferase assay for CasRx effectors with non-targeting gxCtrl gRNA, *CD44v6*-targeting gx_v6_1, gx_v6_2 (single gRNA) and gx_v6_1+2 (double gRNA), or *Rluc*-targeting gxRen1-5. RLU values were normalized to respective non-targeting control and are shown as mean from triplicates ± SD. Statistical analysis was performed using one-way ANOVA with Dunnett's multiple comparison test.

I evaluated the downregulation of CD44 variant exon v6 specifically by creating the psiCheck-2_CD44v6 reporter, which contains the v6 exon sequence in the 3'-UTR of *Rluc* in psiCheck-2. To compare knockdown efficiencies, I employed both shRNA and CasRx effectors to directly target *Rluc* mRNA or to target the v6 sequence within the 3'-UTR of *Rluc* mRNA. In this transfection assay, both shRNA (Figure 9C) and CasRx effectors

(Figure 9D) demonstrated stronger knockdown with *CD44v6*-targeting effectors as compared to *Rluc*-targeting effectors. Apart from shRen2 (targeting *Rluc*), all effectors yielded significant knockdown compared to the respective non-targeting control.

To evaluate *knockdown* of an endogenous mRNA instead of a reporter, I targeted the murine *CD44s* mRNA directly (Figure 10A). Knockdown was performed by transducing *CD44*-expressing Hepa1-6 cells (mouse hepatoma cell line) with AAV-DJ vectors expressing either shRNA or CasRx effectors (MOI of $10^5$). Cells were harvested at day 3 post-transduction, followed by RNA extraction and RTqPCR-based quantification of *mCD44* expression compared to the *GAPDH* housekeeping reference gene (Figure 10B). For shRNA effectors, each scAAV vector expressed either a single shRNA from a U6 promoter or three different shRNAs from three different RNA polymerase III promoters (U6, H1 and 7SK) using the TRISPR format [274] (Figure 10C). For CasRx (Figure 10D), ssAAV vectors also expressed either a single gRNA or three gRNAs within the array structure introduced in Figure 7. For shRNA-based knockdown, a pre-screening of seven shRNA sequences targeting *CD44* was conducted jointly with Emma Gerstmann (B.Sc. student). This led to the selection of three shRNAs with strong *CD44* knockdown efficiency in Hepa1-6 cells. To increase knockdown efficiency even further, I then multiplexed these three shRNAs in the TRISPR format using two different configurations A and B. Configuration A uses the U6 promoter for expression of sh3, H1 promoter for sh2, and 7SK promoter for sh1. For configuration B, sh2 and sh3 are switched (sh2 is now under the U6 and sh3 under the H1 promoter). AAV-DJ transduction of Hepa1-6 cells with single- or triple-shRNA vectors demonstrated a significant knockdown of relative *CD44* expression for all effectors as compared to the non-targeting shCtrl vector (Figure 10E). Highest knockdown efficiency was achieved for triple-vector A (relative *CD44* expression of 0.10, *i.e*., 90% knockdown), followed closely by triple-vector B and single-shRNA vector sh1 (both giving 86% knockdown). For CasRx effectors, *CD44*-targeting gRNAs did not yield significant knockdown of *CD44* expression in Hepa1-6 cells (Figure 10F), with a mean expression of 83-99% as compared to the non-targeting gxCtrl (*i.e*., 1-17% knockdown).

In an attempt to improve CasRx-induced knockdown, I multiplexed the three strongest gRNAs (with relative *CD44* expression of 83% for gx1, 83% for gx2 and 87% for gx3; all non-significant) into a triple-gRNA array. Transduction with either of these single or the triple-gRNA CasRx effector again failed to induce significant knockdown (Figure 10G), with a mean *CD44* expression of 54%-61% compared to the non-targeting gxCtrl. Thus,

while showing promising knockdown rates in luciferase reporter assays, I did not observe significant knockdown of the endogenous *CD44* target mRNA with CasRx vectors.

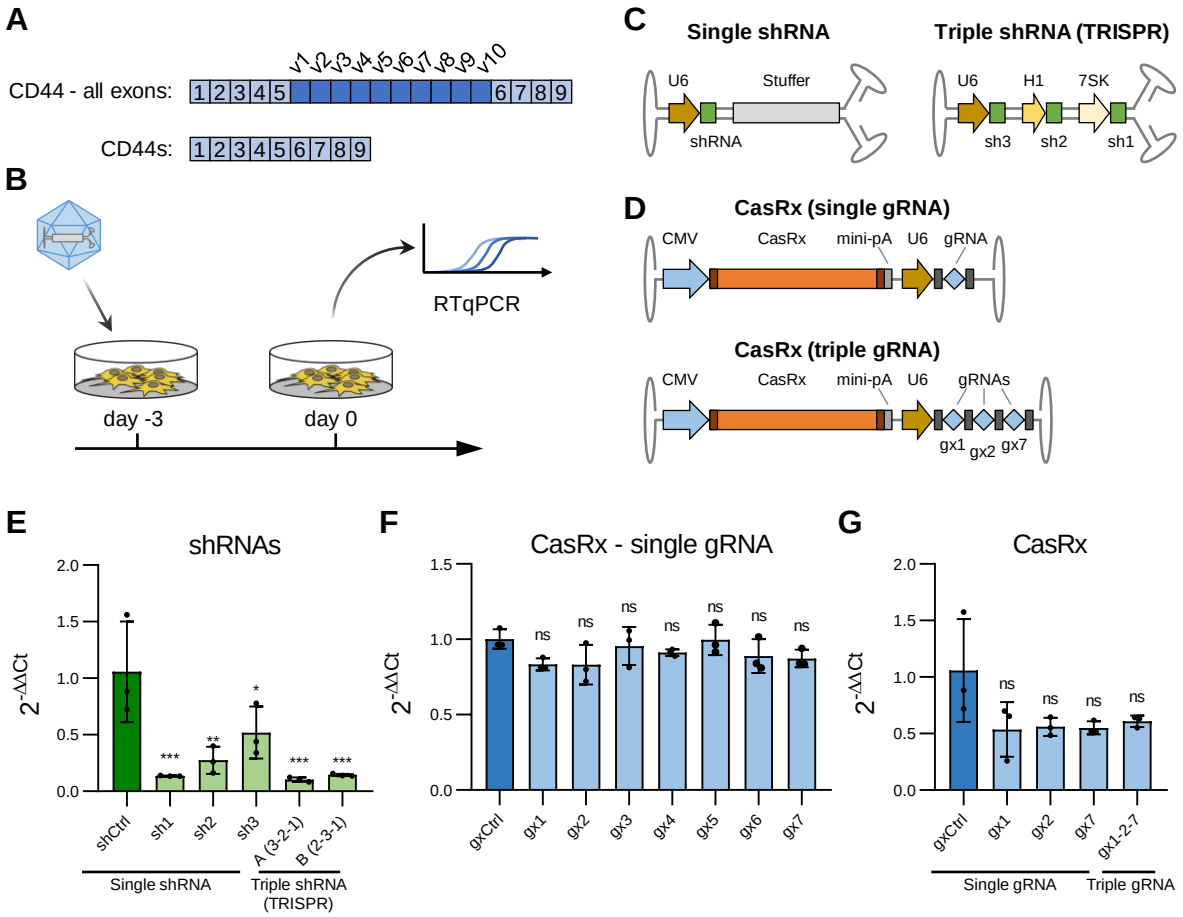



**Figure 10: Knockdown of *CD44* standard isoform is successful with shRNAs but fails with CasRx. (A)** The *CD44* standard isoform *CD44s* contains all constant exons and none of the variant exons v1-v9. Effector shRNAs and CasRx gRNAs were designed to bind *CD44s* mRNA within any of the constant exons. **(B)** Analysis of *CD44s* knockdown through transduction of Hepa1-6 cells with AAV-DJ vectors (three wells per condition), harvesting after three days and RTqPCR measurement of *CD44* expression compared to a *GAPDH* reference. **(C)** Effector shRNAs were cloned either as single- or triple-shRNA vectors. For triple-shRNA constructs, the TRISPR format was used with three shRNA expression cassettes comprising U6, H1 and 7SK promoters. **(D)** CasRx effectors were constructed with either a single gRNA or a triple-gRNA array. **(E)** Relative expression of *CD44* in Hepa1-6 cells after treatment with shRNA effectors. Values are normalized to the mean of non-targeting shCtrl. **(F)** Relative *CD44* expression after transduction with single-gRNA CasRx effectors. Values are normalized to the mean of non-targeting gxCtrl. **(G)** *CD44* expression after transduction with single- or triple-gRNA CasRx effectors. Values are normalized to gxCtrl. Bars represent mean values with error bars indicating standard deviation of triplicates. Statistical analysis was performed using one-way ANOVA with Dunnett's multiple comparison test.

### 3.1.2 Repressing SARS-CoV-2 infection *via* combinatorial AAV/shRNA vectors

Viral infection with SARS-CoV-2 can pose serious risks of immediate and long-term illness. As SARS-CoV-2 is an RNA virus, RNA-targeting effectors may prevent or inhibit infections. I explored this strategy using shRNA vectors in a preventive setting, by pre-treating cells with anti-SARS-CoV-2 shRNAs before infection with SARS-CoV-2. This project was conducted in collaboration with the group of Steeve Boulant, where SARS-CoV-2 infections and infection quantification were conducted by Megan Stanifer. The results presented in this chapter have been published together with additional data and a more comprehensive discussion in Molecular Therapy [290].

#### *3.1.2.1 Single-shRNA vectors for blocking SARS-CoV-2 infection*

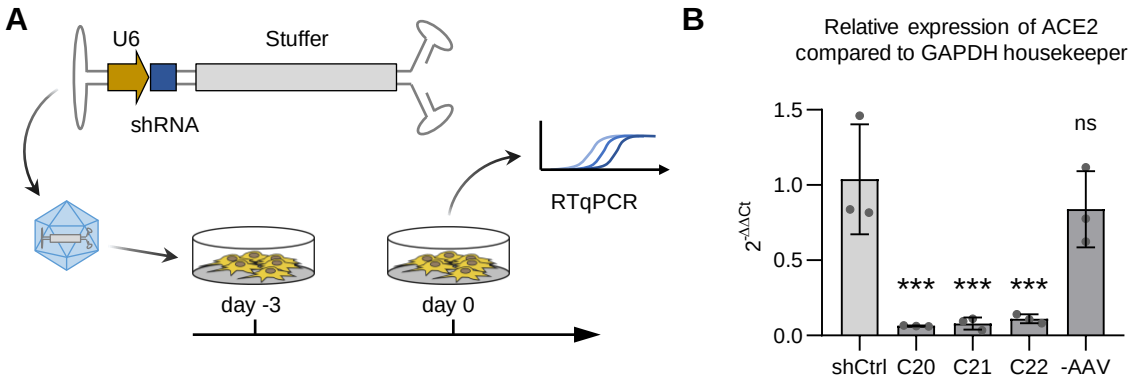



**Figure 11: Short hairpin (sh)RNA-induced knockdown of the SARS-CoV-2 receptor ACE2.** **(A)** Vero E6 cells were transduced with shRNA-expressing AAV-LK03 vectors with an MOI of $10^5$ vg per cell (n = 3 wells per condition). Cells were harvested three days later for quantification of the *ACE2* target and the *GAPDH* housekeeper expression *via* RTqPCR. **(B)** Relative expression of ACE2 compared to the *GAPDH* reference and normalized to the non-targeting shCtrl. C20, C21 and C22 are *ACE2*-targeting shRNA vectors. Bars represent mean normalized expression ± SD. Statistical analysis was conducted with one-way ANOVA and Dunnett's multiple comparison test. A modified version of this figure has been published [290].

As the conventional shRNA system for AAV-mediated RNA knockdown showed more robust knockdown results compared to CasRx (section 3.1.1.3), all following experiments were conducted with shRNA-based downregulation. Before targeting SARS-CoV-2 RNA directly, I tested the shRNA knockdown system again for silencing its entry receptor. The human (h)ACE2 cell surface protein is used by SARS-CoV-2 as a receptor during cell infection [245]. Thus, I tested knockdown of *hACE2* mRNA (Figure 11) in order prevent SARS-CoV-2 infection. Vero E6 cells were transduced with AAV-LK03 vectors (MOI of $10^5$) expressing a non-targeting shCtrl shRNA, or *ACE2*-targeting C20, C21 or C22

shRNAs. I harvested the cells three days post-transduction for RTqPCR-based analysis of ACE2 expression relative to a *GAPDH* housekeeping reference (Figure 11A). All three *ACE2*-targeting shRNAs yielded significant knockdown, with the mean relative ACE2 expression reduced to 6% - 11% of the non-targeting shCtrl condition (Figure 11B).

Since the shRNA system proved efficient for inducing *hACE2* knockdown in Vero E6 cells, the next step focused on targeting SARS-CoV-2 RNA directly (Figure 12). Based on a pre-selection of 19 SARS-CoV-2-targeting shRNAs using crude lysate AAV vectors (data of pre-selection not shown; details can be found in Becker *et al*. [290]), seven shRNAs binding within SARS-CoV-2 *RdRP* (C8, C9, C12) and *N* genes (C3, C5, C16, C17) were selected for production of purified AAV-LK03 vectors (Figure 12A). Transduction of Vero E6 cells with AAV-LK03 vectors was followed by infection with SARS-CoV-2 on day three post-treatment and infection analysis after another 24 h (Figure 12B). Infection rates of Vero E6 cells were quantified by Megan Stanifer, detecting double-stranded (ds)RNA as a surrogate for viral replication using an in-cell immunofluorescence (IF) assay (Figure 12C). As compared to non-transduced cells, a significant reduction in the number of infected cells was observed following pre-transduction with the non-targeting control shRNA vector (shCtrl; 72% of non-transduced). The reduction in infection was, however, much stronger with targeting shRNAs C3, C8, C9 and C12 (8%-13% relative to non-transduced). Pre-transduction with the *ACE2*-targeting C22 shRNA vector also resulted in reduced SARS-CoV-2 infection (33% compared to -AAV). Quantification of SARS-CoV-2 genome copy numbers compared to input (also conducted by Megan Stanifer) resulted in similar results (Figure 12D): Compared to non-transduced cells, a strong reduction in SARS-CoV-2 genomes was observed especially for C3, C8 and C12 shRNA vectors. Since these three shRNAs yielded the most efficient reduction in SARS-CoV-2 infection, I selected them for subsequent multiplexing into a combinatorial triple-shRNA vector.

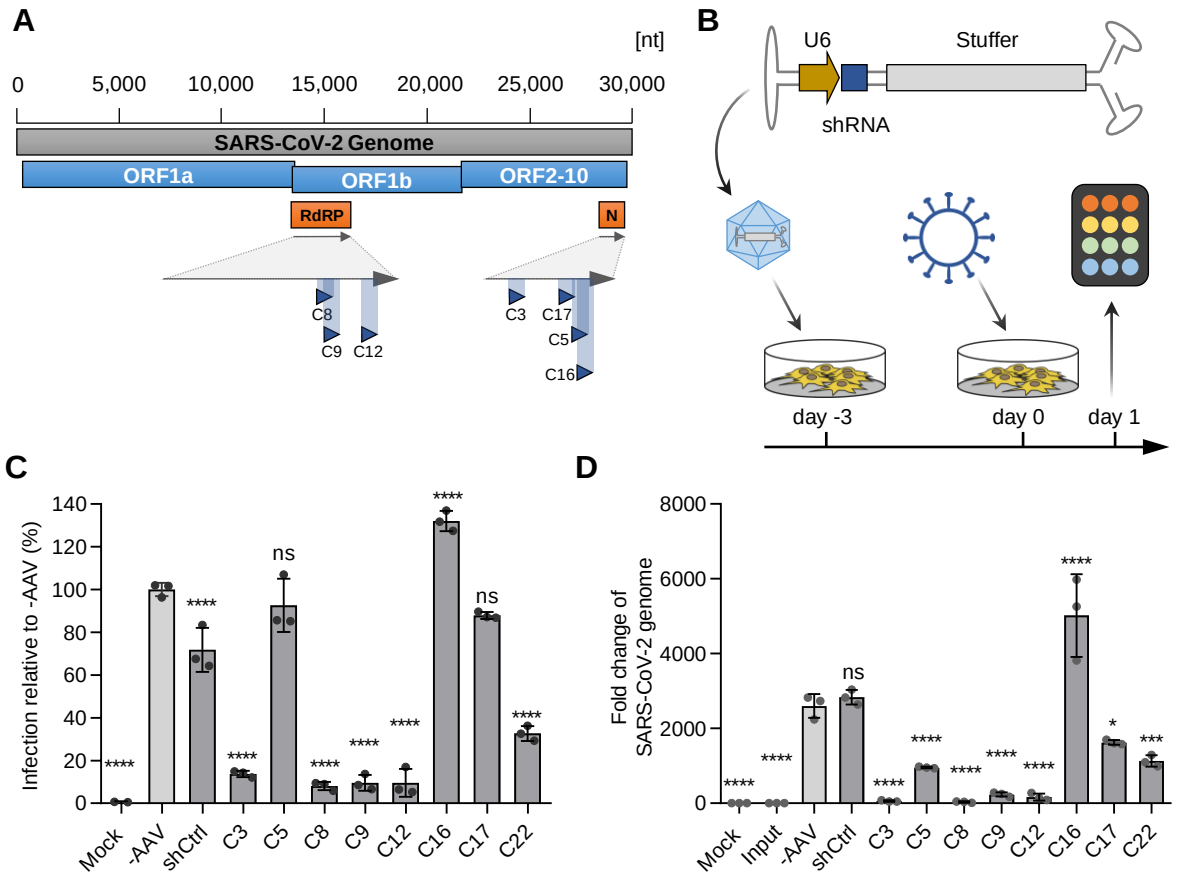



**Figure 12: Efficient knockdown of SARS-CoV-2 *via* AAV/shRNA vectors. (A)** Knockdown of SARS-CoV-2 was conducted using shRNAs to target the *RdRP* and *N* genes. Based on a crude lysate pre-screening, shRNAs C3, C5, C8, C9, C12, C16, C17 (targeting SARS-CoV-2 directly) and C20 (targeting *hACE2*) were selected for purified AAV-LK03 vector production. **(B)** Vero E6 cells were transduced with $10^5$ vector copies per cell (three wells per condition). Three days later, cells were super-infected with SARS-CoV-2 and harvested after another 24 h for analysis of SARS-CoV-2 infection by indirect immune-fluorescence (IF) assay to detect double-stranded (ds)RNA and by RTqPCR. **(C)** Relative SARS-CoV-2 infection rate of Vero E6 cells compared to non-transduced cells (-AAV) as measured by in-cell detection of dsRNA *via* IF assay. **(D)** Fold-change of SARS-CoV-2 genome relative to input as measured by RTqPCR. SARS-CoV-2 infection and infection analysis were conducted by Megan Stanifer. Triplicate measurements are presented as mean ± SD. Statistical analysis was conducted using one-way ANOVA with Dunnett's multiple comparison test. A modified version of this figure has been published [290].

### *3.1.2.2 Multiplexing of shRNAs improves SARS-CoV-2 repression*

As a rapidly evolving RNA virus, direct targeting of SARS-CoV-2 RNA or proteins for preventative or therapeutic applications is likely to trigger mutational escape. Hence, simultaneous targeting of multiple viral targets may increase the selective pressure and improve viral suppression. To test this, I combined the three shRNAs C3, C8 and C12 into a triple-shRNA vector using the TRISPR format. Before assessing their combined

knockdown efficiency, I first tested the integrity of packaged scAAV vectors containing one, three or no shRNA cassette(s) (Figure 13).

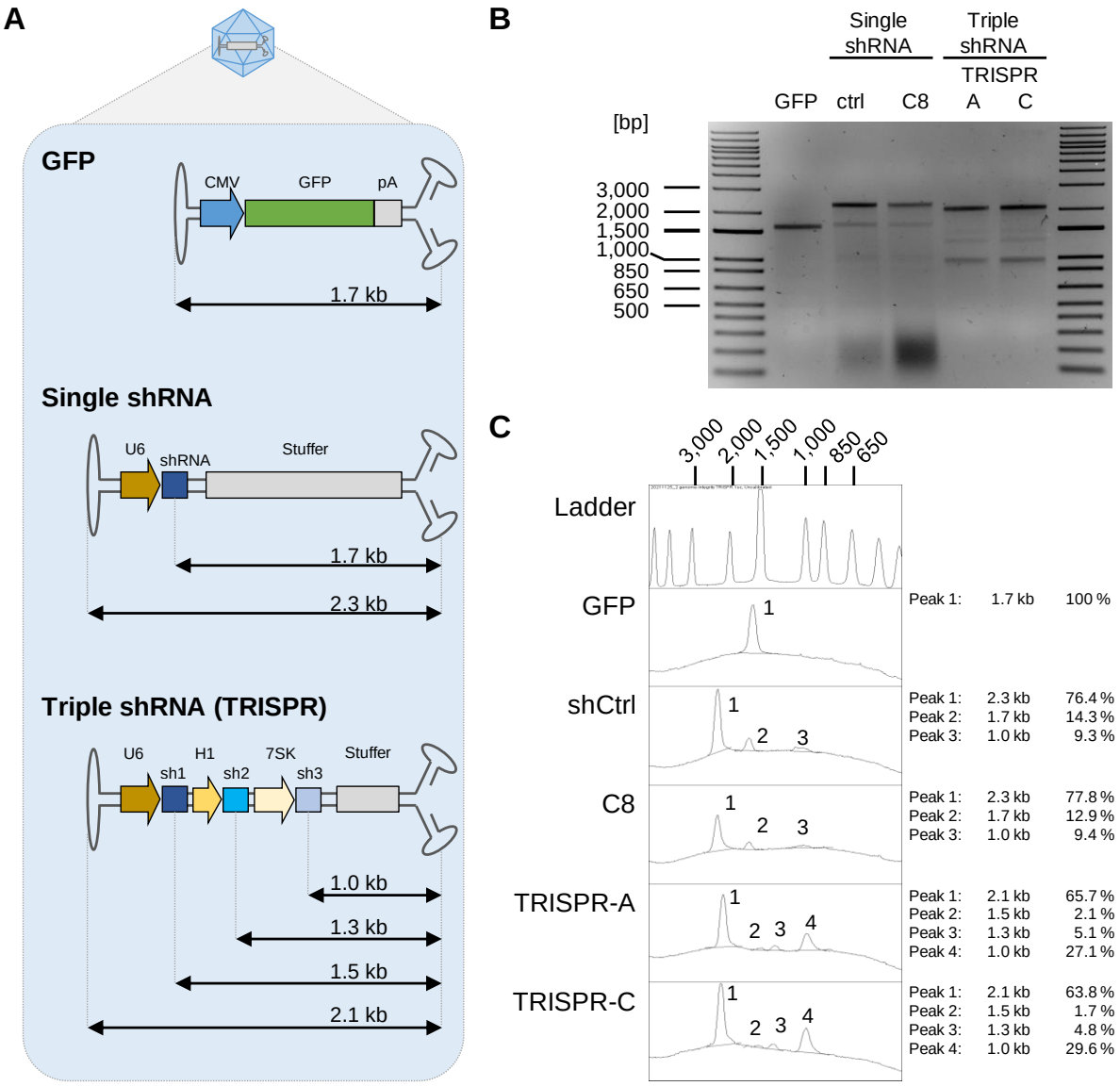



**Figure 13: Packaging of sub-genomic vector species is observed in single- and triple-shRNA vectors. (A)** Three different types of scAAV constructs were packaged into the AAV-LK03 capsid: “GFP” contains a CMV-GFP-pA cassette, “Single shRNA” contains a U6 promoter-driven shRNA expression cassette, “Triple shRNA” contains three shRNA expression cassettes in the TRISPR format using U6, H1 and 7SK promoters. Genomic and sub-genomic sizes are indicated as kilobases (kb). **(B)** For each construct type, $5\times10^{11}$ vg were extracted after packaging into AAV-LK03 and analyzed *via* native agarose gel electrophoresis. For single-shRNA vectors, shRNAs shCtrl and C8 were investigated. For triple-shRNA vectors, TRISPR A (C8-C12-C3) and TRISPR C (triple shCtrl) were used. The agarose gel was stained using ethidium bromide for DNA visualization. **(C)** Band densities of the agarose gel were quantified using ImageJ. Peaks are labeled, giving approximate peak sizes (compared to ladder) and relative peak volumes for each lane. A modified version of this figure has been published [290].

Strong secondary structures of transgenes (such as shRNA cassettes containing stem-loop-stem structures) can hinder AAV packaging and lead to vectors particles containing partial genome fragments [291]. To test whether single- or triple-shRNA vector genomes retain structural integrity after packaging compared to a vector with no shRNA cassette, I packaged all three of these vector types into AAV-LK03. Specifically, I compared vectors with no shRNA cassette ("GFP" vector), with a single shRNA cassette (shRNAs shCtrl or C8) or a with a triple shRNA cassette (TRISPR A with C8-C12-C3 or triple shCtrl) (Figure 13A). After vector production and purification, I extracted packaged vector genomes and analyzed their integrity *via* native agarose gel electrophoresis. This showed a single genomic band for the GFP vector (Figure 13B), represented by a single peak during band density quantification (Figure 13C). This peak at ~1.7 kb matches the full-sized vector genome. For both single-shRNA vectors (shCtrl and C8), three bands/peaks were visible: (i) a major peak at ~2.3 kb matching the full genome size, (ii) a smaller peak at ~1.7 kb matching in size to a sub-genomic fragment spanning from the shRNA to the right ITR (ITR2), and (iii) a minor peak at ~1.0 kb. For both triple-shRNA vectors, density quantification showed four peaks: (i) a major peak at ~2.1 kb matching the entire genome size, (ii) a faint peak at ~1.5 kb (matching in size to a sub-genomic fragment spanning from the first shRNA to the right ITR), (iii) a peak at ~1.3 kb (matching in size to the distance of the second shRNA to the right ITR), and (iv) a peak at 1.0 kb (matching the distance from the third shRNA to the right ITR). Peaks 2-4 gradually increased in density (2<3<4), whereas peak 1 equaled approximately one third of the combined peak density, suggesting that two thirds of packaged genomes were full-length constructs.

With the majority of TRISPR vectors showing correct genome size, the next experiment was a functional test for triple-shRNA vectors targeting SARS-CoV-2. Therefore, I constructed TRISPR vectors with the anti-SARS-CoV-2 shRNAs C3, C8 and C12 with two configurations TRISPR A (C8-C12-C3) and TRISPR B (C3-C12-C8). To allow assessment of each individual shRNA within these combinations, I cloned 12 more TRISPR vectors (TRISPR C-N) as permutations of TRISPR A and B containing a non-targeting shRNA (ctrl) at each of the three positions, including the fully non-functional TRISPR C (ctrl-ctrl-ctrl) (Figure 14A). Vectors were packaged into AAV-LK03 for pre-treatment of Vero E6 cells prior to infection with SARS-CoV-2 as above (Figure 14B-C).

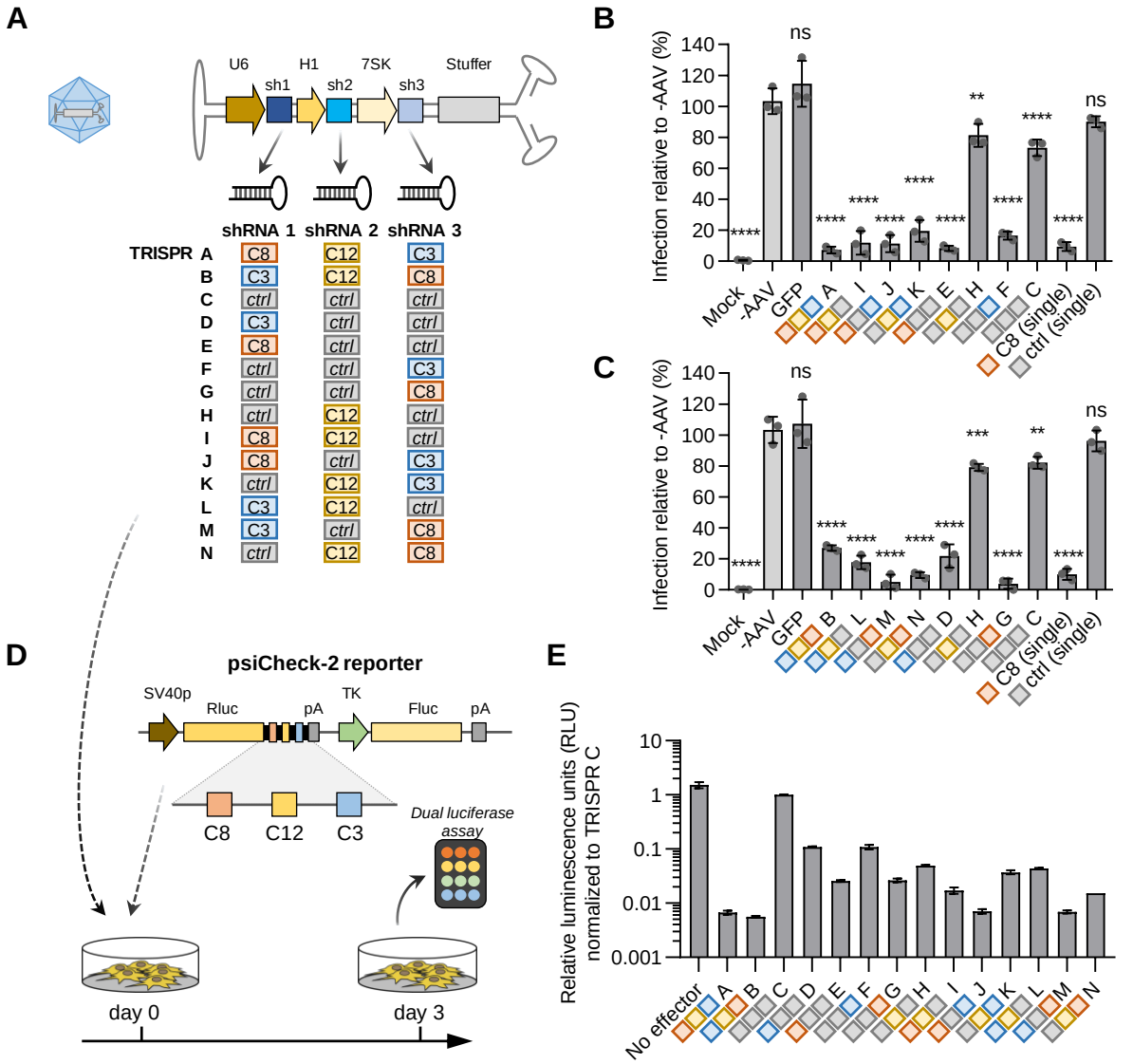



**Figure 14: Multiplexing of anti-SARS-CoV-2 shRNAs into AAV/TRISPR vectors allows parallel targeting of three genomic target regions. (A)** The anti-SARS-CoV-2 shRNAs C3, C8 and C12 were combined into TRISPR vectors with 14 different configurations (TRISPR A-N). These extend the two fully functional TRISPR A (C8-C12-C3) and TRISPR B (C3-C12-C8) permutations by replacing effector shRNAs with a non-targeting "ctrl" shRNA. **(B-C)** SARS-CoV-2 knockdown was tested by pre-treatment of Vero E6 cells with AAV-LK03 (including TRISPR vectors, single-shRNA vectors C8 and ctrl, no-shRNA control vector "GFP", and non-transduced "-AAV" control). Three days later, cells were infected with SARS-CoV-2 ("Mock": no SARS-CoV-2 in non-transduced cells) followed by quantification of infection with a dsRNA antibody as before. Panel **B** depicts the TRISPR A lineage, while panel **C** shows TRISPR B and derivates. SARS-CoV-2 infection and quantification were conducted by Megan Stanifer. Values show mean infection relative to non-transduced control ("-AAV"; ±SD). Statistical analysis was conducted by one-way ANOVA with Dunnett's multiple comparison test. **(D-E)** Specific knockdown for each TRISPR construct was tested using transfection-based dual luciferase assay in Hek293T cells with C8, C12 and C3 target sites placed in the 3'-UTR of *Rluc* in the psiCheck-2 reporter. Knockdown was quantified four days after transduction and is presented after normalization to non-functional TRIPSR C (mean of triplicates ± SD). Modified versions of this figure were published [290].

Compared to a non-transduced control (“-AAV”), all TRISPR vectors (including non-functional TRISPR C) induced significant reduction of SARS-CoV-2 infection. Single-shRNA ctrl and no-shRNA vector “GFP” did not significantly reduce SARS-CoV-2 infection. TRISPR vectors containing C3 and/or C8 generated a strong knockdown comparable to single-shRNA C8, while C12 expressed from the weak H1 promoter did not noticeably contribute to reduction in SARS-CoV-2 infection as seen for TRISPR H. To assess whether C12 shRNA is functional within the TRISPR format, I conducted a dual luciferase reporter assay in Hek293T cells using a psiCheck-2 reporter containing the C8, C12 and C3 binding sites in the 3’-UTR of *Rluc* (Figure 14D-E). This assay showed almost equal knockdown efficiencies between the pairs of TRISPR A and B (fully functional vectors), TRISPR D and F (containing only the C3 shRNA under the U6 or 7SK promoter, respectively), TRISPR E and G (containing only the C8 shRNA under the U6 or 7SK promoter, respectively), TRISPR I and N (encoding both C8 and C12 shRNAs, but switched between the U6 and 7SK promoters), TRISPR J and M (encoding both C3 and C8 shRNAs, but switched between the U6 and 7SK promoters), and TRISPR K and L (encoding both C3 and C12 shRNAs, but switched positions for C3 between the U6 and 7SK promoters). The similar knockdown efficiencies for shRNAs expressed from either the U6 or 7SK promoter imply a similar expression rate for these two promoters in Hek293T cells. Furthermore, the luciferase assay clearly demonstrated a functional knockdown by each effector within the TRISPR setting.

Simultaneous targeting of three different genomic loci by a triple-shRNA vector may block the evolution of SARS-CoV-2 variants with point mutations in either of the three target sites. To test whether TRISPR A would impede mutational escape of SARS-CoV-2 as compared to a single-shRNA vector, a passaging experiment was performed. To this end, Vero E6 cells were again pre-treated with AAV-LK03 vectors and subsequently infected with SARS-CoV-2. After 24 h, supernatant was collected and used to infect a new batch of pre-treated cells (Figure 15A). This was continued for a total of eight passages with three replicates per condition, each time measuring the infection rate with an in-cell IF assay as above.

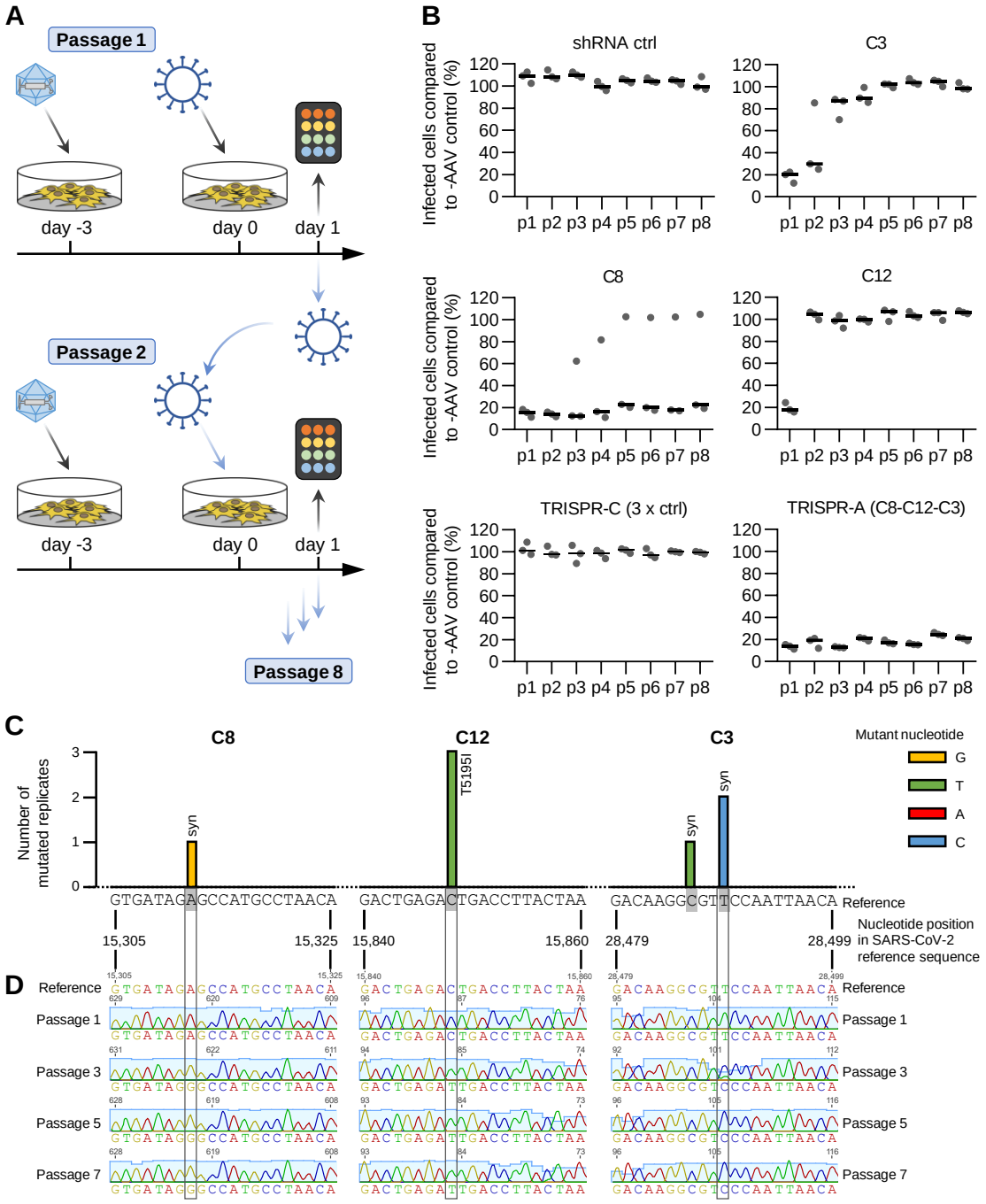



**Figure 15: Triple shRNA-based knockdown prevents mutational escape of SARS-CoV-2.** **(A)** Vero E6 cells were pre-treated with AAV vectors and infected with SARS-CoV-2 (three wells per condition). One day later, infection rates were quantified by an in-cell IF assay. Virus in supernatants was used for infection of the next batch of pre-treated cells. This was continued for a total of eight passages. SARS-CoV-2 infection and IF assay were conducted by Megan Stanifer. **(B)** Infection rates over eight passages relative to non-transduced (-AAV) control for cells pre-treated with single shRNA (ctrl, C8, C12 or C3) or triple-shRNA TRISPR vectors (C: 3x ctrl; A: C8-C12-C3). **(C-D)** SARS-CoV-2 RNA was extracted from supernatants in passage 1, 3, 5 and 7, followed by cDNA synthesis and Sanger sequencing of shRNA binding sites. **(C)** Point mutations and number of occurrences thereof among the three replicates treated with C8, C12 or C3 shRNAs. **(D)** Exemplary Sanger sequencing chromatograms for one mutated replicate per shRNA over passages 1, 3, 5 and 7. A modified version of this figure was published [290].

Relative to non-transduced cells (-AAV), the SARS-CoV-2 infection rate was constant at ~100% for all eight passages for the non-targeting single-shRNA ctrl and TRISPR C (Figure 15B). Following single-shRNA treatment with C3 and C12, a strong knockdown in infection rate was observed initially, followed by an increase in SARS-CoV-2 infection rates over the next passages until ~100% infection was reached (at passage 5 for C3 and passage 2 for C12). For C8, only one of the three replicates showed an increase in infection rate starting at passage 3, while the other two replicates remained constant at ~20% infection as compared to the non-transduced control. Only for TRISPR A (C8-C12-C3), a continued knockdown of SARS-CoV-2 infection was observed with little variance between all eight passages. To assess whether the increase in infection rates observed for single-shRNA vectors is owing to SARS-CoV-2 variants mutated within the respective shRNA binding sites, I extracted viral RNA from supernatants of passages 1, 3, 5 and 7, and converted it to cDNA for Sanger sequencing of shRNA binding sites (Figure 15C-D). C8, C12 and C3 binding sites were not mutated in the control or TRISPR A samples. In contrast, in the “escape” replicates treated with C8, C12 or C3 shRNA, single point mutations were found within shRNA binding sites after passage 3. Following C8 treatment, only the replicate that showed increased infection also exhibited a mutation (A>G; synonymous) in the C8 binding site for passages 3, 5 and 7. Following pre-treatment with C12 vectors, all three replicates escaped and showed the same point mutation (C>T). This represents a threonine to isoleucine mutation at position 5,195 of the ORF1ab polyprotein. For the C3 shRNA, one replicate showed a C>T point mutation (nucleotide position 28,486) while the other two replicates had a T>C mutation at nucleotide position 28,489 (both mutations are synonymous). Formation of these escape mutants demonstrates the necessity for simultaneous targeting of multiple genomic target sites in order to prevent SARS-CoV-2 mutational breakthrough. To test whether mutated binding sites would still show general downregulation, I performed dual luciferase reporter assays with single shRNA and TRISPR A effectors (Figure 16). I introduced perfect or point-mutated binding sites for the C8, C12 and C3 shRNAs into the 3’-UTR of *Rluc* in psiCheck-2 (Figure 16A). In each configuration, point mutations resulted in lower knockdown rates with the respective shRNA as compared to perfect binding sites. For all reporters, TRISPR A exhibited the strongest knockdown efficiency, showing 89% knockdown (normalized RLU of 0.11) even for the triple point-mutated reporter (8*-12*-3*) (Figure 16B). Thus, triple-shRNA effectors can induce efficient knockdown even for point mutations present in multiple target sites.

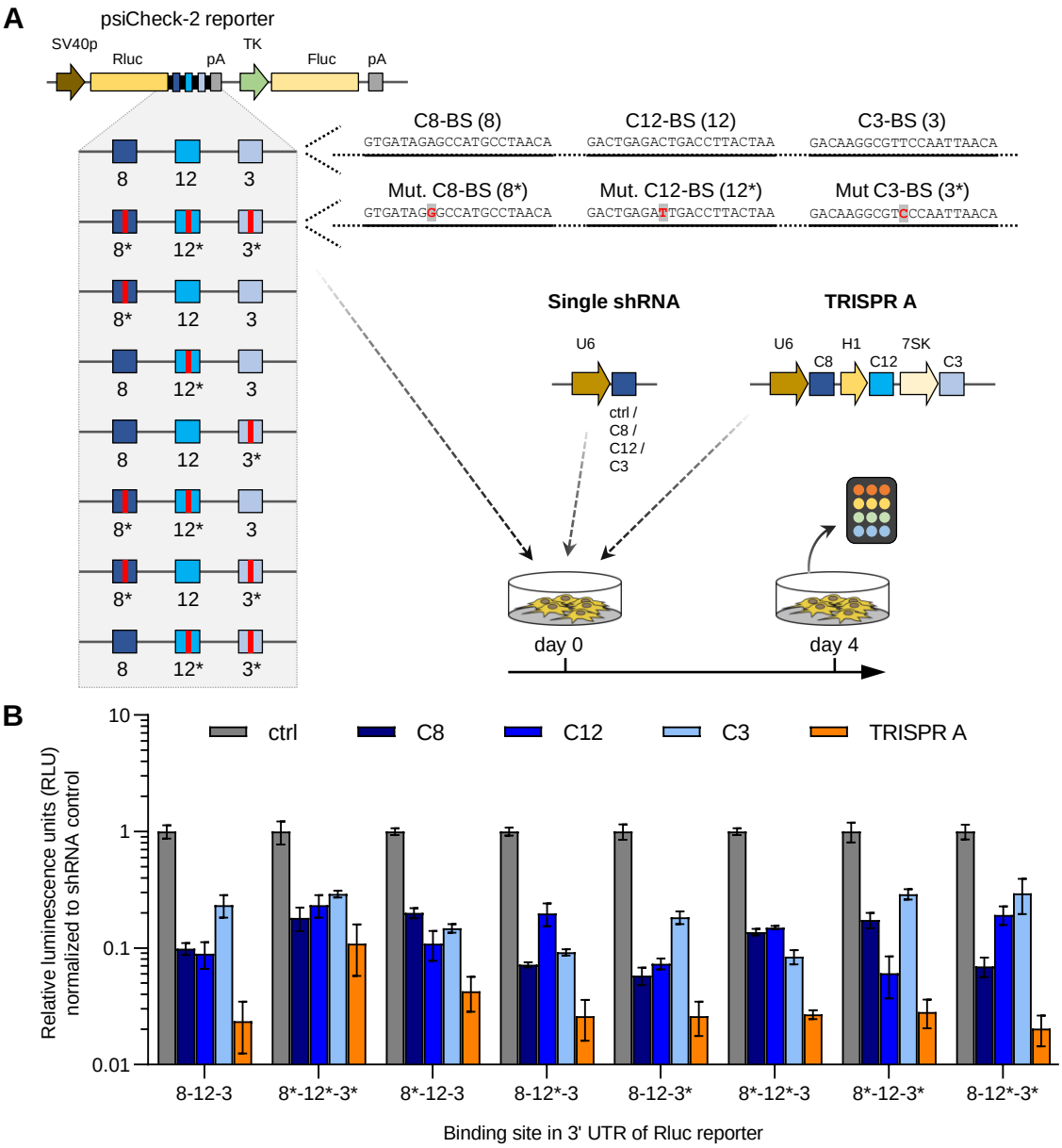



**Figure 16: Triple-shRNA strategy maintains efficient knockdown for point-mutated target sites. (A)** Binding sites (BS) for C8, C12 and C3 were cloned into the 3'-UTR of *Rluc* in psiCheck-2. Point mutations (*) were introduced into each binding site, yielding eight reporter configurations. Knockdown with single shRNA C8, C12 and C3 or triple-shRNA construct TRISPR A was tested by plasmid transfection in Hek293T cells and dual luciferase assay. **(B)** Relative luminescence normalized to non-targeting ctrl shRNA. Values represent mean of triplicates ± SD. A modified version of this figure was published [290].

Overall, shRNAs proved to induce robust knockdown for reporter, endogenous and viral target RNAs while still being sensitive to point mutated target sites and mutational escape. Multiplexing of three shRNAs into a single vector could prevent the formation of escape mutations and enabled continued SARS-CoV-2 suppression.

## 3.2 Advancing AAV vector evolution through screening and engineering of promoters and capsids

High-dose administration of AAV vectors can cause adverse effects and even lead to fatality [120]. Hence, major goals of vector development are to increase vector efficiency and specificity. Improving these factors can enable low-dose administration while still reaching therapeutic efficacy. This requires maximizing transgene expression in on-target tissues and simultaneously lowering transduction in all other tissue types.

The observed level of transgene expression is the combined result of vector delivery into the nucleus of target cells (also referred to as functional transduction) and the ability of the expression cassette to drive transcription. Both delivery and expression of transgenes in the intended target tissue are strongly restricted by (i) the employed vector capsid and (ii) modulatory elements of the transgene expression cassette (enhancer, promoter, intron, polyadenylation signal, *etc.*). On the level of vector delivery, efficiency relates to the ability to direct a large fraction of the vector particle input to the intended tissue, while specificity describes the selectivity thereof, *i.e.*, which quantity is delivered to on- vs. off-target cells. Regarding transgene expression, efficiency defines the rate of expression (how many RNA copies are transcribed per DNA input), while specificity represents the selectivity of expression across different cell types (*e.g.*, ubiquitous vs. tissue-specific promoters).

Both delivery and expression can be optimized by choosing suitable components matching the given task, such as a CNS-specific capsid combined with a neuron-specific promoter for a neuron-targeted gene therapy vector. For most applications, however, optimal capsids and modulators are yet to be discovered or engineered.

### 3.2.1 Promoter screening identifies liver-directed transgene expression with the GFAP promoter

AAV vectors offer a wide range of available natural or synthetic capsids with widely differing transduction profiles. For most capsids, though, transgene delivery is not restricted to one intended target tissue. Thus, tissue- and cell-type-specific modulators such as enhancer/promoter combinations can be employed to drive a cell- or tissue-selective expression [174]. To identify and compare suitable promoters in parallel, a combined readout of efficiency and specificity is necessary. This was achieved by former lab member Claire Domenger in a barcoded promoter screen comprising 53 different enhancer/promoter combinations and variants. Following the generation and screening

of the initial promoter library by Claire Domenger, I contributed to the analysis by (i) generating a validation dataset, (ii) using the validation dataset to optimize the promoter library data normalization, and (iii) following up on the surprising finding of superior GFAP promoter-driven transgene expression in the liver.

#### *3.2.1.1 Validation and analysis of a barcoded promoter library*

The initial promoter library screen (Figure 17A) was performed by Claire Domenger by (i) cloning 53 promoter constructs into ssAAV vectors with construct-specific barcode sequences, (ii) individual packaging into AAV9 and combining variants into a promoter library, (iii) screening in C57BL/6 mice, and (iv) assessing barcode distribution in genomic (g)DNA and complementary (c)DNA (from mRNA) of 16 different tissue types. Analysis of promoter-specific barcodes placed in the 3'-UTR of the *eYFP* transgene was conducted by high-throughput sequencing (NGS). This allowed quantification of relative proportions of each promoter and derived transcripts in each tested tissue and yielded distinct expression profiles for different promoters, with some showing more ubiquitous or more specialized transcription. Although most promoters demonstrated expression as described in literature, others showed unexpected results. The GFAP (gfa2) promoter, for instance, was a major outlier. This promoter of the human glial fibrillary acidic protein is commonly used for astrocyte-specific expression within the CNS [164,292]. Within the promoter library, however, the GFAP promoter demonstrated major transgene expression in the liver.

To assess the robustness and validate the findings of the resulting dataset, four promoter variants were selected for individual assessment. These were CMV, LP1, SPc5-12 and the aforementioned GFAP. CMV is the human cytomegalovirus immediate-early enhancer and promoter, and is widely used due to its strong ubiquitous expression profile in mammalian cells [176,177]. LP1 is an artificial promoter with a liver-specific expression profile consisting of elements from the *HCR* and *hAAT* promoters [189,293]. SPc5-12 is also a short artificial promoter but exhibits a muscle-specific expression profile [188]. This sub-selection was chosen based on the promoter library dataset and includes a combination of viral, human, and synthetic promoters with well-studied ubiquitous or selective expression profiles.

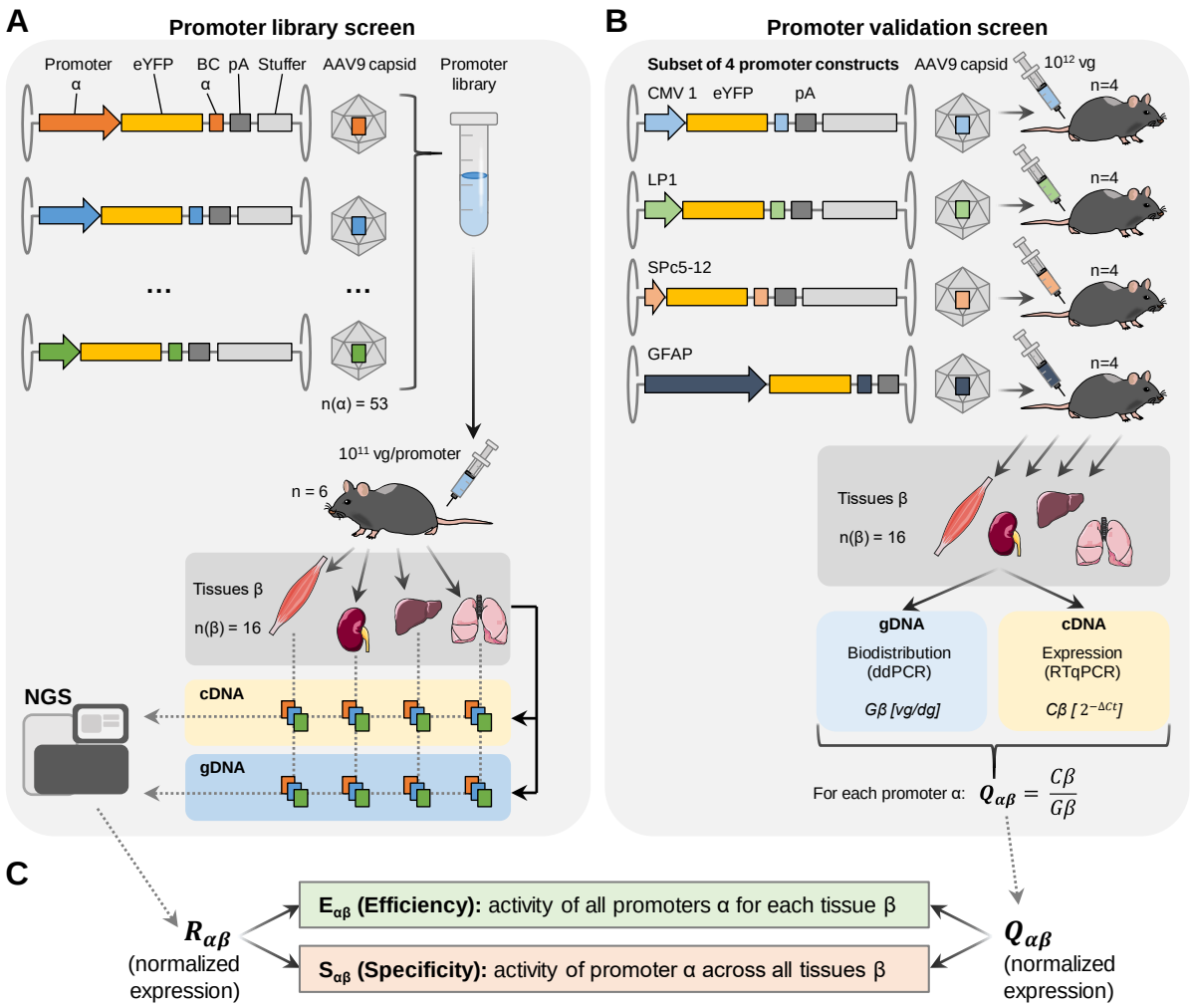



**Figure 17: Barcode- and reporter RNA-based evaluation of promoters *in vivo*. (A)** The original promoter library screen was designed and conducted by Claire Domenger. A set of 53 promoter-*eYFP* constructs was cloned into ssAAV constructs with promoter-specific barcodes. Inert *lacZ* stuffer sequences were added after the bgh-polyadenylation signal (pA) to achieve uniform vector size. Promoter constructs were individually packaged into AAV9 capsids and mixed subsequently. The resulting promoter library was screened in six C57BL/6 mice. Two weeks after injection, 16 major tissue types were harvested for DNA/RNA extraction and barcode analysis by NGS on the cDNA and gDNA level. **(B)** To confirm the results of the library screen, I conducted a promoter validation screen using the four promoter-*eYFP* constructs CMV, LP1, SPc5-12 and GFAP. Each was packaged into AAV9 and injected into four mice. After two weeks, tissues were harvested for DNA/RNA extraction and analysis of vector biodistribution by ddPCR on gDNA (Gβ) and *eYFP* transgene expression by RTqPCR on cDNA (Cβ). Normalized expression Qαβ was calculated for each promoter α in each tissue β. **(C)** Normalized expression values from the library screen (Rαβ, relative barcode concentration) and from the validation screen (Qαβ) were deconstructed into efficiency score Eαβ and specificity score Sαβ to enable a direct comparison of the two screening methods. Efficiency scores compare the activity of different promoters in each individual tissue, while specificity scores depict the expression profile of each promoter across different tissues.

I assessed the subset of four promoters individually for AAV-based transgene expression by utilizing the same promoter-eYFP-BC constructs used in the initial library screen. After packaging into AAV9, vectors were administered by tail-vein injection to four C57BL/6 mice per construct with $10^{12}$ vg per mouse. After two weeks, 16 major tissues (brain, diaphragm, eye, fat, gut, heart, kidney, liver, lung, lymph nodes, quadriceps femoris muscle, ovaries, pancreas, skin, spleen and stomach) were harvested for DNA/RNA extraction followed by analysis of vector distribution by ddPCR and relative expression by RTqPCR (Figure 17B).

The goal of both datasets was to generate a promoter expression profile independent of the vector biodistribution, which would thus be applicable to vectors whose transduction properties differ from the employed AAV9. Therefore, data from both, the original library and the validation screen, was analyzed by normalization of relative expression (measured by NGS readout of barcodes in cDNA for the library setting or by RTqPCR for the validation setting) to the vector biodistribution (measured by NGS of barcodes in gDNA or by ddPCR) (Figure 17C). These normalized expression values from the library and validation datasets were additionally deconstructed in order to represent either efficiency or specificity of reporter expression. This enables useful comparisons between the strengths of different promoters in a single tissue (efficiency) or between different tissues for each individual promoter (specificity).

For all tested constructs from the validation setting (testing promoter expression individually), normalized expression was mostly consistent between animal replicates. I observed the highest normalized expression values in liver tissue for the LP1, SPc5-12 and GFAP promoters, while CMV showed the strongest normalized expression in pancreas (Figure 18A-D). Inspecting vector presence alone, I found the highest transduction rates in liver tissue for all promoter constructs (Supplementary figure 1).

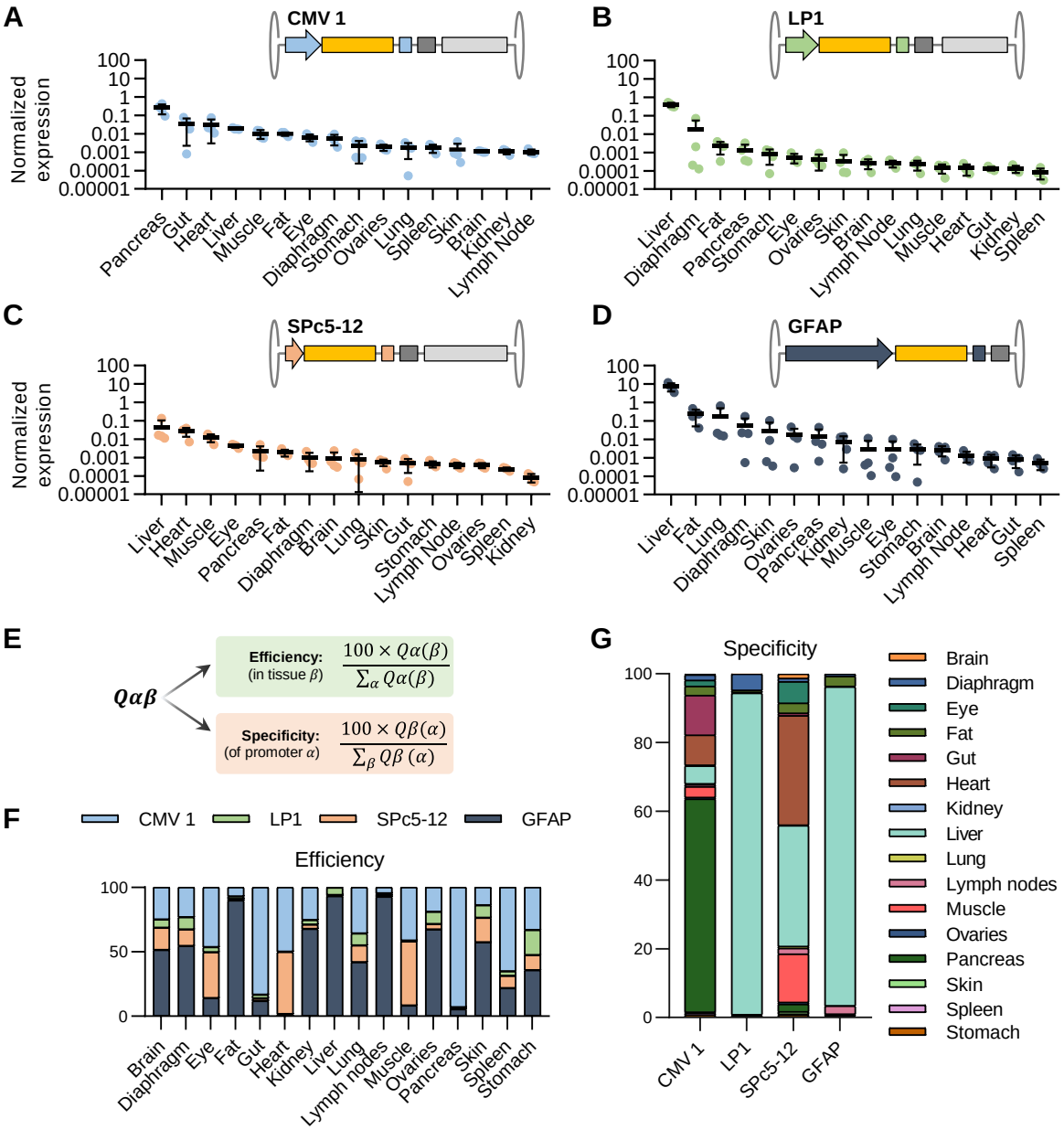



**Figure 18: Efficiency and specificity of CMV, LP1, SPc5-12 and GFAP promoters for AAV-based reporter expression. (A-D)** Results of the promoter validation screen as described in Figure 17B. Normalized expression (Qαβ) is calculated by dividing relative expression (Cβ; $2^{-\Delta Ct}$ of *eYFP* relative to *PolR2A* housekeeper) to vector distribution (Gβ; vg/dg) for each sample (Cβ and Gβ are individually depicted in Supplementary figure 1). Individual values of four mice per promoter construct are depicted with mean ± SD for CMV (**A**), LP1 (**B**), SPc5-12 (**C**) and GFAP (**D**) promoters. **(E)** Normalized expression Qαβ is deconstructed into efficiency and specificity scores by normalizing to the sum over all promoters in each individual tissue or to the sum over all tissues for each individual promoter, respectively. **(F)** Mean efficiency scores of CMV, LP1, SPc5-12 and GFAP for all 16 tissues. **(G)** Mean specificity scores for the four tested promoters across tested tissues.

To better understand and compare promoter expression profiles, normalized expression Qαβ was deconstructed into efficiency (Eαβ) and specificity (Sαβ) scores (Figure 18E). Efficiency scores were calculated for each individual tissue β by normalizing average Qα(β) (normalized expression for each promoter α within this tissue) to $\Sigma_\alpha$ (sum of Qα(β) over all tested promoters α). Specificity scores, on the other hand, were calculated for each promoter α individually by normalizing Qβ(α) (normalized expression for this promoter in each this tissue β) to $\Sigma_\beta$ (sum of Qβ(α) over all assessed tissues β). Both efficiency and specificity scores are scaled to values from 0 to 100.

Depiction of the efficiency scores (Figure 18F) enabled a good representation of relative expression strengths of each promoter, showing, for instance, superior activity of GFAP over other promoters in fat, liver, and lymph nodes. Specificity scores (Figure 18G) demonstrated that both LP1 and GFAP promoter expression was mostly focused on liver tissue, whereas the CMV promoter was most active in pancreas and SPc5-12 expressed primarily in heart, liver, and muscle tissues.

Individual screening helped to establish the expression profiles for the CMV, LP1, SPc5-12 and GFAP promoters by testing *eYFP* RNA expression for each respective vector construct in mice. The next step was now to compare this validation dataset to the data from the primary promoter library screen. Thus, I extracted the relative proportions of barcodes for the subset of CMV, LP1, SPc5-12 and GFAP promoters from the library dataset for cDNA and gDNA. Next, I calculated the ratio Rαβ by dividing barcode proportions in cDNA to the proportion in gDNA (Figure 19A). Efficiency and specificity scores were calculated from Rαβ as was done for Qαβ in the validation setting above. Comparing efficiency scores between library and validation data demonstrated a strong positive correlation between the two datasets, with a Pearson correlation coefficient of 0.91 (Figure 19B).

To calculate specificity scores, a comparison of Rαβ across different tissue samples is necessary. This requires scaling with an additional weighting factor ω to account for varying bulk transduction rates across different tissues. I tested different weighting factors based on either bulk relative expression Cβ ($2^{-\Delta Ct}$ of *eYFP* vs *PolR2A* housekeeper as measured by RTqPCR for each tissue), vector distribution Gβ (vg/dg as measured by ddPCR) or normalized expression Qβ (Cβ/Gβ). Comparing the resulting library specificity scores to the specificity scores derived from the validation dataset showed the best correlation for normalization *via* relative expression Cβ, with a Pearson correlation coefficient of 0.95 (Figure 19C).

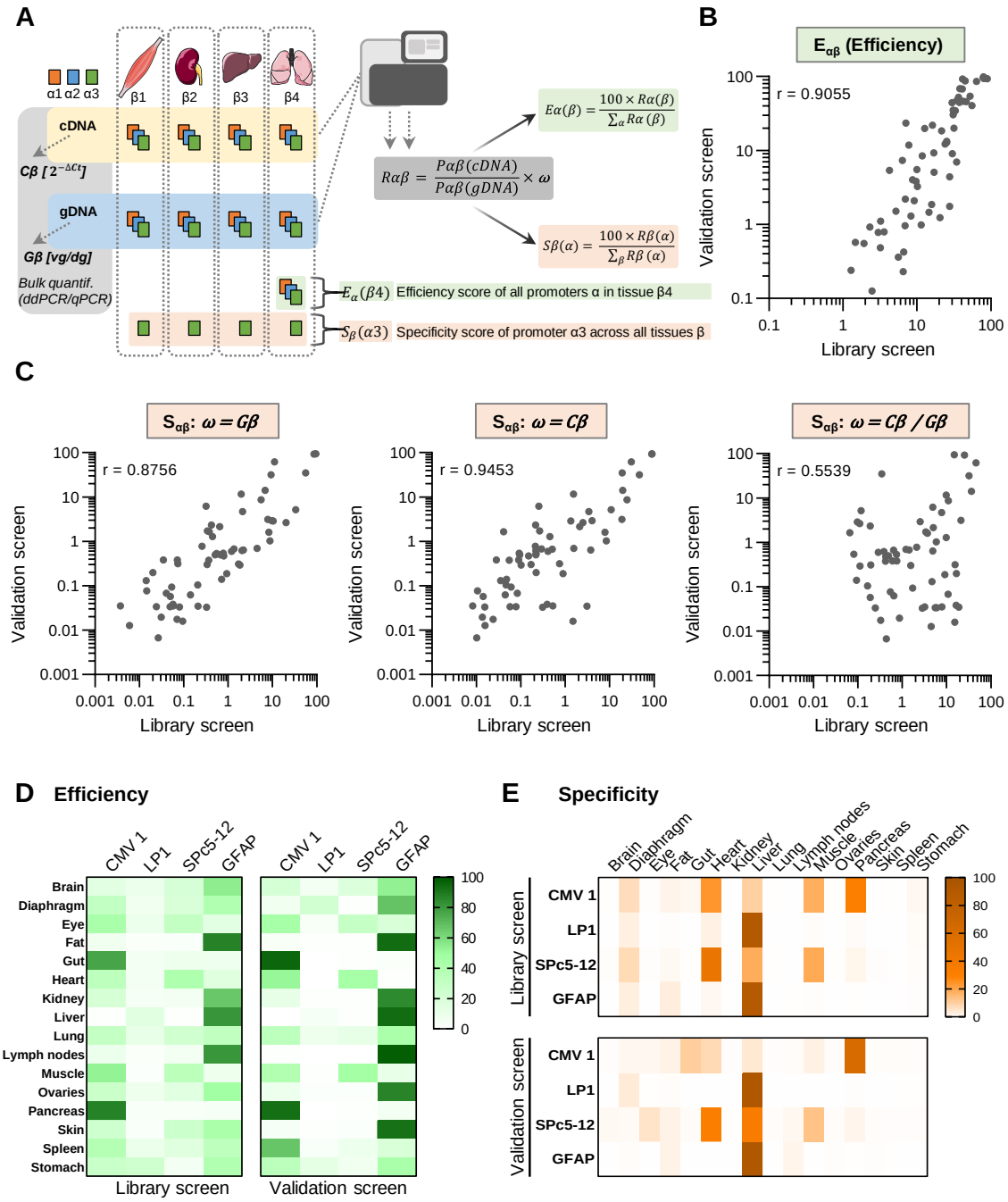



**Figure 19: NGS barcode readout of promoter library correlates well with single-promoter validation data. (A)** Data normalization conducted for the promoter library dataset. Concentrations of individual barcodes α (each representing a promoter variant) are measured *via* NGS in cDNA and gDNA of individual tissues β. Bulk relative expression Cβ and vector distribution Gβ are quantified by RTqPCR and ddPCR, respectively. For promoter analysis, barcode concentrations in cDNA (Pαβ(cDNA)) are normalized to values in gDNA (Pαβ(gDNA)), to yield normalized ratios Rαβ for each sample. An additional scaling factor **ω** is needed to allow comparisons between different tissue samples, before calculating efficiency (Eαβ) and specificity (Sαβ) scores. **(B)** Comparison between efficiency scores for the CMV, LP1, SPc5-12 and GFAP promoter for all 16 tissues as derived from the library and validation screens. **(C)** Comparing specificity scores between the library and validation screens using different scaling factors ω. **(D-E)** Heatmaps showing mean efficiency and specificity scores for the library and validation screens. The data of the library screen was acquired by Claire Domenger.

The strong correlations between library and validation data are also well depicted by heatmaps for efficiency and specificity scores in Figure 19 D and E. For efficiency scores (Figure 19D), the library screen correctly predicted the strongest promoter for 14 out of 16 tested tissues (the exceptions were heart and muscle). For specificity scores (Figure 19E), library and validation data both showed the strongest expression for the CMV promoter in pancreas tissue. For the LP1 and GFAP promoters, both datasets showed the strongest expression to be derived from liver tissue. For the SPc5-12 promoter, however, the library screen showed the strongest expression in heart, whereas the individual validation showed a stronger expression in liver tissue with this promoter.
Overall, the validation data largely confirmed the results from the promoter library screen and thus demonstrated the robustness of the NGS-based barcode readout for parallel assessment of promoter expression profiles.

#### *3.2.1.2 GFAP promoter drives specific reporter expression in hepatocytes*

Surprisingly, both library and single-promoter screens showed a pronounced activity of the GFAP promoter in liver tissue. Since GFAP is usually used as an astrocyte-specific promoter for expression in the CNS [164,292], its prominent expression activity in liver was unexpected. To identify cell type(s) contributing to GFAP promoter-driven expression in liver, a collaborative experiment was conducted together with Pervinder Choksi (Willenbring lab, UCSF) (Figure 20A).
The AAV9/GFAP-eYFP vector was injected into two *Lrat-Cre*$^{+/-}$*R26-RFP*$^{+/+}$ mice. Two weeks later, mice were sacrificed, followed by fluorescence-activated cell sorting (FACS)-based isolation of different populations of liver cells. Isolated cells include hepatocytes (Hep), cholangiocytes (Chol), hepatic stellate cells (HSC), macrophages (Mac) and liver sinusoidal endothelial cells (LSEC) (mouse injections and cell type isolation were conducted by Pervinder Choksi, Willenbring lab). For each of these cell populations, I performed DNA/RNA extractions, as well as quantification of vector distribution and relative expression. Subsequent calculations of normalized expression (Qβ) and specificity scores were conducted as explained above. Notably, mouse 1 showed a lower vector copy number in hepatocytes than mouse 2 (9.2 vs. 41.7 vg/dg), and also showed lower relative expression than mouse 2 in these cells. However, normalized expression and specificity scores were almost identical for hepatocytes between the two replicate mice (Figure 20B). This again demonstrates the usefulness of normalizing relative expression to vector distribution for generating a robust readout of promoter activity. The

comparison between different liver cell types demonstrated robust GFAP promoter-driven *eYFP* expression that was mostly limited to hepatocytes.

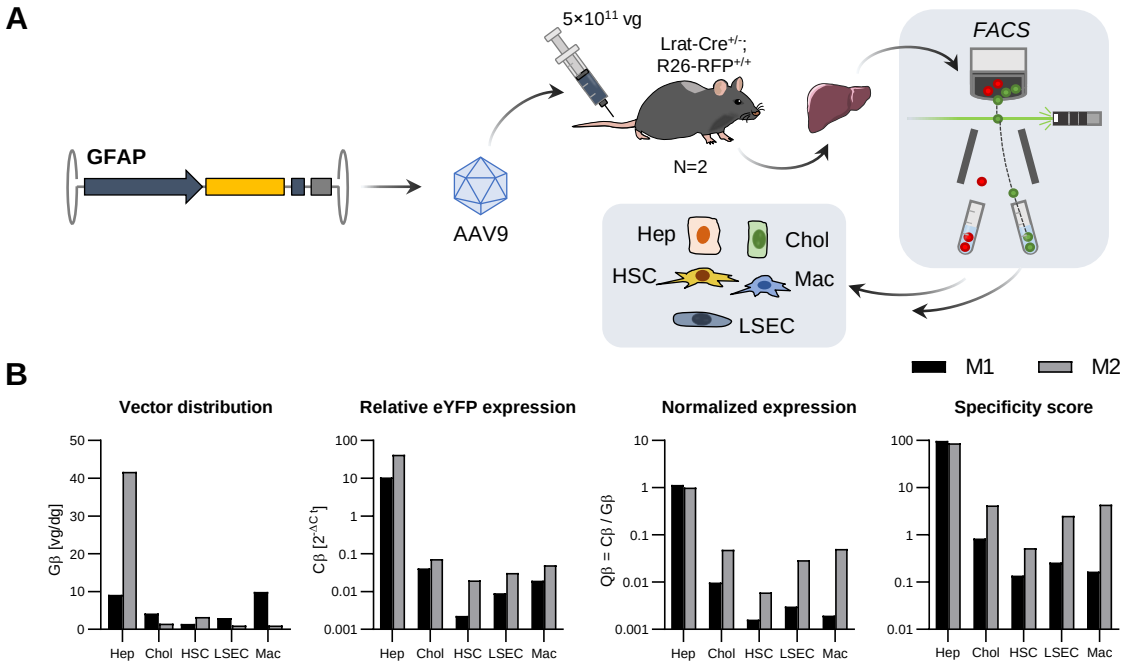



**Figure 20: GFAP promoter drives robust expression in murine hepatocytes. (A)** Activity of GFAP promoter-driven *eYFP* reporter expression was tested by injecting $5\times10^{11}$ vg of AAV9/GFAP-*eYFP* into two *Lrat-Cre*$^{+/-}$*R26-RFP*$^{+/+}$ mice. After two weeks, mice were sacrificed and hepatocytes (Hep), cholangiocytes (Chol), hepatic stellate cells (HSC), macrophages (Mac) and liver sinusoidal endothelial cells (LSEC) were isolated from liver by FACS (mouse injection and cell isolation were conducted by Pervinder Choksi, Willenbring lab). **(B)** From each cell population, I isolated DNA and RNA to determine vector distribution (Gβ) by ddPCR and relative *eYFP* expression (Cβ) by RTqPCR. Normalized expression Qβ and specificity scores were calculated as before.

To better understand the high activity of the GFAP promoter in hepatocytes, the promoter library dataset was revisited. Apart from the full-length GFAP promoter, a truncated version thereof (gfaABC1D [164]) had also been included in the promoter library screen. Out of all 53 included promoters, GFAP yielded the highest efficiency score in liver, whereas gfaABC1D was ranked 23rd (full library dataset not shown). The stronger expression observed for GFAP compared to gfaABC1D suggests that regions present only in the full-length GFAP promoter may contain binding sites for transcription factors responsible for its strong activity in hepatocytes. To identify these putative regions, I selected 20 transcription factors known for high activity in hepatocytes [282-287] and interrogated their binding profiles for matches within the GFAP promoter sequence (Figure 21A and Table 26). This was done for human and mouse transcription factors

with annotated position frequency matrices (PFMs) found within the JASPAR 2022 database [288]. Putative binding sites were identified for ten out of the 20 tested transcription factors, most of which lie within regions that are present in GFAP but not gfaABC1D. To dissect the activity of GFAP compared to gfaABC1D, and to identify potentially crucial transcription factor interactions, five truncation variants GFAPdel1-5 were created. Each of these variants is missing a region present in GFAP but not gfaABC1D (Figure 21B). Equivalent to other promoter constructs, I cloned GFAPdel1-5 into a ssAAV vector with an *eYFP* transgene and a variant-specific barcode in the 3'-UTR. Promoter mini-libraries were created by packaging the GFAP variants (full length GFAP, gfaABC1D and GFAPdel1-5) and benchmark promoters from the validation screen (CMV, LP1 and SPc5-12) into either AAV9, for broad transduction *in vivo*, or AAV-DJ, which offers strong transduction *in vitro* [133,294]. After mixing, library compositions were assessed by NGS. This demonstrated near equivalent contributions of each of the ten included promoter variants, with slightly lower contributions for GFAP (Figure 21C).

I tested the *in vitro* activity of the promoter variants by transducing Hepa1-6 cells (mouse hepatoma), Huh7 cells (human hepatoma) and primary human hepatocytes (PHH) with the AAV-DJ mini-library (Figure 21D). Three days later, DNA and RNA were extracted for NGS-based readout of barcode composition and calculation of efficiency scores as was done above. This demonstrated a strong activity of the CMV promoter *in vitro*, whereas expression from GFAP was not as prominent as seen in the mouse studies before. For each cell type, truncated versions of GFAP showed lower activity than their full-length counterpart, with the lowest activity observed for gfaABC1D and intermediate efficiencies for GFAPdel1-5. All of the GFAPdel1-5 variants showed similar expression levels, with the highest activity measured for GFAPdel4 in both Hepa1-6 cells and PHH. For PHH, I also assessed promoter activity by RTqPCR after transduction with individual promoter constructs (Figure 21E). This demonstrated highly similar trends as compared to the library setting (Pearson correlation coefficient between library setting and mean expression in RTqPCR: r = 0.9987), although the expression levels between the GFAP variants cannot be clearly distinguished.

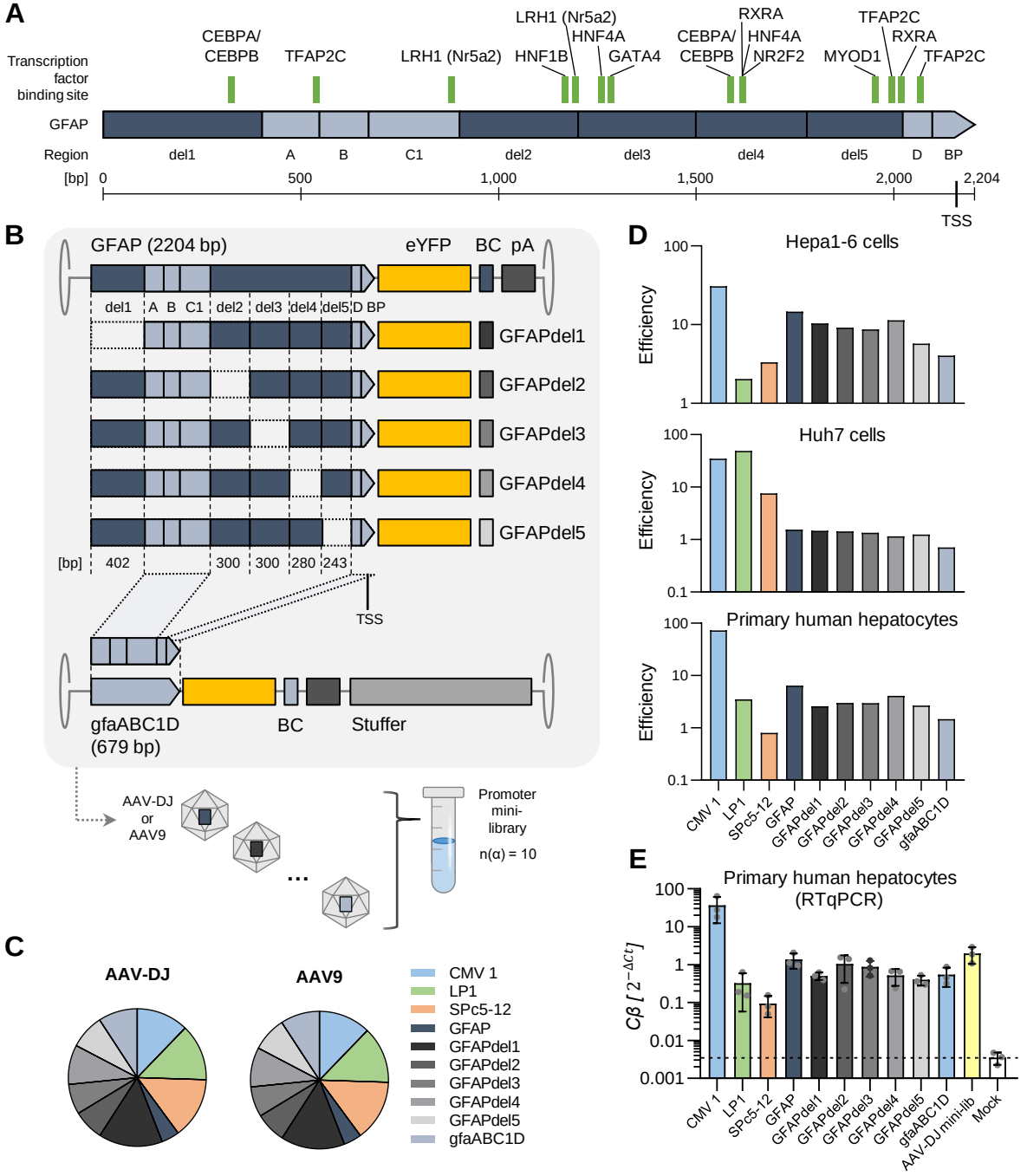



**Figure 21: Dissecting GFAP promoter activity in murine and human hepatocytes *in vitro*.** **(A)** Annotation of putative hepatocyte-specific transcription factor binding sites within the GFAP promoter. Details can be found in Table 26. Light blue regions (A, B, C1, D and basic promoter BP) exist also in gfaABC1D. **(B)** Truncated versions of GFAP with variant-specific barcodes were cloned by deleting five regions (del1-5) present in GFAP but not gfaABC1D. Promoter mini-libraries were produced with AAV-DJ or AAV9 containing GFAP variants as well as CMV, LP1 and SPc5-12 benchmarks. **(C)** Compositions of promoter mini-libraries in AAV-DJ and AAV9. **(D)** Hepa1-6, Huh7 cells and primary human hepatocytes (PHH) were transduced with the AAV-DJ mini-library (MOI of $10^5$). Promoter efficiency scores were calculated as before from the barcode composition in cDNA vs. gDNA (extracted three days after transduction). **(E)** PHH were transduced with individual AAV-DJ vectors or the AAV-DJ mini-library (three wells per condition). Three days after transduction, relative *eYFP* expression compared to the *HPRT* housekeeper was measured by RTqPCR.

**Table 26: Putative binding sites for transcription factors (TF) with major activity in hepatocytes.** Annotation of binding position within the GFAP promoter for hepatocyte-specific transcription factors. For each included transcription factor, binding position, region, and sequence within the GFAP promoter are annotated, as well as the MEME ID (JASPAR 2022 database)[288], species origin and literature reference for hepatocyte-directed TF expression. Depending on availability, human and/or mouse transcription factor memes were analyzed.

| Transcription factor | GFAP position (start/ end) | | GFAP region | Binding sequence | Meme ID | Species | Ref. |
|---|---|---|---|---|---|---|---|
| CEBPA | 322/ | 333 | del1 | TGGCGCAACCAC | MA0102.1 | *Mus musculus* | 282,283 |
| CEBPB | 324/ | 334 | del1 | GGTTGCGCCAC | MA0466.1 | *Homo sapiens* | 284 |
| TFAP2C | 532/ | 546 | A | CCGCCCCCCAGGGCC | MA0524.1 | *Homo sapiens* | 282 |
| LRH1 (NR5A2) | 875/ | 889 | C1 | AAGTCCAAGGACACA | MA0505.1 | *Mus musculus* | 295 |
| HNF1B | 1165/ | 1176 | del2 | CTCATGTGTAAC | MA0153.1 | *M.m. and H.s.* | 285 |
| LRH1 (NR5A2) | 1188/ | 1202 | del2 | GCTGTCAAGGCCTGG | MA0505.1 | *Mus musculus* | 295 |
| HNF4A | 1265/ | 1279 | del3 | CAGGACTTTAGCCCC | MA0114.2 | *Homo sapiens* | 282,295 |
| HNF4A | 1265/ | 1280 | del3 | GGGGCTAAAGTCCTGA | MA0114.3 | *Mus musculus* | 282,295 |
| GATA4 | 1276/ | 1287 | del3 | TGCCTTATCAGG | MA0482.2 | *Homo sapiens* | 282 |
| CEBPA | 1585/ | 1595 | del4 | ATTTCATAACC | MA0102.3 | *Homo sapiens* | 282,283 |
| CEBPB | 1586/ | 1596 | del4 | GATTTCATAAC | MA0466.1 | *Homo sapiens* | 284 |
| PPARA::RXRA | 1608/ | 1625 | del4 | AAGCAGGTCAGAGGTCAT | MA1148.1 | *Homo sapiens* | 282,295 |
| HNF4A | 1611/ | 1625 | del4 | ATGACCTCTGACCTG | MA0114.2 | *Homo sapiens* | 282,295 |
| HNF4A | 1611/ | 1626 | del4 | CAGGTCAGAGGTCATC | MA0114.3 | *Mus musculus* | 282,295 |
| NR2F2 | 1616/ | 1626 | del4 | CAGAGGTCATC | MA1111.1 | *Homo sapiens* | 282 |
| RXRA | 1616/ | 1626 | del4 | CAGAGGTCATC | MA0512.1 | *Mus musculus* | 282,295 |
| MYOD1 | 1950/ | 1962 | del5 | TACCACCTGCCTC | MA0499.2 | *Homo sapiens* | 283 |
| MYOD1 | 1951/ | 1963 | del5 | ACCACCTGCCTCA | MA0499.1 | *Mus musculus* | 283 |
| TFAP2C | 1989/ | 2003 | del5 | CTCTGCCTCTGGGCA | MA0524.1 | *Homo sapiens* | 282 |
| RXRA | 2005/ | 2015 | del5 | CTGAGGTCACT | MA0512.1 | *Mus musculus* | 282,295 |
| TFAP2C | 2061/ | 2075 | D | CACCCCCTCAGGCTA | MA0524.1 | *Homo sapiens* | 282 |
| TFAP2C | 2063/ | 2077 | D | CATAGCCTGAGGGGG | MA0524.1 | *Homo sapiens* | 282 |
| AHR::ARNT | – | – | – | | MA0006.1 | *Mus musculus* | 282 |
| FOXA1 | – | – | – | | MA0148.1 | *Homo sapiens* | 282,283 |
| FOXA2 | – | – | – | | MA0047.3 | *Homo sapiens* | 286 |
| FOXA2 | – | – | – | | MA0047.2 | *Mus musculus* | 286 |
| GATA4 | – | – | – | | MA0482.1 | *Mus musculus* | 282 |
| HNF1A | – | – | – | | MA0046.2 | *Homo sapiens* | 282,283 |
| HNF1A | – | – | – | | MA1991.1 | *Mus musculus* | 282,283 |
| LEF1 | – | – | – | | MA0768.1 | *Homo sapiens* | 283 |
| LEF1 | – | – | – | | MA0768.2 | *Mus musculus* | 283 |
| ONECUT1 | – | – | – | | MA0679.1 | *Homo sapiens* | 282,295 |
| ONECUT2 | – | – | – | | MA0756.1 | *Homo sapiens* | 285 |
| TCF7 | – | – | – | | MA0769.2 | *Homo sapiens* | 287 |
| TCF7 | – | – | – | | MA0769.1 | *Mus musculus* | 287 |
| XBP1 | – | – | – | | MA0844.1 | *Homo sapiens* | 282 |

To gain insights into the *in vivo* activity of truncated GFAP variants, and to assess differences between promoter activity between human and murine cells, the AAV9 mini-library was screened in a humanized liver FRGN mouse model [279] (Figure 22A). FRGN mice are immune-deficient and allow partial repopulation of the liver with primary human hepatocytes. Therefore, they are well suited for comparing *in vivo* expression of promoters in hepatocytes of both species. This *in vivo* study was again performed in collaboration with the Willenbring lab, with the injection and the FACS-based human- and mouse-hepatocyte isolation being performed by Pervinder Choksi. Efficiency scores calculated from barcode frequencies in cDNA and gDNA were highly similar between murine and human hepatocytes (Pearson r of 0.9961) (Figure 22B). The humanized mouse study also showed similar trends for expression in bulk liver tissue as those found in the initial promoter library screen (data not shown), with expression levels in the order of GFAP > LP1 > gfaABC1D > SPc5-12. The only exception is CMV, which showed weaker expression in liver than all other mentioned benchmarks in the primary library screen but yielded stronger expression than gfaABC1D and SPc5-12 in the humanized mouse model. The expression profile in the FRGN mouse did not replicate the superior activity of CMV that was observed in the *in vitro* screens above, but confirmed the finding of superior GFAP promoter activity observed by the previous mouse studies. Truncated versions of GFAP all showed lower expression than the full-length promoter, with the lowest expression observed for gfaABC1D. GFAPdel1 and GFAPdel4 showed slightly higher expression levels than GFAPdel2/-3/-5.

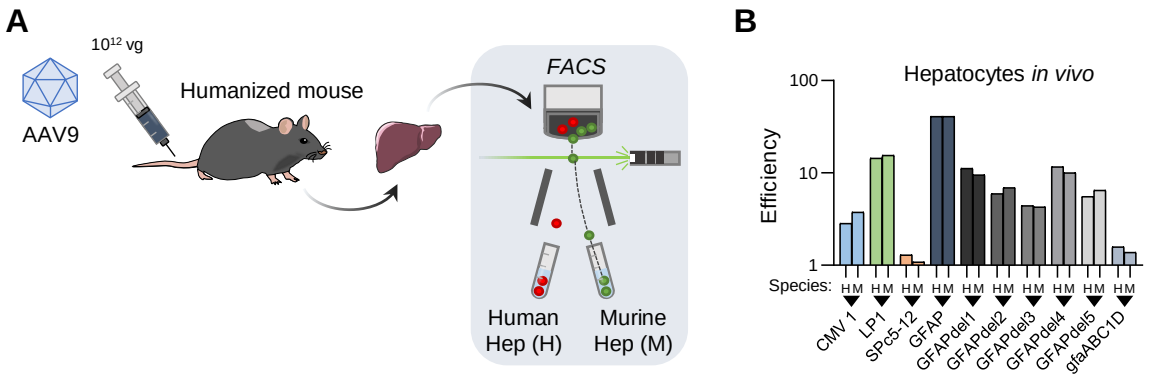



**Figure 22: Promoter activity in human and murine hepatocytes *in vivo*. (A)** The AAV9 promoter mini-library containing GFAP variants and benchmarks was injected into a FRGN mouse model engrafted with human hepatocytes. Two weeks later, liver was harvested for FACS-based isolation of human and mouse hepatocytes (mouse injection and hepatocyte isolation were conducted by Pervinder Choksi, Willenbring lab). **(B)** As before, promoter efficiency scores were calculated from the barcode composition in cDNA vs gDNA of human (H) and murine (M) hepatocytes.

Overall, dissection of GFAP promoter activity did not identify a specific promoter sub-region that could explain its prominent expression in hepatocytes. Instead, each truncated variant showed decreased expression as compared to the full-length counterpart but still outperformed gfaABC1D, indicating a combinatorial contribution of each of these regions to GFAP promoter activity in the liver. Barcode-based promoter readout was once again shown to generate robust results highly comparable to expression data collected by individual promoter screens.

### 3.2.2 Controlling directed evolution of AAV vectors through negative selection

While the application of tissue-specific promoters can focus transgene expression to a tissue of interest, the inherent transduction profile of the employed vector ultimately defines the required dose and the level of vector presence in on- and off-target tissues. Most naturally occurring AAV capsids offer broad, non-selective transduction of multiple tissue types after systemic application with an enrichment in the liver. If robust transduction of tissues other than the liver is intended, such as skeletal muscle for applications in muscular dystrophy or spinal muscular atrophy, high vector doses are required to achieve sufficient delivery. AAV vectors with prominent off-target transduction in the liver can induce severe adverse events such as hepatotoxicity [120].

Capsid engineering can help to create vectors with specialized transduction profiles. This is commonly achieved by directed evolution, where naturally occurring AAV capsids are diversified into capsid libraries and selected by *in vivo* screening [124]. These screens are usually performed iteratively, by identifying capsid variants enriched in the on-target tissue and selecting these for follow-up enrichment in secondary selection rounds. One problem of this approach is an imbalance in selection pressures [125]. Applying purely positive on-target selection benefits variants that are generally good at transduction but do not necessarily target the intended tissue specifically. The addition of a negative selection pressure to “punish” the transduction of off-target tissues may improve this imbalance. One approach to achieve this negative selection was explored during my Master thesis [275], in which I developed the “depletion of off-targeting AAVs from on-target libraries” (DEPOOL) as a means to remove liver-targeting (off-targeting) variants from targeted capsid selections in lung or pancreas (on-targets). Here, the negative selection is applied by using Cas9-based *in vitro* cleavage of *cap* gene libraries prior to subcloning and vector library production. For the current study, I followed up on the results of the

Master thesis project, by improving the negative selection conditions and exploring different applications for the DEPOOL procedure.

#### *3.2.2.1 Cas9-based targeting of peptide insertion sites*

A prominent method of capsid diversification is the insertion of randomized peptides into the capsid surface to generate new potential receptor interactions [132]. The initial iteration of DEPOOL performed during my Master thesis intended to use sgRNA libraries matching the peptide insertions in *cap* genes in order to apply a Cas9-based negative selection. After testing the functionality of insertion-site-directed Cas9 cleavage *in vitro*, the functionality of DEPOOL had been assessed using *in vivo* screening of an AAV8 peptide library (Figure 23A). Following an initial *in vivo* selection round, *cap* variants with peptide insertion were isolated from liver tissue as the defined off-target. From these variants, a sgRNA library was cloned using a multi-step cloning procedure. With this cloning strategy, each peptide-display *cap* variant yields a sgRNA that can bind to its template sequence. Thereby, the cloning templates (peptide display variants) found in the liver can facilitate their own respective depletion. Design of the sgRNA spacer sequence and binding to peptide insertion sites in the AAV8 library are depicted in Figure 23B. Secondary *in vivo* selection was performed for *cap* variants rescued from either lung or pancreas as the designated on-targets in two separate screens. For each of these screens, selection was either performed with the conventional non-depleted setting or with a previous Cas9-depletion to follow the DEPOOL protocol. Conventional selection was performed by (i) PCR-based “rescue” of *cap* gene variants found in the on-target tissue (lung or pancreas), (ii) sub-cloning into ITR-flanked rep-cap constructs, (iii) AAV library production, and (iv) biopanning by injection into NMRI mice. The DEPOOL protocol modifies this procedure by incubating the *cap* amplicons from step (i) with Cas9 and the sgRNA library, thereby cleaving and removing variants with matches to the sgRNA library. Although NGS-based readout demonstrated lower off-target presence for candidates from the depleted screen, an experimental validation of top candidates was necessary to determine the benefits of the DEPOOL procedure for candidate enrichment. All experiments from this point forward are part of the current study.

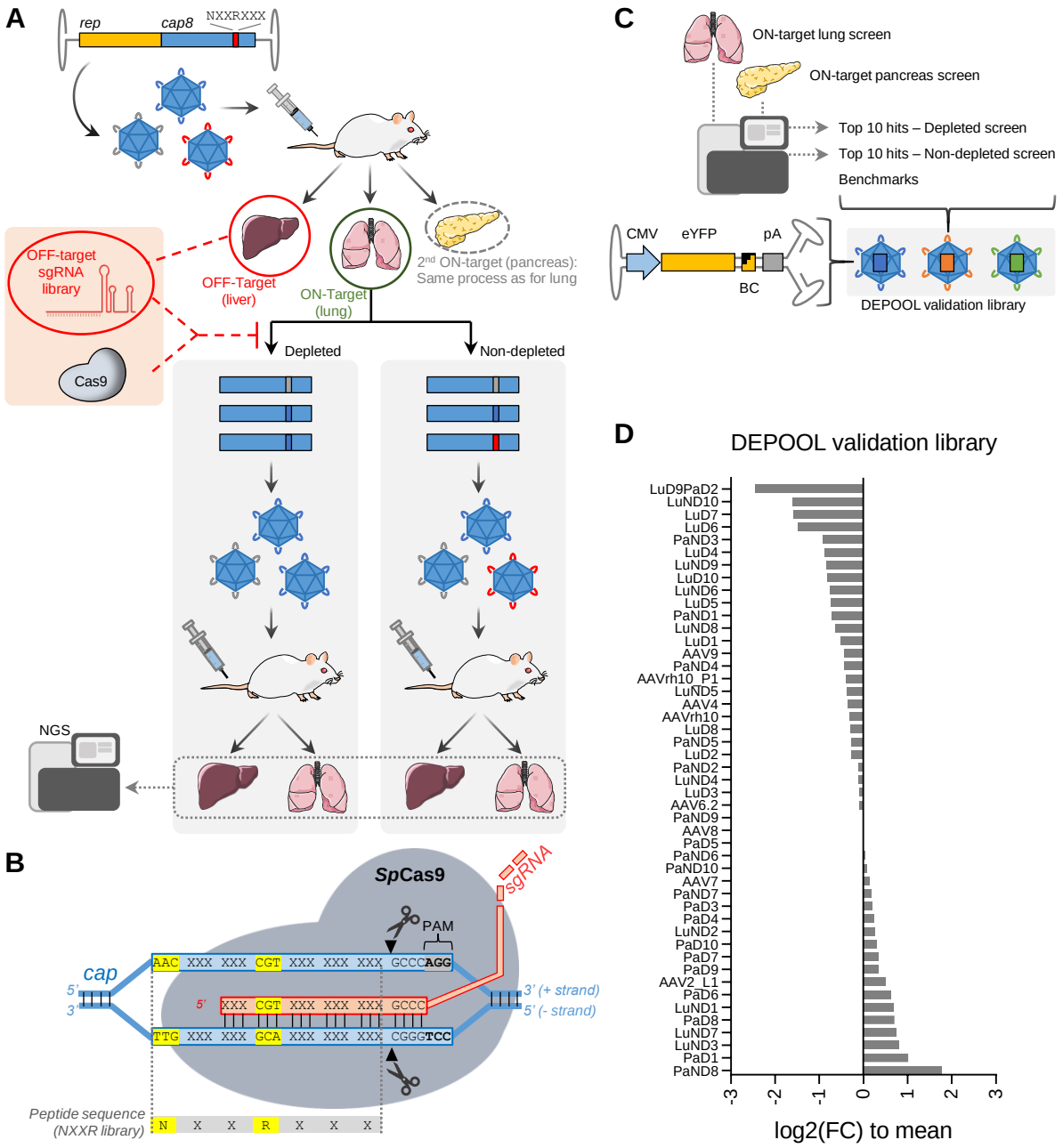



**Figure 23: Initial DEPOOL screen and candidate validation setting. (A)** Outline of the initial DEPOOL screen conducted prior to this study [275]. An AAV8 peptide library was screened in NMRI mice. *Cap* variants with peptide insertion enriched in lung and pancreas (on-target tissues) were rescued for a second selection round. From *cap*/peptide variants found in liver (off-target), a sgRNA library was generated for targeting of matching variants found in the on-target libraries. Secondary selection rounds were performed with the conventional non-depleted setting (sub-cloning of *cap* variants, vector production and biopanning) or with the depleted setting using prior Cas9-based removal of off-targeting variants using the sgRNA library. Variant enrichment was analyzed by NGS of peptide insertion. **(B)** Targeting of peptide insertion site in *cap* genes by matching sgRNA library. **(C)** Setting for the validation screen. The top ten candidates from each screen (depleted/non-depleted; selection in lung/pancreas) were selected together with benchmark capsids to generate a barcoded capsid library. **(D)** Composition of the DEPOOL validation library as determined by NGS analysis of barcode variants after vector production and library mixing.

To evaluate candidates enriched *via* DEPOOL and conventional pipelines, I selected the top ten most enriched variants found in on-targets (lung (Lu) or pancreas (Pa)) of each depleted (D) and non-depleted (ND) screen for a secondary validation experiment (Figure 23C). Since one candidate (LuD9PaD2) was found in both depleted lung and pancreas screens, the final selection panel included 39 capsid candidates. Following the procedure of Weinmann *et al*. [142], scAAV constructs with a CMV-eYFP-barcode cassette were packaged into each of the selected capsids as well as eight benchmark variants (AAV2_L1, AAV4, AAV6.2, AAV7, AAV8, AAVrh10 and AAVrh10_P1). Each transgene contained a variant-specific barcode in the 3'-UTR of *eYFP* to enable NGS-based readout on the DNA and RNA levels. I determined variant composition of the DEPOOL validation library by NGS (Figure 23D). This demonstrated that only LuD9PaD2 and PaND8 were outliers with especially low or high contributions, respectively, while most other variants showed similar ratios.

Next, the DEPOOL validation library was systemically administered to five NMRI mice (Figure 24A). After two weeks, mice were sacrificed for collection of liver, lung, pancreas, spleen, muscle, and kidney tissue. I quantified barcode compositions in gDNA (Figure 24B) and cDNA (Figure 24C) of these tissues and normalized these to (i) their relative abundance in the input library (Figure 23D) and to (ii) bulk vector biodistribution (Figure 24D). Highest abundance in lung gDNA and cDNA was achieved with the benchmark capsid AAV2_L1, a variant evolved for specific lung endothelial transduction [149]. Several candidates from both lung and pancreas screens outperformed benchmarks other than AAV2_L1 and AAV4 for lung transduction. Pancreas transduction was mostly lower for candidates as compared to benchmarks, but two candidates from the depleted pancreas screen (PaD1 and PaD3) outperformed all other capsids in hepatic expression except for AAV9. Both PaD1 and PaD3 still showed higher expression in liver than pancreas.

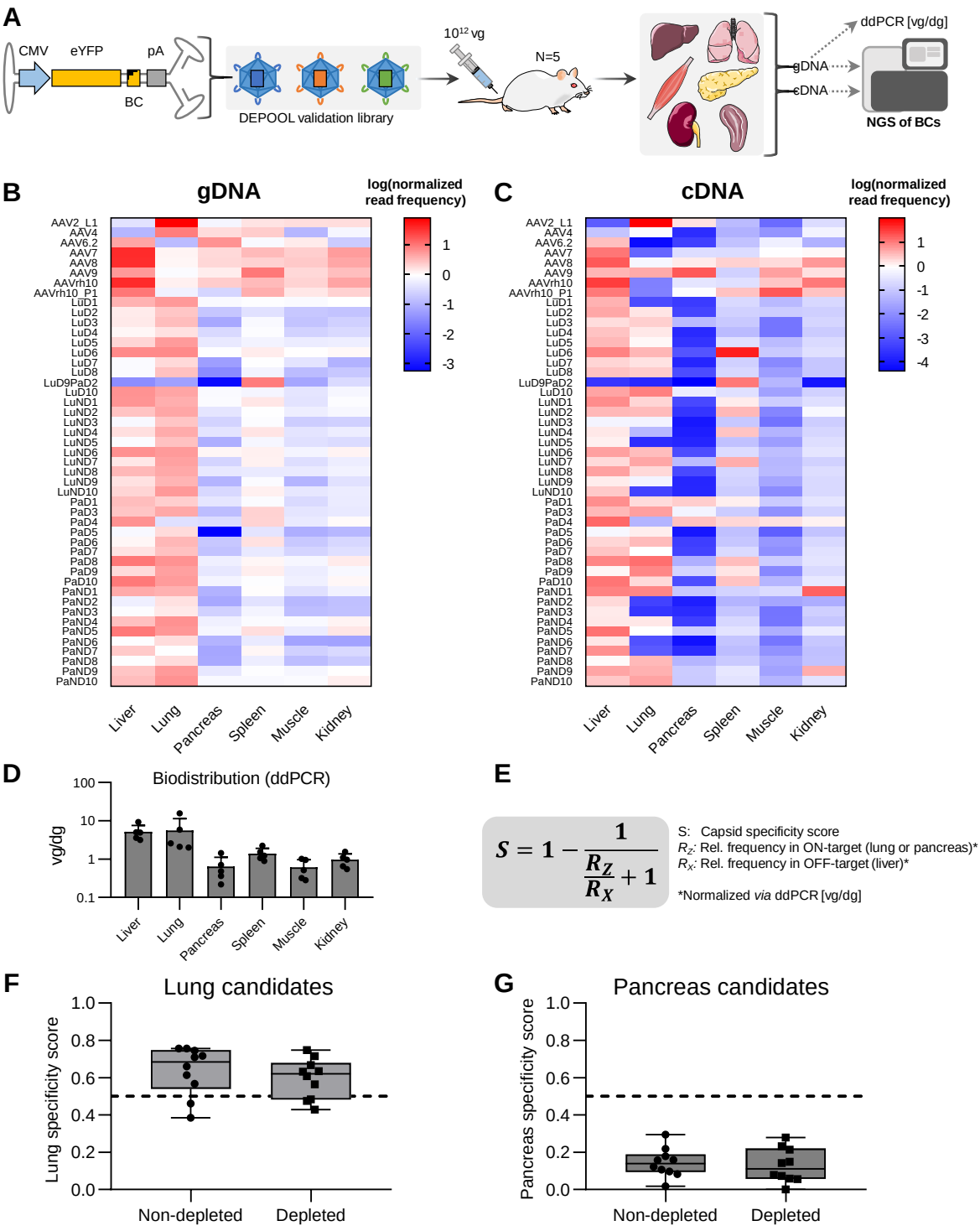



**Figure 24: Validation of initial DEPOOL screen does not show liver-detargeting for top candidates. (A)** The DEPOOL validation capsid library was injected into five NMRI mice with $10^{12}$ vg per mouse. Mice were sacrificed two weeks after injection for harvesting of liver, lung, pancreas, spleen, muscle (quadriceps femoris) and kidney tissue. **(B-C)** Relative variant frequency in cDNA and gDNA after normalization to the input library and to bulk vector distribution. **(D)** Bulk vector biodistribution [vg/dg] measured from gDNA by ddPCR. **(E)** Formula for calculation of capsid specificity scores S according to Körbelin *et al*. [149]. **(F-G)** Specificity scores of top 10 candidates for lung and pancreas from non-depleted and depleted screens. These compare vector presence in gDNA of on-target (lung/pancreas) vs. liver.

I calculated capsid specificity scores S according to Körbelin *et al*. [149] to compare variant frequency in on-target (lung or pancreas) vs. off-target (liver) (Figure 24E). This was calculated for vector presence in gDNA since Cas9-depletion and rescue were performed as selection on the gDNA level. Most lung candidates from both depleted and non-depleted screens showed S > 0.5, reflecting higher vector presence in lung than in liver and thus successfully achieving on-target selectivity (Figure 24F). However, all S scores of pancreas candidates were below 0.5, owing to higher vector presence in liver as compared to the pancreas on-target (Figure 24G). For both lung and pancreas candidates, no significant differences were found between S scores of non-depleted and depleted candidates. Since no improvement in on-target selectivity was found for DEPOOL candidates compared to conventionally enriched variants, a benefit of the DEPOOL procedure could not be verified.

#### *3.2.2.2 Improving DEPOOL parameters for peptide targeting*

Several reasons that may explain the negative outcome of the initial DEPOOL screen relate to sub-optimal selection conditions and will be addressed below. One problem was an incomplete coverage of the sgRNA sequences to the peptide insertion sites due to the use of a distant protospacer adjacent motif (PAM). This left an entire variable codon unmatched by the sgRNAs and thus contributed to a substantial non-specific depletion (Figure 23B). To improve sgRNA coverage of peptide insertion sites, cleavage was next attempted for a different peptide library design. This time, the peptide insertion site within the cap6-536 library was targeted, which contains a proline-flanked randomized 7mer peptide instead of the NXXRXXX peptide design used for the initial DEPOOL screen. For the cap6-536 library, sgRNAs can be designed to cover the entire randomized peptide insertion site, by using a directly neighboring PAM sequence within the 5’ proline codon (Figure 25A).

I tested the functionality of the sgRNA design for targeting of the cap6-536 peptide insertion site by using a plasmid cleavage assay. To this end, a target plasmid (cH001 peptide variant in CMV-cap6 backbone; linearized with HindIII prior to cleavage) was incubated with Cas9 and *in vitro* transcribed sgRNAs. This demonstrated correct cleavage patterns for incubation of cH001 plasmid with cH001 sgRNA, yielding cleavage bands of 5.1 kb and 2.1 kb for the 7.2 kb linear plasmid (Figure 25B). Incubation with the non-matched cH002 sgRNA did not produce cleavage bands, indicating specificity of target cleavage. As expected, incubation with a mix of nine sgRNAs (including cH001 and

cH002) also produced a cleavage pattern for the cH001 plasmid, although the cleavage bands were less prominent as compared to incubation only with the cH001 sgRNA.

Another problem with the initial DEPOOL screen was a suboptimal depletion protocol. For the initial screen, large quantities of *cap* amplicons were incubated with Cas9 and sgRNAs for depletion, followed by purification of non-cleaved amplicons through agarose gel electrophoresis. This mostly yielded DNA products for subcloning with poor quality and low quantity. A different Cas9-depletion approach was introduced by Hardigan *et al*.[280] for removal of unwanted variants from RNA sequencing libraries. I applied this method to the cap6-536 library using the above-mentioned sgRNAs (Figure 25C). Instead of applying Cas9 cleavage to large quantities of PCR amplicons, depletion was integrated into an amplification pipeline by performing (i) pre-amplification for only five PCR cycles, (ii) Cas9 cleavage followed by removal of Cas9 and sgRNAs, and (iii) final amplification (35 cycles). By using this protocol, over-amplification of unwanted fragments is prevented, resulting in a depletion of these variants from the amplified pool. This was tested using primers with Illumina overhangs to enable direct submission of amplicons for NGS analysis of variant frequency. Depletion was performed with either no sgRNA, the cH001 sgRNA or the 9x sgRNA mix (Figure 25D). As expected, amplicons from Cas9 depletion showed the same size as amplicons from a control amplification (-Cas9, -sgRNA). NGS analysis of treated amplicons (+Cas9, no sgRNA / cH001 sgRNA / 9x sgRNA mix) demonstrated highly effective depletion of targeted variants for either cH001 only or all nine targeted variants within the 9x sgRNA mix (Figure 25E-H). Depletion of non-targeted variants was not observed, demonstrating the applicability of both sgRNA design and depletion protocol for efficient and specific variant removal within peptide display libraries.

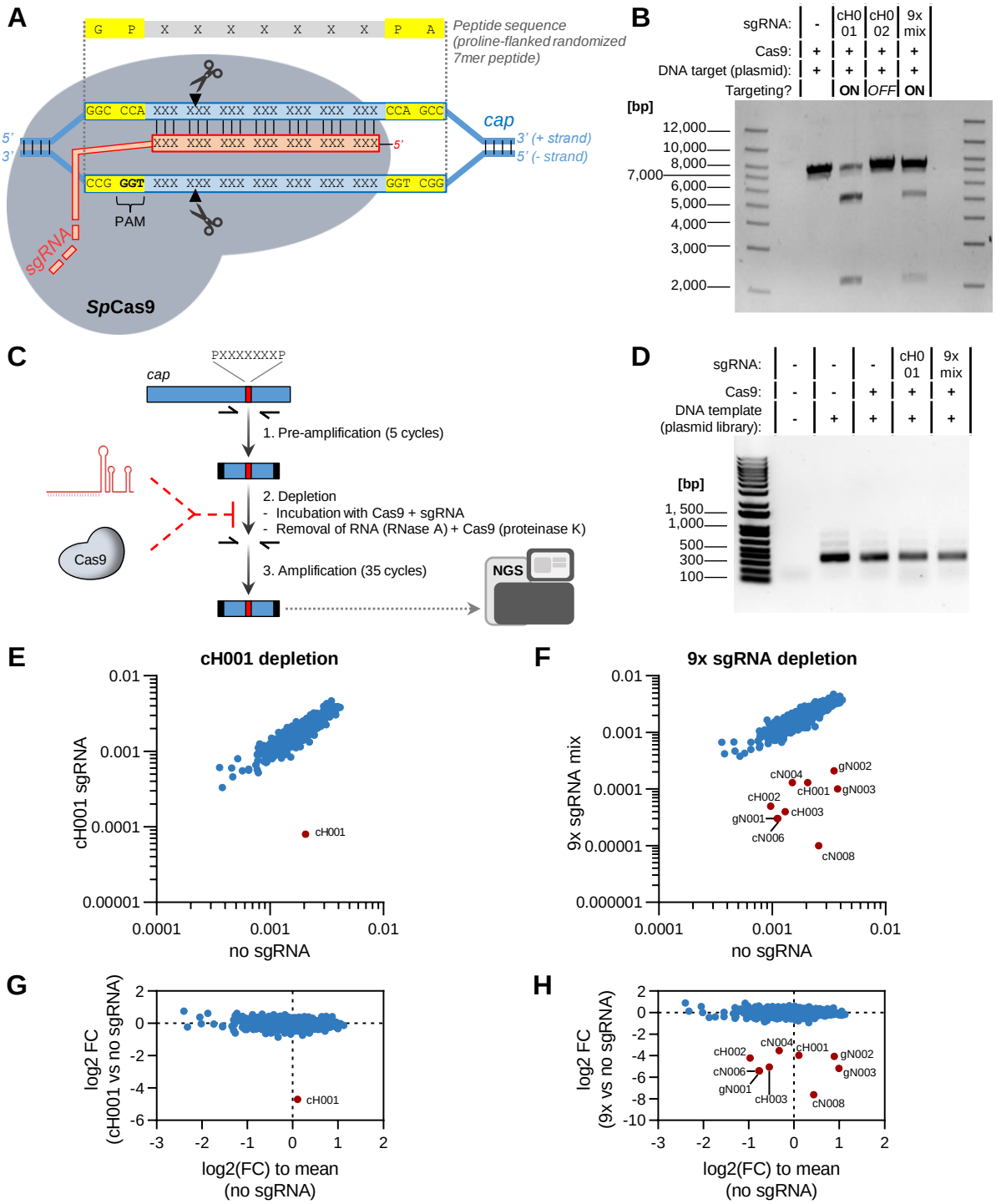



**Figure 25: Advanced targeting of peptide insertion sites for DEPOOL applications. (A)** Design of sgRNA spacer and binding to peptide insertion in the cap6-536 library. **(B)** Agarose gel after plasmid cleavage of cH001 (cap6-536 variant) with Cas9 and sgRNAs matching cH001 (targeting), cH002 (not targeting) and 9x sgRNA mix including cH001. **(C)** Pipeline for Cas9-based depletion during PCR amplification. Here, PCR amplification is used to generate NGS amplicons, adding Illumina adapters through primer overhangs (black bars in amplicons). **(D)** Agarose gel after pipeline from (**C**) conducted for cap6-536 library using sgRNAs for cH001 only or a mix of nine sgRNAs ("9x mix"). Amplicons from +Cas9/+template were submitted to NGS (-sgRNA, cH001 sgRNA or 9x mix sgRNA). **(E-F)** Variant frequencies of cap6-536 libraries compared between incubation with no sgRNA and cH001 sgRNA (**E**) or 9x sgRNA mix (**F**). **(G-H)** Log2-fold change (FC) after incubation with cH001 (**G**) or 9x sgRNA mix (**H**).

### 3.2.2.3 Targeting of barcode sequence in 3'-UTR of cap genes

Apart from AAV peptide display, other methods such as DNA family shuffling, variable region (VR) shuffling and random mutagenesis also enable diversification of AAV capsids into libraries for subsequent directed evolution [124]. As DEPOOL was so far only applicable for targeting of peptide insertion sites, a combination with other diversification methods that do not rely on randomized insertions with adjacent PAM sequences was not possible. To address such other methods as well, and to generalize the DEPOOL approach for other directed evolution applications, a new sgRNA target site was introduced into the *rep-cap* library selection backbone. By adding a barcode sequence into the 3'-UTR of *cap*, and linking *cap* identity to barcode sequence as was done by Pekrun *et al*. [141], a removal of unwanted *cap* variants is feasible through targeting of their associated barcode sequences with DEPOOL (Figure 26).

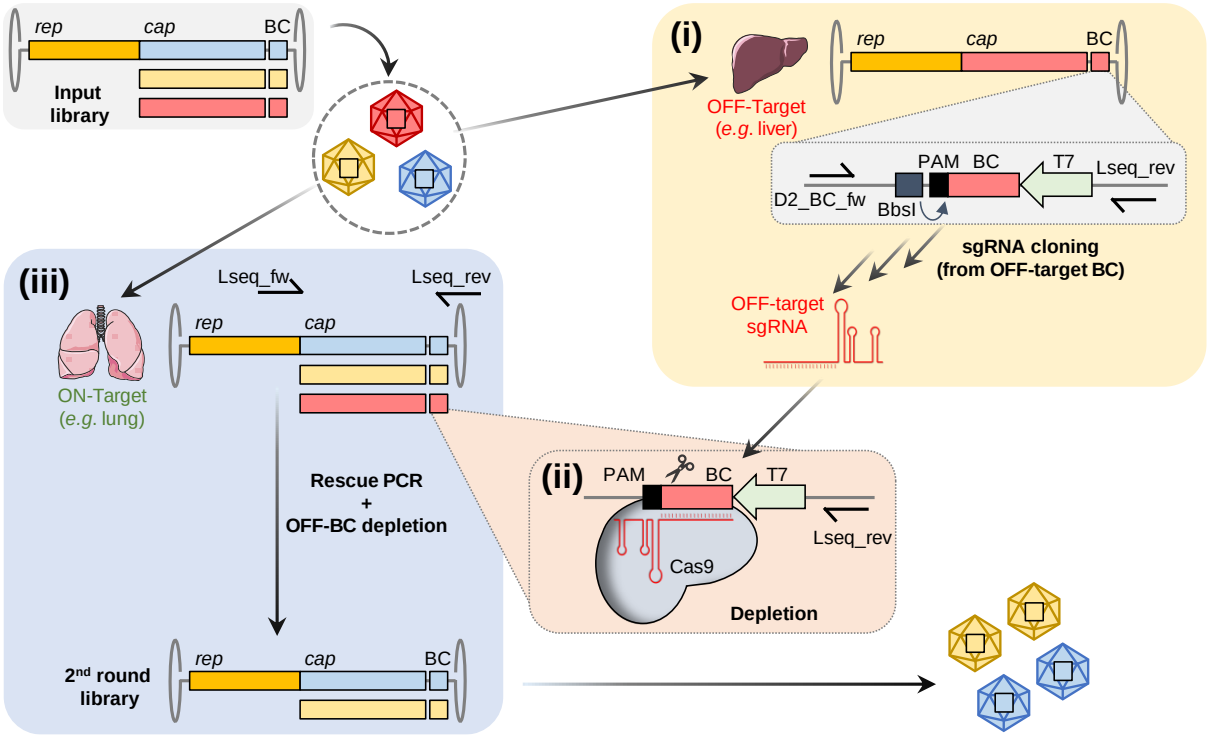



**Figure 26: Concept for depletion of variants based on barcode sequences.** An AAV *cap* library is generated where *cap* variants are associated with a specific barcode in the 3'-UTR of *cap*. The virus library is screened *in vivo* for enrichment in on- and off-target tissues. (**i**) From variants found in the off-target tissue, a sgRNA library is cloned. Within this library, each sgRNA variant is capable of binding to its respective cloning template for Cas9-based depletion thereof (**ii**). A detailed cloning procedure for the generation of self-targeting sgRNAs from barcode templates is depicted in Figure 5 (Methods section). To create a secondary screening library, enriched *cap* variants are recovered from on-target tissue by rescue PCR (**iii**). Variants within this pool that were also enriched in the off-target can be depleted through incubation with Cas9 and the off-target sgRNA library (**ii**).

An issue for the initial DEPOOL screen was the quality of the sgRNA library, which only poorly reflected variant distribution within the off-target tissue. However, this parameter is crucial for the simultaneous depletion of multiple variants, ideally based on their relative abundance in the off-target pool. To optimize cloning of sgRNAs from barcodes that are capable of targeting their own respective template, I designed sgRNA-targeted barcodes for the new DEPOOL iteration as follows (Figure 26 section i and ii). First, 21 nt barcodes were cloned with an adjacent 5' PAM site and an upstream BbsI site for restriction-based removal of the PAM during sgRNA cloning. Downstream of the barcode, a T7 promoter was placed for *in vitro* transcription of sgRNAs. This construct design allows cloning of a sgRNA library from off-target variants without a need for NGS readout or synthesis of oligonucleotide pools. A detailed procedure for sgRNA cloning from such templates can be found in the Methods section and is additionally depicted in Figure 5.

I tested cloning of sgRNAs from barcode templates and targeting thereof for three exemplary barcode sequences (BC1-3) inserted in a rep-cap2 AAV construct (Figure 27A). Following individual vector production, I assembled mini-libraries by combining BC1:BC2:BC3 in a ratio of 1:1:1 (L1) or 10:1:1 (L2). To test rescue and depletion, Hek293T cells were transduced with individual or pooled vectors followed by extraction of gDNA three days later. I again applied the PCR-based depletion protocol from above, with the modification of using primers to amplify the full *cap*-BC sequence (Figure 27B). This allowed sub-cloning of *cap*-BC and subsequent readout by Sanger sequencing. Rescue of *cap*-BC2 from gDNA of transduced Hek293T cells validated the strategy by showing an almost complete block of amplification with a BC2-sgRNA but not a control BC1 sgRNA (Figure 27C). The same result was found for the BC1 and BC3 templates, with depletion (block of amplification) only occurring for matching barcode and sgRNA combinations (Figure 27D). In the library setting, depletion with BC1 sgRNA still generated sufficient amplicons for sub-cloning (Figure 27E). Sanger sequencing was used for readout of barcode variants in bacterial colonies derived after sub-cloning from BC1-depleted or non-depleted L1 and L2 libraries (Figure 27F). This showed a successful depletion of the targeted BC1 variant. Overall, the DEPOOL procedure was found to be applicable, in principle, to barcode sequences as well as peptide insertion sites. A targeted removal of variants from *cap* libraries is thus possible, offering a novel means of applying negative selection pressure during iterative selection.

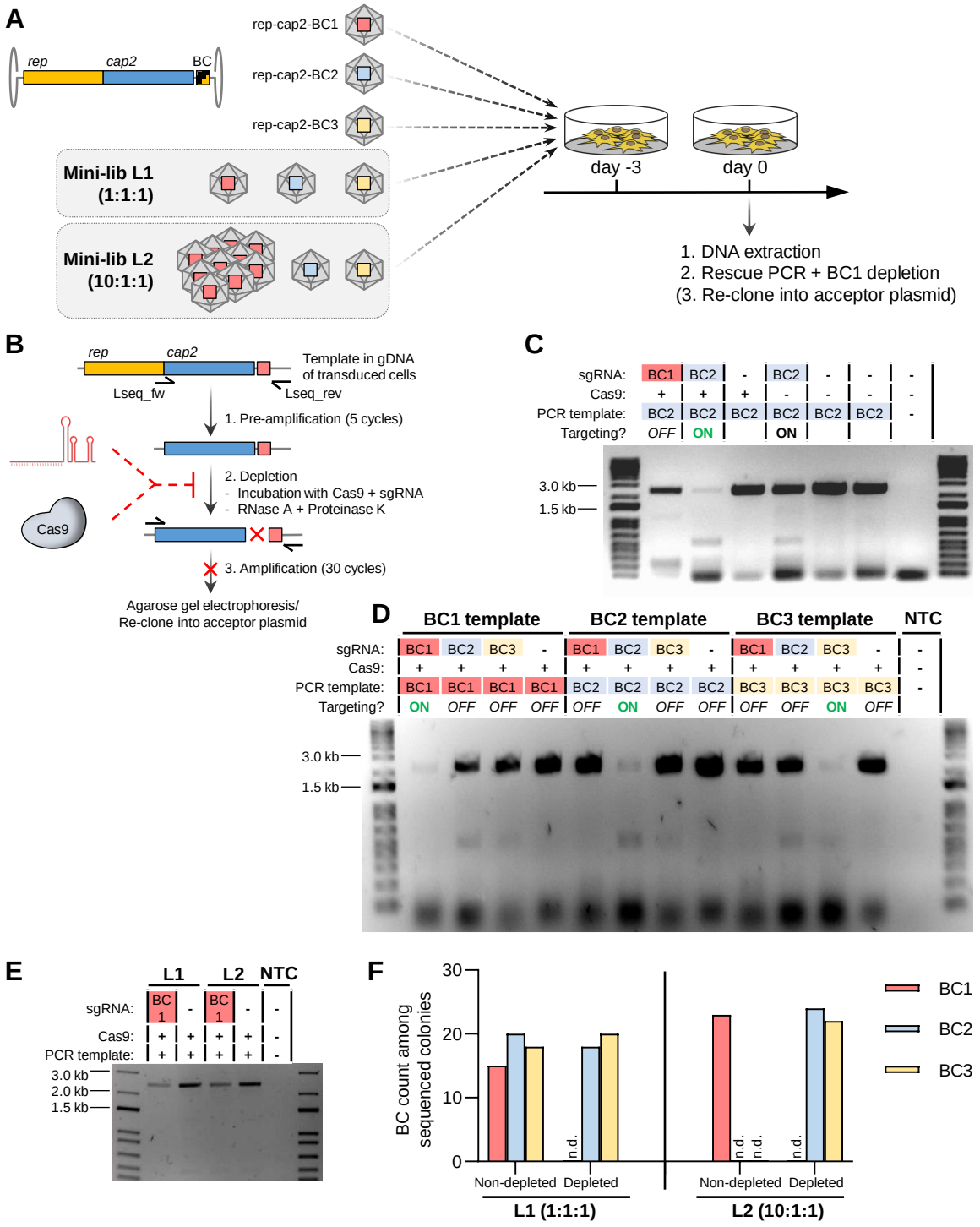



**Figure 27: Depletion of targeted barcode variants with DEPOOL in a rescue-PCR context.** **(A)** Experimental setup. Three different barcodes (BC1-3) were cloned into a rep-cap2 backbone for individual AAV production. Libraries were mixed at ratios of 1:1:1 (L1) or 10:1:1 (L2). Hek293T cells were transduced with individual vector constructs or libraries L1/L2 (MOI of $10^4$). Three days later, DNA was extracted for rescue PCR and depletion of cap-BC constructs. **(B)** Depletion/rescue protocol for sgRNA-based targeting of barcode sequences. **(C)** Agarose gel of amplicons after test depletion for rep-cap-BC2. **(D)** Test depletion for rep-cap-BC1/-2/3, using sgRNAs for each target. **(E)** Depletion of libraries L1 and L2 with or without BC1 sgRNA. **(F)** BC1-depleted and non-depleted amplicons from L1 and L2 were sub-cloned. Barcode variants in bacterial colonies were identified *via* Sanger sequencing.

### 3.2.3 RNA-based capsid evolution in mouse liver cells

Multiple successful examples have shown that directed evolution of AAV capsids can generate highly potent vector candidates [124]. While DEPOOL can contribute to directed capsid evolution by removing unwanted variants, it is still limited to selection on DNA level. If selection of enriched variants is performed by PCR rescue from genomic DNA in target tissue, then rescued capsid variants will likely have an advantage for DNA deposition. This does not automatically translate to correct cellular uptake, trafficking to the nucleus, and transgene expression for the DNA-enriched variants, all of which are required for functional transduction [154]. To achieve a more functional selection, several groups have started to perform *cap* selection on the mRNA level [74,146,147]. Tracing in RNA offers readout only for variants that have successfully reached the nucleus and induced transgene expression. Importantly, as reasoned in the DEPOOL chapter, transduction of off-target tissue must be monitored as well to prevent the selection of unspecific vector candidates. Thus, my aim was to explore a directed evolution system for AAV capsids that allows functional readout from the mRNA level in both on- and off-target cells by employing a ubiquitous expression of the *cap* gene. I implemented and tested this through (i) *in vitro* evaluation of different *cap* selection constructs, (ii) utilization of such construct for primary RNA-based screening of a *cap6* peptide library *in vivo*, and (iii) secondary *in vivo* screening across different liver cell types.

#### *3.2.3.1 Exploration of mRNA-based functional capsid selection constructs*

Wild-type-like ITR-flanked rep-cap constructs that are commonly used for capsid selection do not express sufficient levels of the *cap* gene in the absence of helper virus co-infection. A modification of the virus selection construct is therefore necessary to boost *cap* gene expression *in vivo* and enable rescue of *cap* variants from total mRNA. I modified the conventional *rep-cap* construct by replacing *rep* with different functional elements with the goals of (i) boosting *cap* mRNA expression after transduction in the absence of a helper virus and (ii) adding a reporter expression cassette to allow FACS-based selection of transduced cells (Figure 28A). Six different constructs were evaluated in comparison to wild-type AAV2 (rep-c2, (1)). Two of them (BYRPc2 (2) and BYxCPc2 (3)) retain the endogenous AAV p40 promoter. BYRPc2 contains a reverse RSV-eYFP-bgh-pA reporter cassette and is similar to the FT platform published by Westhaus *et al*.[74] in that it allows an RNA- and reporter-based *in vivo* selection. Westhaus *et al*., however, used the SFFV promoter for reporter expression instead of RSV. The other p40 promoter-

containing construct, BYxCPc2, lacks a promoter for *eYFP* expression, but contains a combination of CMV and p40 promoters to drive *cap* expression. This *cap* expression cassette is similar to constructs used by Nonnenmacher *et al*. [146], although their *in vivo* selection employed CNS-specific promoters instead of the ubiquitous CMV promoter used in BYxCPc2. For three other constructs (BYxCc2S (4), BYECc2 (5) and MYECc2S (6)), the p40 promoter was removed as well, and *cap* expression was solely driven by the CMV promoter. Two of them (BYxCc2S (4) and MYECc2S (6)) include an SV40 polyadenylation signal akin to Tabebordbar *et al*. [147] (although they used a tissue-specific promoter instead of CMV for *in vivo* selection) to stabilize *cap* mRNA. The last construct (CAGc2M, (7)) does not contain a reporter cassette but uses the 1.6 kb CAG promoter and a minimal polyadenylation signal for *cap* expression. For each construct, vector packaging was tested including assessment of virus titers, capsid assembly, and genome integrity (Figure 28B-E). For the rep-c2 (WT) construct, production was achieved by co-transfection with the adenoviral helper plasmid, while all other constructs missing *rep* were produced by co-transfection with the pDGΔVP helper that contains adenoviral helper genes as well as *rep2*. Virus production was performed in a setting equivalent to *cap* library productions. Selection of *cap* gene variants from capsid libraries requires packaging of each *cap* variant into its encoded capsid to create a genotype-phenotype linkage. As these variants are usually present in diverse plasmid libraries during transfection, self-packaging must be promoted by limiting the number of available plasmid copies per cell. As discussed and tested by others [139,296], limiting library-transfection to 5,000 plasmid copies per cell (or less) can achieve sufficient genotype-phenotype linkage. Thus, this setting was used for vector production with the above-mentioned *cap2* constructs.

Titration after vector purification demonstrated reduced vector quantity for RNA-selection constructs as compared to rep-c2 (WT), with a 71-90% reduction in yield after removing the *rep* gene (Figure 28B). A native dot blot with normalized vector concentrations (diluted based on titration) demonstrated similar quantities of assembled capsids (detected by A20 antibody) (Figure 28C). To test whether the modification of the *cap* expression cassette would alter ratios of VP1:VP2:VP3 capsid proteins in assembled capsids, a silver stain of PAGE-separated capsid proteins was conducted (Figure 28D). This demonstrated similar ratios compared to wild-type (rep-c2) and thus indicated assembly of functional capsids. Extraction of genomes from purified vectors and analysis by native agarose gel electrophoresis did not indicate packaging of sub-genomic fragments for any

construct (Figure 28E). In summary, all tested constructs showed functional assembly and packaging, albeit at a lower level as compared to the wild-type control.

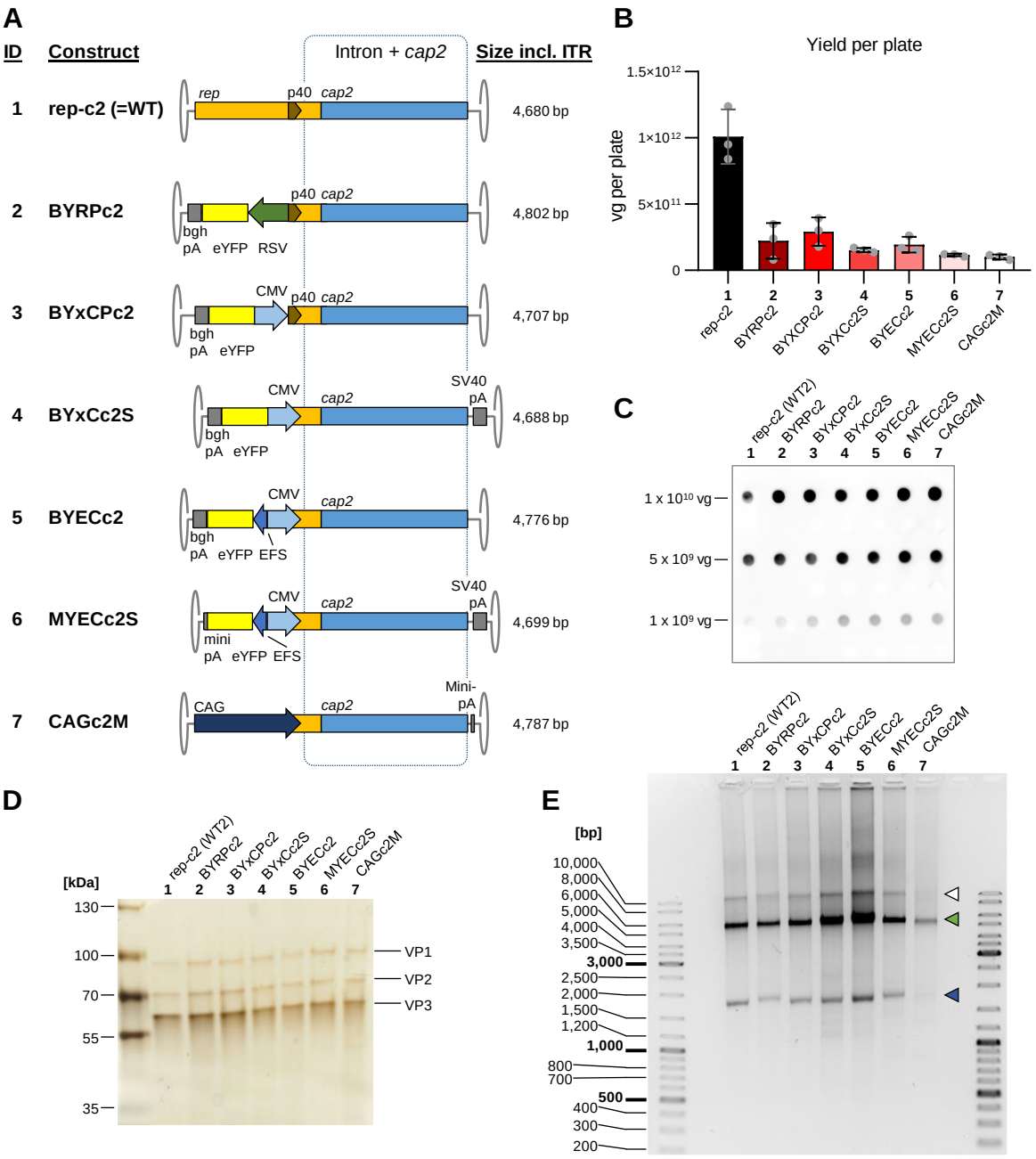



**Figure 28: Exploration of potential constructs for RNA-based selection of *cap* libraries. (A)** Design of constructs created for boosted *cap* mRNA and *eYFP* reporter expression. From the AAV2 wild-type (WT) construct, the *rep* gene was partially replaced by functional components for *eYFP* reporter expression and/or CMV or CAG promoters to boost *cap* expression. The p40 promoter was either retained or replaced. All constructs contain the full AAV intron and *cap2* gene. **(B)** Titration of purified AAV vectors conducted with ddPCR using a *cap2* primer/probe set. Shown are titers from three separate productions/purifications per construct, each done with 15 cm dishes ("plates"). **(C)** Native dot blot for assembled AAV2 capsids detected with the A20 antibody. Each construct was loaded at $1\times10^9$, $5\times10^9$ or $1\times10^{10}$ vg. **(D)** Silver staining of $5\times10^9$ vg of purified AAV constructs. Capsid proteins VP1, VP2 and VP3 are indicated. **(E)** Native agarose gel of AAV vector genomes extracted after virus production and purification. Single- (blue arrow), double-stranded (green arrow) and concatemeric (white arrow) DNA bands are indicated.

Next, I tested the ability of these constructs to achieve expression of *cap2* and *eYFP* reporter. This was conducted by transfection of Hek293T cells, as well as transduction of Hek293T, Hepa1-6, Huh7 and Caco2 cells followed by RTqPCR analysis of *cap2* and *eYFP* expression three days post-treatment (Figure 29A-B). An EFS-eYFP vector (Figure 29C) was included to serve as a positive control for *eYFP* expression and negative control for *cap2* expression. In transfected Hek293T cells, all constructs missing *rep* outperformed the rep-c2 wild-type control for expression of *cap2* (Figure 29A). This indicates the intended boost of *cap* expression necessary for RNA-based screening. However, transduction in different cell types showed a different trend, with *cap2* expression being barely higher and, in several instances, even lower for the modified constructs as compared to rep-c2. Especially BYRPc2, where *cap2* is only driven by the endogenous p40 promoter, demonstrated poor *cap2* expression after transduction although showing ~3× higher *cap2* expression than rep-c2 after plasmid transfection. Highest *eYFP* expression was observed for the BYECc2 and MYECc2S constructs, with BYECc2 partially exceeding the EFS-eYFP control (Figure 29B). Constructs missing a promoter for *eYFP* (BYxCPc2 and BYxCc2S) still showed low levels of *eYFP* expression, indicating a reverse activity of the CMV promoter.

To investigate why BYRPc2 showed lower *cap2* expression than rep-c2 after transduction despite carrying the same *cap*2 expression cassette, I followed two different hypotheses: (i) the *eYFP* cassette may interfere with *cap2* expression in BYRPc2, or (ii) Rep proteins that are missing in BYRPc2 may benefit *cap2* expression in the rep-c2 construct after transduction. To address these questions, I cloned new constructs (Figure 29D) including one in which (i) the RSV promoter or the entire RSV-eYFP-pA cassette in BYRPc2 were replaced with inert *lacZ* stuffers, yielding BYzPc2 and zPc2. This did not alter the titer after vector production compared to BYRPc2 (Figure 29E). Transduction of Hek293T cells did also not show a significant difference in *cap2* expression between BYRPc2 and *lacZ* derivates (Figure 29F), indicating that the *eYFP* cassette has no major influence on *cap2* expression of BYRPc2. To see whether Rep proteins limit *cap2* expression, I created a second construct that (ii) only expresses Rep proteins but not *cap2* (rep-z, Figure 29D). However, co-transduction of BYRPc2 or zPc2 with rep-z did not increase *cap2* expression (Figure 29F). Thus, neither hypothesis could explain the reduced *cap2* expression observed for BYRPc2 after transduction.

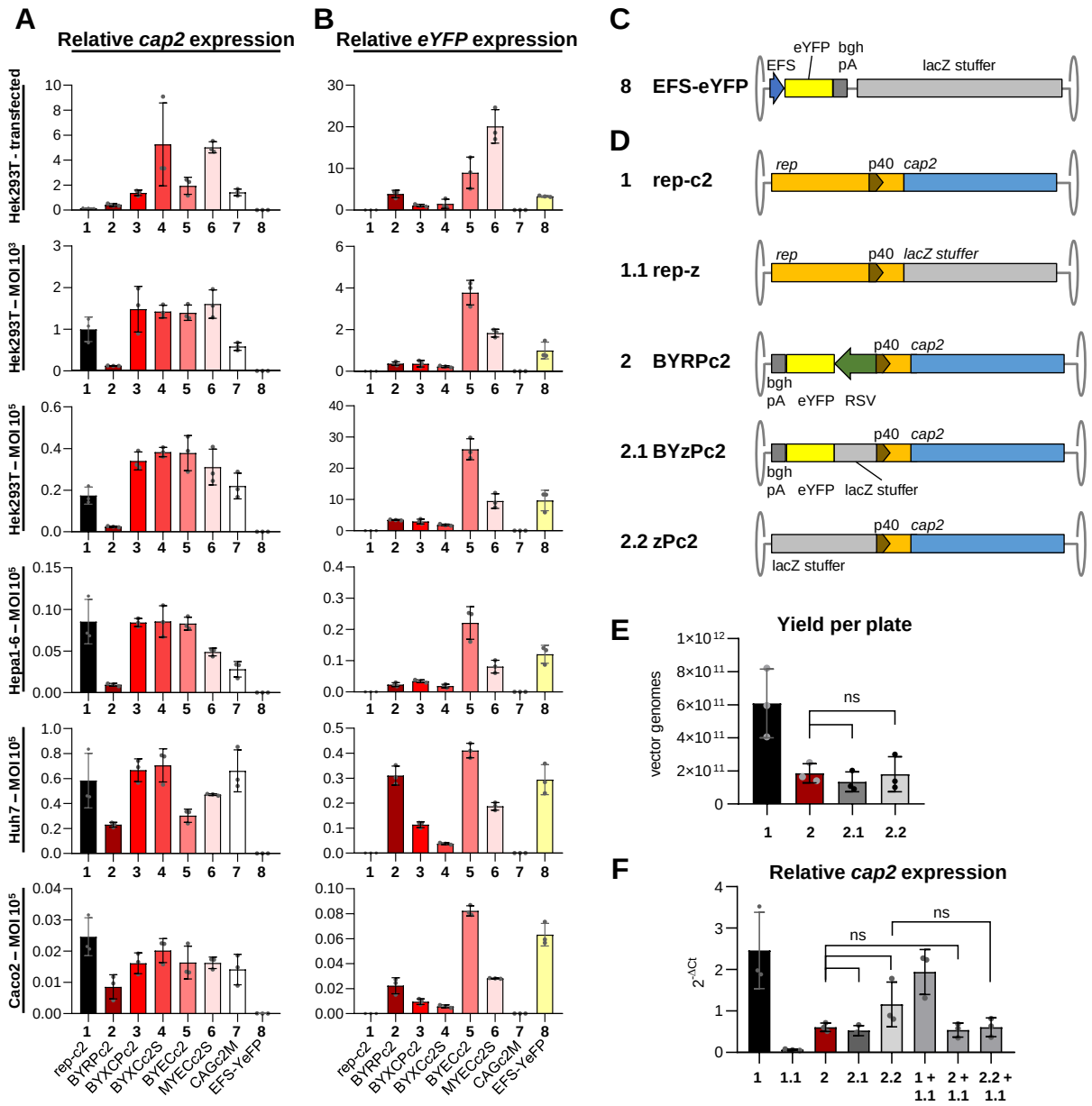



**Figure 29: CMV-*cap* vector constructs demonstrate functional transduction but non-uniform boost in efficiency of *cap* expression. (A-B)** Transfection of Hek293T cells or transduction of Hek293T, Hepa1-6, Huh7 or Caco2 cells. Relative expression of *cap2* (**A**) or *eYFP* (**B**) compared to the *GAPDH* housekeeper was quantified with RTqPCR three days after treatment of cells. Expression is presented as $2^{-\Delta Ct}$, with bars indicating mean ± SD of three replicates. **(C)** EFS-eYFP control construct used for comparison of *eYFP* expression. **(D)** Constructs rep-c2 (1) and BYRPc2 (2) were modified by replacing *rep*, RSV or RSV-eYFP-pA with inert *lacZ* stuffers. **(E)** Yields per plate (15 cm dish) for AAV production of constructs rep-c2 (1), BYRPc2 (2), BYzPc2 (2.1) and zPc2 (2.2). Each construct was produced in triplicates with three plates per production. **(F)** Relative *cap2* expression in Hek293T cells transduced with rep-c2, BYRPc2 and derived *lacZ* stuffer constructs. Transduction was conducted in triplicates with an MOI of $5\times10^4$ vg per construct. Statistical analyses were conducted with one-way ANOVA with Dunnett's multiple comparison test.

MYECc2S, a construct that contains both a functional *eYFP* reporter cassette and a CMV promoter-driven *cap* with an SV40 polyadenylation signal for *cap* mRNA stabilization demonstrated superior *cap2* expression after Hek293T transfection combined with robust reporter expression. Thus, I selected this construct for subsequent experiments in which I evaluated mRNA-based *cap* selection. To test rescue of *cap* genes from mRNA, I transduced Hek293T cells with MYECc2S (MOI of $10^5$) and harvested cells three days later. RNA was extracted, followed by reverse transcription (RT) using either the High-Capacity cDNA Reverse Transcription kit, the SuperScript VILO kit or the SuperScript First Strand Synthesis kit with different RT primers (Figure 30A). A rescue PCR with cap2_fw and cap2_rev primers showed better amplification for the SuperScript VILO kit as compared to the High-Capacity Reverse Transcription kit. All tested RT primers showed similar results (Figure 30B). Amplification of the entire *cap* gene with Lseq_fw/rep2_resc_fw and cap2_rev primers was successful as well (Figure 30C). For the rep2_resc_fw primer, amplicons were detected for both spliced and non-spliced mRNA/cDNA templates (green and white arrows in Figure 30C, respectively). As the SuperScript VILO kit uses pre-mixed random hexamer and oligo-dT primers, I tested target-specific reverse transcription again for the SuperScript First Strand Synthesis kit with different RT primers (Figure 30D). After amplification with Lseq_fw and cap2_rev, no clear differences were visible between using oligo-dT or target-specific RT primers, but random hexamers showed a lower abundance of the target amplicon and a slightly higher abundance of unspecific PCR products. For all rescue PCRs, controls missing reverse transcriptase (“-RT”) did not produce target amplification, indicating successful and specific rescue from *cap* cDNA.

In summary, although *in vitro* transduction did not show a clear boost of *cap* expression with different constructs intended for functional selection on the mRNA level, a *cap* rescue from mRNA was feasible with the MYECc2S construct after transduction of Hek293T cells. Thus, MYECc2S may allow functional selection for directed capsid evolution by providing both ubiquitous *cap* expression and additional *eYFP* reporter expression for enrichment of transduced cells.

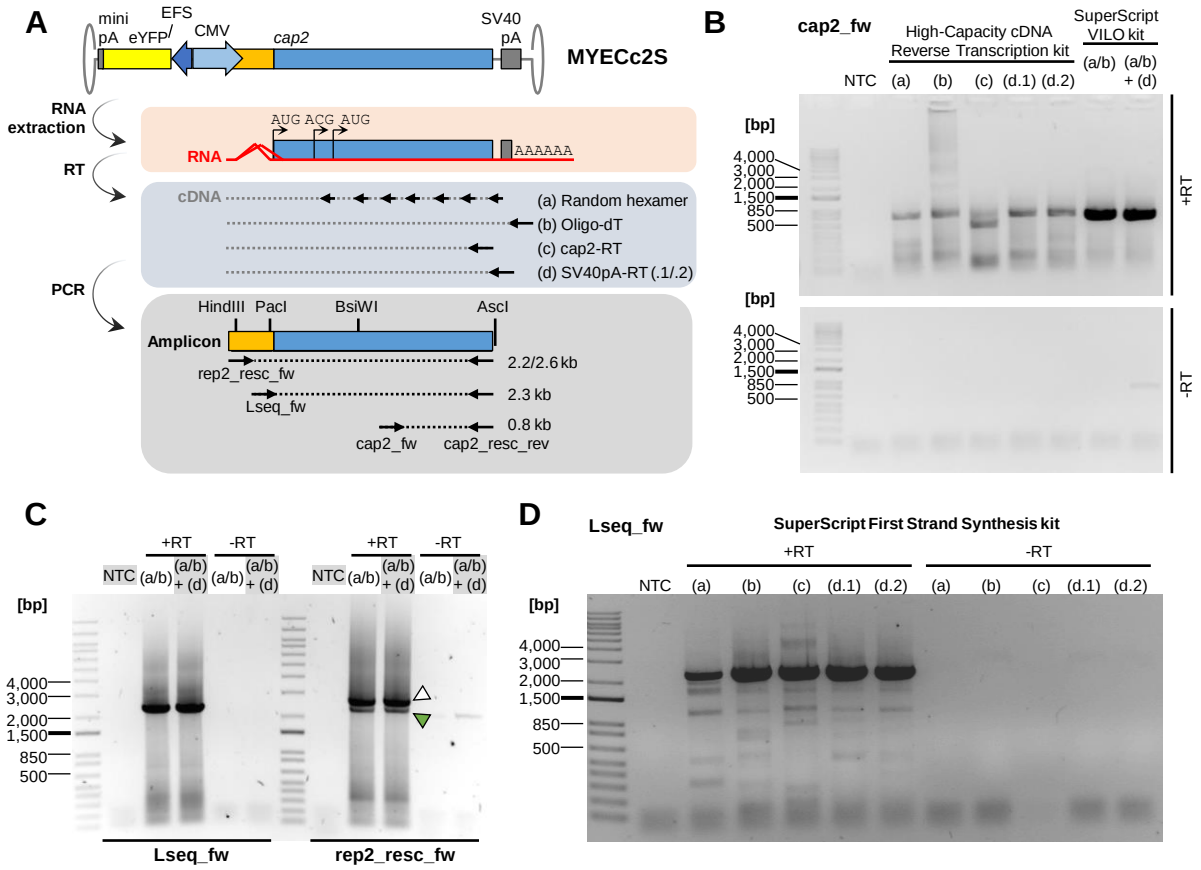



**Figure 30: Functional rescue of *cap* gene from RNA is possible for a CMV-*cap* construct.** **(A)** Hek293T cells were transduced with MYECc2S (MOI of $10^5$), followed by RNA extraction on day 3 post-transduction. Reverse transcription (RT) was conducted using (a) random hexamer, (b) oligo-dT, (c) cap2-RT or (d) SV40pA-RT primers. From the resulting cDNA, rescue PCR was conducted using rep2_resc_fw, Lseq_fw or cap2_fw primers with the cap2_resc_rev primer, yielding amplicons with the indicated sizes. For the rep2_resc_fw primer, amplicons can be either 2.2 kb for amplification of cDNA derived from spliced mRNA or 2.6 kb for cDNAs derived from non-spliced mRNA. **(B)** Agarose gel of cap2_fw rescue PCR amplicons. RT was conducted with the High-Capacity cDNA Reverse Transcription kit using RT primers a-d, or with the SuperScript VILO kit using either a mix of a+b RT primers (supplied by kit) or a+b+d RT primers. **(C)** Rescue PCR after reverse transcription with the SuperScript VILO kit by using Lseq_fw or rep2_resc_fw primers. **(D)** Rescue PCR with Lseq_fw PCR primer after RT with the SuperScript First Strand Synthesis kit and different RT primers.

### *3.2.3.2 RNA-based selection of a cap6 peptide library in vivo*

While most AAV capsids are capable of transducing liver tissue in general, cell-type-specific transduction is far more difficult. Vectors enabling selective transduction of individual non-parenchymal liver cell-types (NPCs) would be beneficial as tools that allow for the dissection of molecular pathways in these cells (*e.g.*, understanding the role of CD44 in the progression of non-alcoholic/metabolic steatohepatitis *via* cell-type-specific knockdown [260]) or for establishing therapeutic transgene expression only in defined target

cells. The goal of evolving capsids for specific transduction of NPC subtypes can be aided by diversifying a capsid already capable of general NPC transduction. As demonstrated by former lab member Jonas Weinmann in his dissertation [297], AAV6 allows efficient but non-specific NPC transduction. Thus, I chose the AAV6 capsid for diversification through random peptide insertion. To enable functional selection in murine NPCs *in vivo*, I employed the MYECc2S construct established above to enable tracing of *cap* variants in total mRNA.

To achieve this, I firstly replaced *cap2* in MYECc2S with the AAV6 *cap* gene (*cap6*) containing a peptide insertion site in VR VIII (described in Börner *et al*. [148]) (Figure 31A). A sequence coding for randomized heptameric peptides with flanking prolines was inserted into this construct, yielding the CMV-cap6-P7P library with a clonal diversity of $6.8\times10^6$ capsid variants (based on bacterial colony number). I then used this library for virus production and subsequent screening in murine liver, from which bulk hepatocytes and NPCs were isolated and separated one week after library injection (mouse work and cell isolation were conducted by Pervinder Choksi, Willenbring lab). Peptide variants were sequenced by NGS in plasmid and virus input libraries, as well as in gDNA and cDNA (mRNA) pools. Comparing variants between plasmid and virus libraries showed low overlap between the two pools (Figure 31B-C), indicating that the diversity of neither library was fully covered by high-throughput sequencing.

EFS promoter-driven *eYFP* expression was not sufficient to enable FACS-based selection of eYFP-positive cells (data not shown). However, vector presence and relative expression were detected for hepatocytes and NPCs. Relative *eYFP* expression and vector distribution were both higher for hepatocytes compared to NPCs (Figure 31D). Normalized expression (relative expression normalized to vector distribution) was equal between the two populations, indicating equivalent functional transduction. PCR amplification of peptide regions for subsequent NGS analysis was successful for both populations on the gDNA and cDNA level, and it produced no amplification for no-target control and -RT control (Figure 31E). This demonstrates the capability of the selected CMV construct to drive sufficient expression for variant monitoring in cDNA of on- and off-target cells.

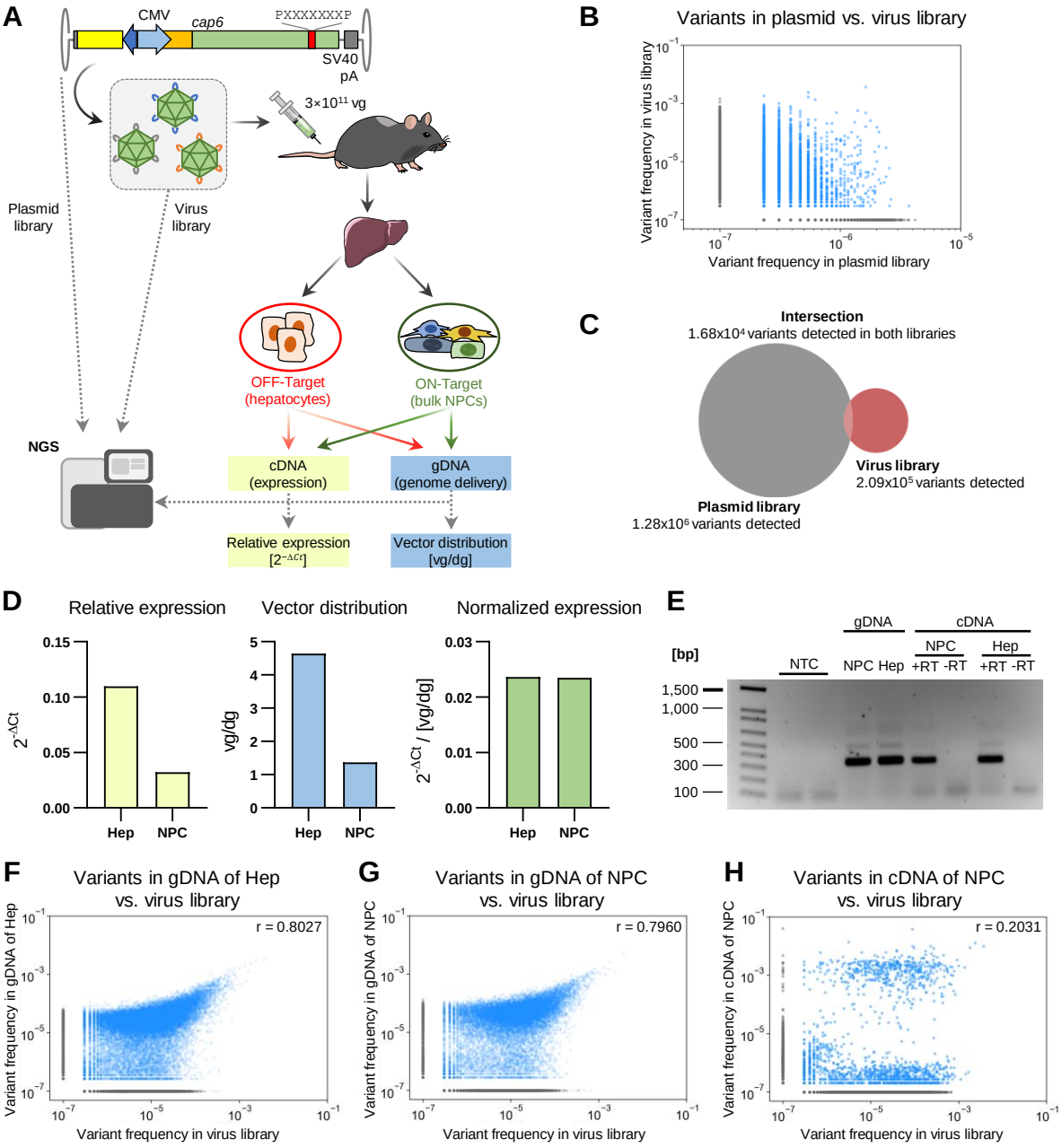



**Figure 31: *In vivo* screening of a CMV-cap6 peptide library on the DNA and RNA level. (A)** The CMV-cap6-P7P library was injected into an *Lrat-Cre*$^{+/-}$*;R26-RFP*$^{+/+}$ mouse followed by harvesting of the liver after one week and dissociation into bulk hepatocytes (off-target) and NPCs (on-target; mouse work by Pervinder Choksi, Willenbring lab). From both cell populations, cDNA and gDNA were extracted for variant tracing by NGS and bulk transduction quantification by RTqPCR and ddPCR. **(B)** Variant frequency compared between plasmid and virus libraries. Gray variants were only detected in one of the two pools. **(C)** Overlap between peptide variants detected in plasmid and virus libraries. **(D)** Bulk measurements of relative *eYFP* expression, vector distribution and normalized expression compared between hepatocytes (Hep) and NPCs. **(E)** Agarose gel after PCR amplification of peptide regions from gDNA and cDNA (± RT) of NPC and Hep. NTC, non-template control. **(F-H)** Comparisons of variant frequencies between virus library (x-axis) and gDNA pool of hepatocytes (**F**), gDNA pool of NPC (**G**), or cDNA pool of NPC (**H**) (all y-axis). Blue variants were detected in both respective pools, gray variants only in one of the two. r, Pearson correlation coefficient.

Comparing variant frequency between the input virus library and gDNA pools from hepatocytes and NPCs demonstrated a positive correlation between the abundance in input and the enrichment in gDNA (Figure 31F-G; Pearson r of 0.80 for both hepatocytes and NPC). For cDNA in NPCs, this correlation was mostly lost (Pearson r of 0.20), but two populations of variants arose instead, namely, one with high frequency in cDNA (*i.e*., high expression) and one with low frequency (Figure 31H). This already indicates that selection on the mRNA (cDNA) level distinguishes more clearly between functional transduction and mere vector genome deposition.

Next, I compared variant frequencies between gDNA and cDNA in order to evaluate functional transduction, *i.e*., the ability of a capsid to achieve vector delivery to the target cell, uptake, trafficking and ultimately transgene expression (Figure 32A). This comparison showed four clusters that appeared in both NPC and hepatocytes: (i) weak candidates with low vector concentration and low expression; (ii) variants with low expression (low frequency in cDNA) but high abundance in gDNA; (iii) an opposite cluster with high frequency in cDNA but low abundance in gDNA (*i.e*., good delivery but poor functional transduction); and (iv) variants with high frequency in cDNA as well as in gDNA. The goal of this screen was to identify variants that combine high transduction efficiency with on-target selectivity. Therefore, I then compared transduction in on- and off-target cells (NPCs and hepatocytes, respectively) (Figure 32B). For gDNA, four sub-groups with poor separation were detected (high or low frequency in NPC or hepatocytes). The group of variants with high frequency in both populations was dominant, and a positive correlation (Pearson r of 0.70) was observed between enrichment in hepatocytes and NPC. This was found to be different for cDNA, where clusters with high and low frequency in hepatocytes and NPCs were clearly separated, with only a minor fraction of variants exhibiting high expression in both groups. No positive correlation was detected between variant frequencies across these two cell populations (Pearson r of -0.14). Hence, analysis of variant frequency in cDNA offers a potent tool for selection of candidates with target specificity.

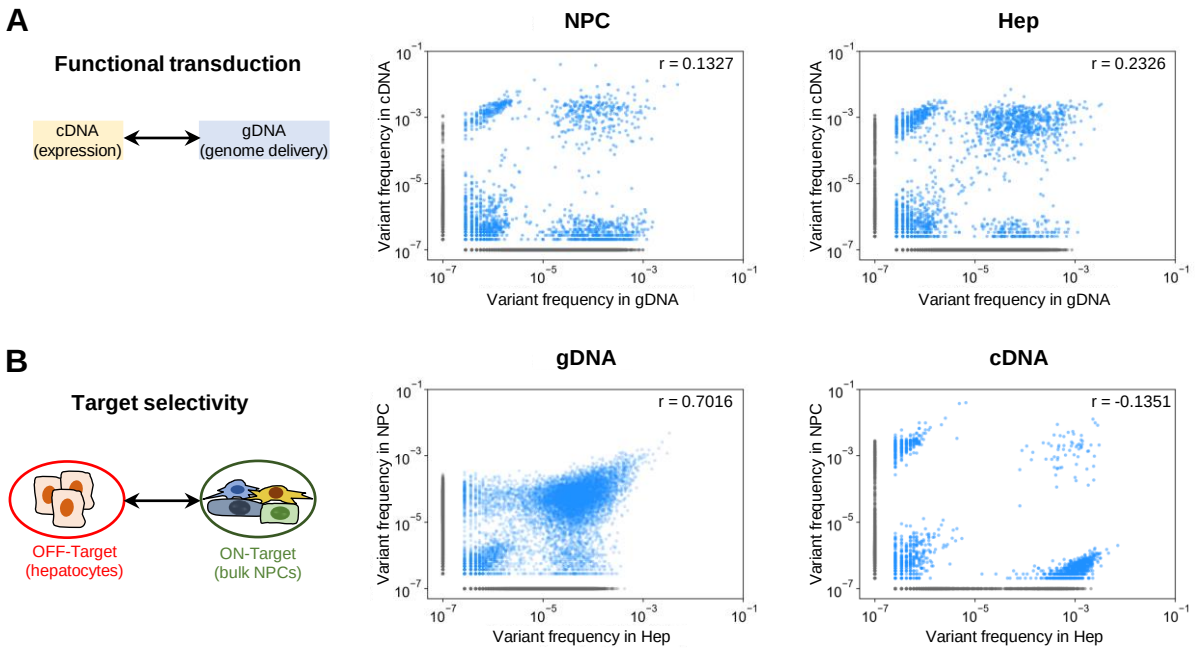



**Figure 32: Comparing functional transduction and target selectivity between DNA and RNA selection settings. (A)** Comparison of variant frequency between gDNA (x-axis) and cDNA (y-axis) for NPC (left) and hepatocyte (Hep, right) pools. **(B)** Comparison of variant frequencies between hepatocytes (Hep, x-axis) and NPC (y-axis) for gDNA (left) and cDNA (right). Blue variants were detected in both respective pools, gray variants only in one of the two. r, Pearson correlation coefficient.

While the CMV promoter-driven *cap* expression was sufficient to enable NGS-based analysis on the RNA level, an iterative selection by rescue of *cap* genes from cDNA would require larger amplicons and thus impose more challenges for the PCR reaction. To demonstrate that amplification from gDNA and cDNA is possible after *in vivo* selection, I tested rescue PCR conditions with a reverse *cap6*-specific primer and different forward primers (Figure 33A). The PCR was designed to amplify either only a region sufficient for sub-cloning of the peptide insertion site (cap6_resc_fw; amplicon of 1.2 kb), or to amplify the entire *cap* gene with forward primers binding within (Lseq_fw; 2.3 kb) or before the intronic sequence (rep2_resc_fw; 2.2 kb for spliced/2.6 kb for non-spliced template). Amplification of the peptide region was successful for both gDNA and cDNA, with no amplicons showing for the no-template and no-RT controls (Figure 33B). Non-specific amplification bands were visible for gDNA and cDNA, but all were less prominent than the 1.2 kb target amplicon. The rescue PCR was also successful for primer sets amplifying the entire *cap* gene (Figure 33C). However, rep2_resc_fw primer did not yield the expected spliced amplicon from cDNA (2.2 kb). In general, amplification from cDNA was weaker than from gDNA with a prominent secondary band. Overall, *cap* gene

amplification was successful and validated the ability of the CMV construct to enable *cap* rescue from gDNA as well as cDNA.

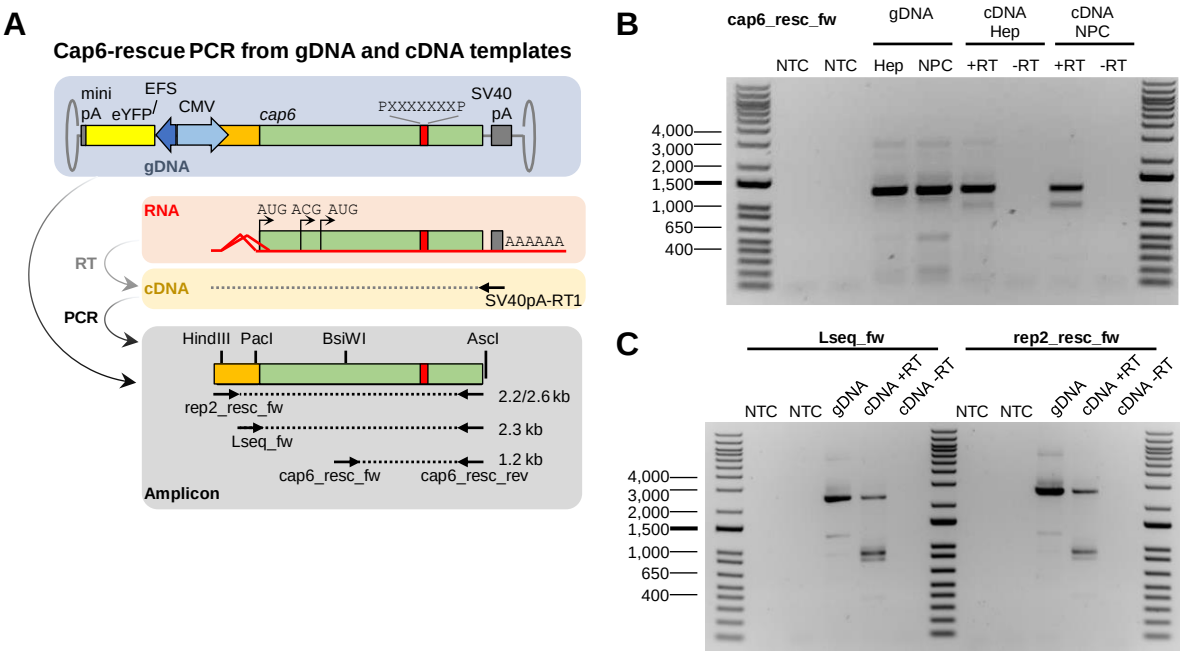



**Figure 33: CMV-*cap* construct allows rescue of *cap* library from RNA *in vivo*. (A)** Depiction of the CMV-cap6-P7P construct with *cap6-P7P* RNA, cDNA, and amplicons for different rescue PCR primers. **(B)** Agarose gel showing amplicons from the rescue PCR with cap6_resc_fw and _rev primers. Amplification was conducted from gDNA and cDNA templates from hepatocytes (Hep) and NPCs. **(C)** Amplicons after rescue PCR with Lseq_fw or rep2_resc_fw primers from gDNA and cDNA templates of hepatocytes. NTC, no-template control.

### *3.2.3.3 Generation and screening of a secondary peptide library*

While rescue PCR-based iterative selection is necessary for capsid diversification methods that include modifications across the entire *cap* sequence, diversification by peptide display is only modifying a small part of the *cap* gene. This offers the possibility to use peptide variant cloning *via* pooled oligonucleotide DNA synthesis instead of rescue PCR for cloning of secondary libraries. Here, selected peptide variants are synthesized as DNA oligonucleotide pools and cloned simultaneously into the *cap* backbone. This strategy offers several advantages, including (i) the possibility to hand-pick selected variants, (ii) relatively similar contribution of each variant as compared to large differences in frequencies after rescue PCR, and (iii) less technical error during subcloning.

As the initial selection round described above was highly stringent, the number of remaining lead candidates was low for CMV-cap6-P7P, thus limiting the pool size for oligonucleotide synthesis. To choose the most promising candidates (Figure 34A), I

selected variants by (i) ranking based on presence in on-target (NPC) cDNA, (ii) differentiating between variants with high vs. low expression, and (iii) identifying candidates with selectivity for NPCs (specific on-target hits, N=464) or non-selectivity (unspecific on-target hits, N=66). Although cloning by oligonucleotide synthesis offers the opportunity to retain only specific on-target hits, the 66 unspecific candidates as well as the top three non-functional on-target hits (high in NPC gDNA but not cDNA) and the top three off-target cDNA hits (high in hepatocyte but not NPC cDNA) were selected as well to serve as control candidates during secondary selection rounds. This resulted in a pool of 536 variants for which a pool of oligonucleotides was synthesized. I cloned this pool into the previously used CMV construct (derived from MYECc2S) as well as into the conventional wild-type-like *rep-cap* construct, yielding the CMV-cap6-536 and rep-cap6-536 plasmid libraries (Figure 34B). Virus production of the rep-cap6-536 library was performed by co-transfection of Hek293T cells with rep-cap6-536 and adenoviral helper plasmid in a 1:20 ratio (Figure 34C).

Notably, several groups have suggested that AAP expression can improve production for rep-less libraries [146,151]. This was tested for CMV-cap6-536 by co-transfecting the library with pDGΔVP and either with or without an additional AAP-expressing plasmid in a 1:20:10 ratio (Figure 34D). After purification, the final yield was $3.7\times10^{10}$ vg/plate for the rep-cap6-536 library, $8.1\times10^{9}$ vg/plate for the CMV-cap6-536 library without AAP (22% of rep library), and $2.1\times10^{10}$ vg/plate for the CMV-cap6-536 library with additional AAP (58% of rep library; n=1 sample per condition). Although not tested in replicates, this preliminary result confirms findings by other groups that suggest improved yields after addition of AAP during *cap* library production [146,151].

Comparison of variant frequencies between the two plasmid libraries demonstrated a good correlation (Pearson r of 0.77; Figure 34E). However, 11 variants (all with a CTG leucine codon at the first variant position) demonstrated poor abundance barely above the detection limit for both libraries, potentially indicating an error in DNA synthesis for these variants. Due to their low abundance, I excluded these variants from following analyses. Although a higher yield was achieved through the addition of AAP during virus production, a comparison of variant frequencies in CMV-cap6-536 produced with or without addition of AAP showed no clear differences and a strong correlation between the two settings (Pearson r of 0.98; Figure 34F). Comparison of variant frequencies between rep and CMV libraries after normalization to the plasmid input showed a lower correlation (Pearson r of 0.8), but still indicated that well-producing variants will give

higher titers in either construct (Figure 34G). Curiously, of the two variants with the lowest relative abundance in the virus library, one (cN001) was the candidate observed with the highest abundance in NPC cDNA of the initial selection round.

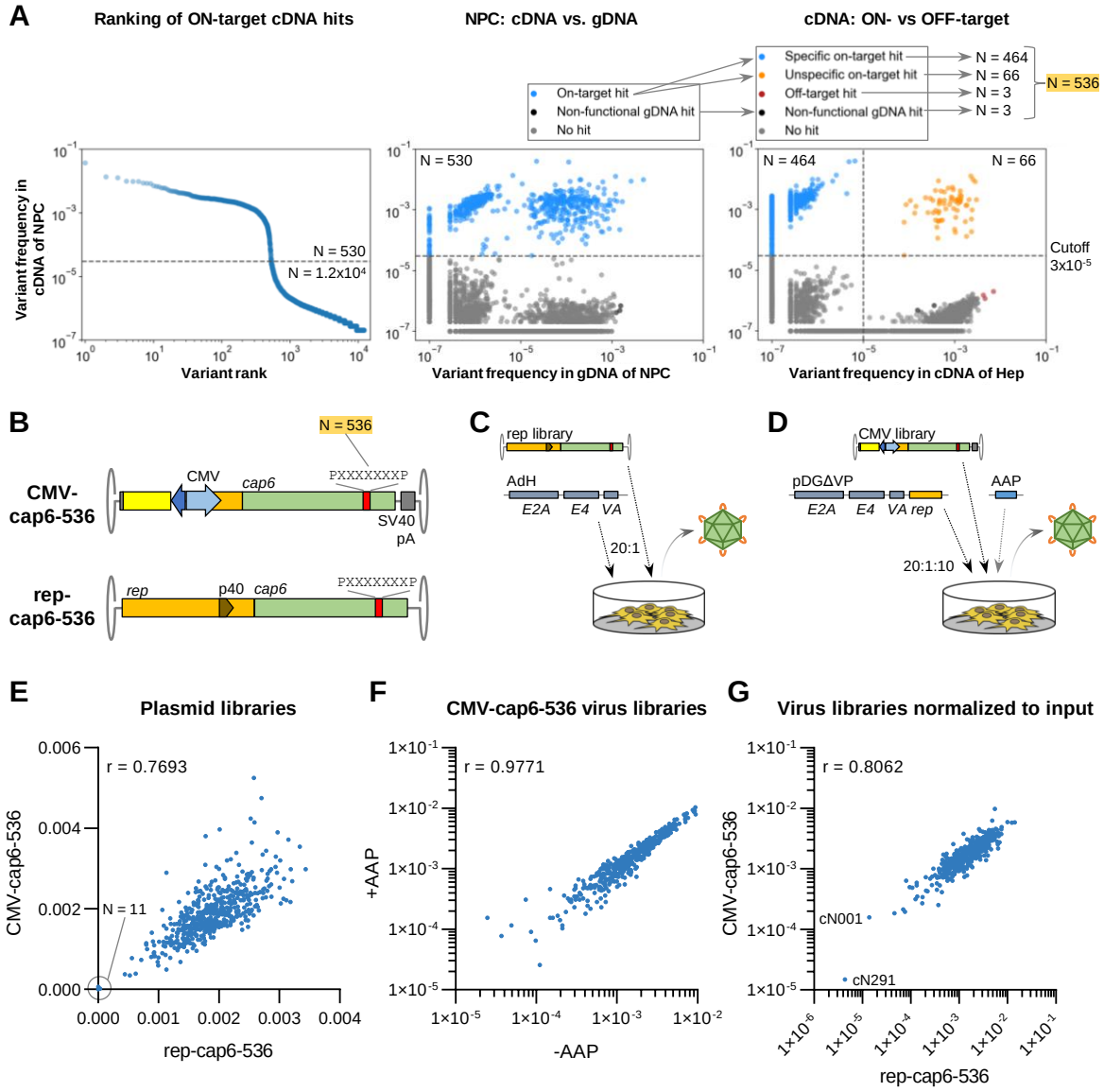



**Figure 34: Cloning and production of a candidate pool for secondary *in vivo* selection. (A)** Variant selection by (i) ranking candidates in on-target (NPC) cDNA, (ii) separating variants with high frequency in cDNA (negative control: three non-functional gDNA hits), and (iii) identifying variants with high frequency in on-target only (specific on-target hit, blue) or in on- and off-target (unspecific on-target hit, orange). The top three off-target hits were selected as well. **(B)** 536 variant sequences were synthesized through pooled oligonucleotide synthesis and cloned into the CMV-cap6 and rep-cap6 constructs. **(C-D)** Virus library production by plasmid transfection. The rep-cap6-536 library (**C**) was produced by co-transfection with the adenoviral helper plasmid, while the CMV-cap6-536 library (**D**) was produced by co-transfection with pDGΔVP containing adenoviral helper and AAV2 *rep* genes. Additional supplementation with a third plasmid for AAP expression was tested as well. **(E)** Comparison of variant frequency between plasmid libraries rep-cap6-536 and CMV-cap6-536. **(F)** Comparison of CMV-cap6-536 virus libraries produced with or without addition of AAP expression plasmid. **(G)** Comparison of rep-cap6-536 and CMV-cap-536 virus libraries as normalized to the respective plasmid input.

To conduct a secondary screening in separate liver cell-types instead of bulk NPCs, the CMV-cap6-536 library was administered systemically to a wild-type C57BL/6 mouse (Figure 35A). One week later, principal cell types were extracted from the mouse liver by employing a custom FACS pipeline (mouse work and cell-type isolation by Pervinder Choksi, Willenbring lab). The isolated cell types were hepatocytes (Hep), cholangiocytes (Chol), endothelial cells (EC; including but not limited to LSECs), hepatic stellate cells (HSC) and macrophages (Mac; including but not limited to Kupffer cells).

I extracted genomic DNA and RNA (converted to cDNA) of each liver cell population for analysis of bulk transduction and variant tracing by NGS. RTqPCR-based analysis of bulk relative *cap6* expression revealed the strongest signal for hepatocytes, with the signal for other cell types was an order of magnitude lower (Figure 35B). ddPCR-based analysis of vector DNA distribution, however, demonstrated that the highest level of vector genomes was found in macrophages, followed by endothelial cells and macrophages. Thus, normalized expression (ratio between signal in gDNA and cDNA) was highest for hepatocytes and lowest for macrophages, indicating a large difference in the functional transduction rate between the two cell types.

Tracing variant abundancies across gDNA revealed a good recovery of most variants across the different cell types with a gradual decrease between the most and least abundant variants (Figure 35C). For all cell types, the most abundant reads in gDNA were derived from gN001, a control variant selected in the first screening round for its high abundance in gDNA. For cDNA (RNA), on the other hand, most variants were not recovered across all tested cell types (Figure 35C). Out of 525 analyzed variants, only 233 were recovered from hepatocytes (44%), 145 from cholangiocytes (28%), 136 from endothelial cells (26%), 156 from hepatic stellate cells (30%) and 138 from macrophages (26%). This indicates a high barrier for functional transduction, but also reveals a strong enrichment of highly abundant candidates. For hepatic stellate cells, for instance, the top ten most abundant variants contribute 95% of all detected reads.

To evaluate replicability across both selection rounds, the relative read frequencies in hepatocytes were plotted against each other for gDNA and cDNA (Figure 35D-E). For gDNA, a weak positive correlation was observed (Pearson r of 0.36; $p<0.0001$), while cDNA showed no detectable correlation between the two rounds of selection (Pearson r of -0.01; $p=0.73$).

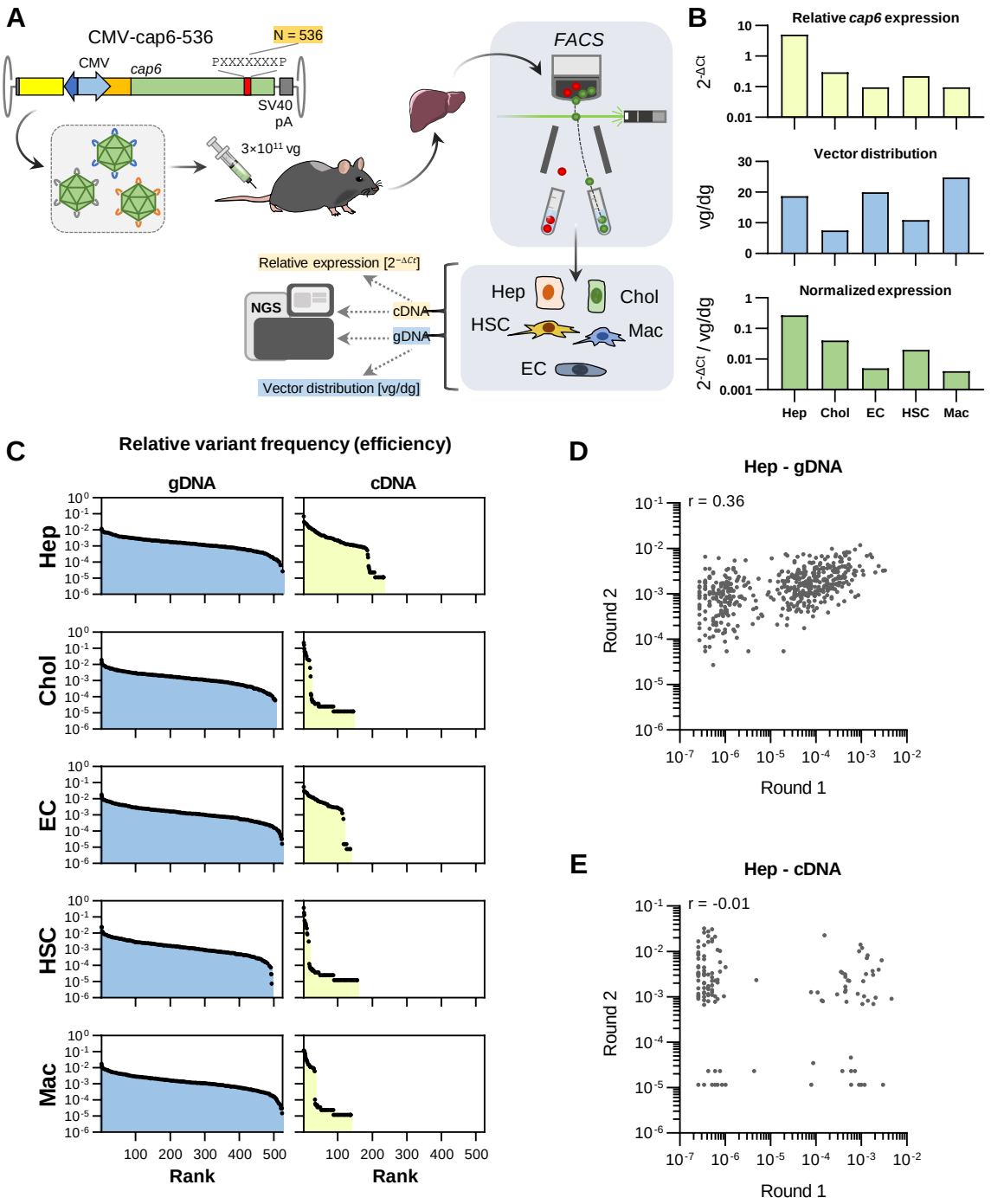



**Figure 35: Secondary *in vivo* selection in separate liver cell types. (A)** The CMV-cap6-536 library was systemically administered to a C57BL/6 mouse at $3\times10^{11}$ vg. One week later, the liver was collected for FACS-based isolation of hepatocytes (Hep), cholangiocytes (Chol), endothelial cells (EC), hepatic stellate cells (HSC) and macrophages (Mac). From each cell type, gDNA and RNA (converted to cDNA) were extracted for analysis of variant distribution by NGS and bulk transduction. Mouse work and cell-type isolation were conducted by Pervinder Choksi (Willenbring lab). **(B)** Bulk relative *cap6* expression as measured by RTqPCR [$2^{-\Delta Ct}$], and bulk vector distribution as measured by ddPCR [vg/dg]. Normalized expression was calculated as the ratio between relative expression and vector distribution. **(C)** Relative variant frequency determined by high-throughput sequencing, ranked according to abundance in gDNA and cDNA. **(D-E)** Comparison of relative variant frequency between first and second screening rounds for hepatocyte gDNA (**D**) and cDNA (**E**).

To select potential candidates for downstream evaluation, the ten variants with the highest read counts in cDNA of each cell type were inspected further (Figure 36A). To better compare variant performance, read frequencies were normalized in two steps. Firstly, variant frequencies were normalized to their relative abundance in the input library. Secondly, reads were scaled by their bulk transduction values (vector distribution for gDNA, and relative *cap6* expression for cDNA) to allow comparisons across cell types. Normalized frequencies for all variants are depicted in Figure 36B, showing broad abundancy for gDNA but low overall recovery for cDNA.

One approach to quantitatively identify cell-type-specific lead candidates is the calculation of combined capsid specificity scores (Figure 36C) [149]. These compare normalized read frequencies between on- and off-target cells, approaching values close to 1.0 for highly specific variants or 0.0 for unspecific variants. Thus, I extracted the normalized read frequencies in cDNA for the top ten candidates of each cell type (Figure 36D) and used these values to calculate the respective combined capsid specificity scores (Figure 36E). While all of the top ten hepatocyte candidates showed high specificity for expression in hepatocytes, other cell types were transduced less selectively. For endothelial cells, for instance, the two candidates with the most abundant reads (cN247 and cN342) showed high specificity scores. The third candidate (cN080), however, also demonstrated similarly high expression in hepatocytes and macrophages, thus resulting in a low specificity score. Especially for macrophages, the top ten candidates exhibited a poor specificity with only cN512 and cN501 showing acceptable selectivity for this cell type (scores of 0.83 and 0.99, respectively).

Overall, the *in vivo* selection of AAV6 peptide display variants demonstrated a higher barrier for enrichment in RNA (cDNA) as compared to gDNA. Although many candidates that gave functional transduction in the first selection round did not perform well in the second round, highly expressing candidates could still be detected for each tested liver cell population. The calculation of specificity scores helps to identify cell-type-specific capsid variants. Based on their functional and cell-type-specific transduction profile, lead candidates can now be selected for in-depth downstream evaluation and comparison to benchmark capsids as well as wild-type AAV variants.

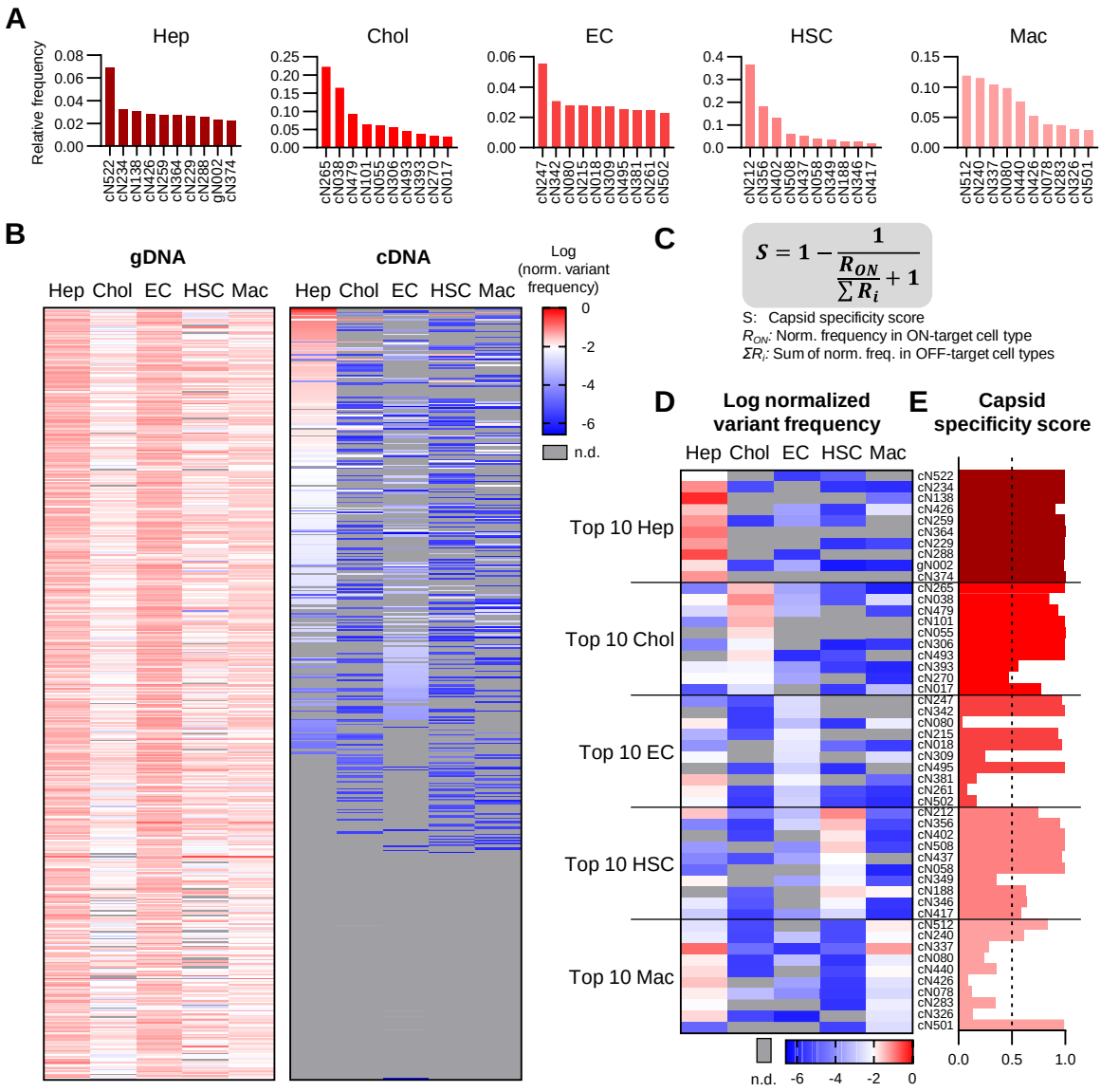



**Figure 36: Identification of cell-type-specific lead candidates. (A)** Relative read frequency in cDNA for the top ten most abundant capsid variants in hepatocytes (Hep), cholangiocytes (Chol), endothelial cells (EC), hepatic stellate cells (HSC) and macrophages (Mac). **(B)** Heatmap of logarithmic normalized variant frequency. Relative variant frequencies were normalized by (i) frequency in input library and (ii) bulk vector distribution (gDNA) or relative *cap6* expression (cDNA). Variants are ranked according to the sum of normalized read frequency in cDNA. **(C)** Formula for calculation of specificity scores derived from Körbelin *et al*. [149]. **(D-E)** Logarithmic normalized variant frequencies (**D**) and respective specificity score (**E**) in cDNA extracted for the top ten candidates in each cell type. n.d., not detected.

# 4 Discussion

With an ever-rising number of clinically approved gene therapy products, AAV vectors are dominant players in the field of vector-based gene delivery. AAVs facilitate safe and efficient transduction of a broad range of tissues with long-lasting transgene expression. This enables genetic correction of otherwise incurable diseases through AAV-based gene replacement, addition, silencing or editing [123]. Although they promise immense benefits, the current generation of AAV vectors is faced with persistent challenges and high expectations. These include faithfully packaging a foreign transgene, yielding high titers, overcoming host immune responses, delivering genetic cargo to the nuclei of target cells without getting sequestered in off-target cells, lysosomes, or proteasomes, and persisting as transcriptionally active episomes. A deficiency in any of these qualities would reduce the vector's therapeutic index and clinical applicability. Suboptimal transduction efficiency lowers therapeutic efficacy and drives the need to administer high vector doses. Although generally well tolerated, this "over"-dosing can induce toxicity even for normally safe AAV vectors. Reports of rare dose- or immune-related toxicities and resulting fatalities in clinical studies have rattled the AAV vector field [120,121,298]. Paired with pre-clinical observations of toxicity in dorsal root ganglia and evidence for a relation to hepatocellular carcinoma in humans [183,299], these concerns are fueling the need for a new generation of vectors with improved safety and transduction profiles. To enable therapeutic efficacy at lower vector doses and thus mitigate risks of adverse events, especially the transduction efficiency of AAV vectors must be increased. To avoid transduction in off-target tissues, this efficiency must be carefully balanced with vector specificity, in order to achieve homing only to intended target cells and spare off-targets. Improving efficiency and specificity are also crucial to optimize transgene expression, and, in the case of gene silencing, to achieve optimal target RNA knockdown.

## 4.1 Multiplexing with single-AAV-delivered knockdown effectors

Although most applications of AAVs in human gene therapy focus on replacing mutated or under-expressed genes [123], these vectors also enable the delivery of effector elements for gene silencing, activation or editing. Knockout and DNA editing strategies prevent expression by introducing mutations in target genes that result in frame shifts, premature stop codons and nonsense-mediated decay of target transcripts. Gene silencing, in contrast, is a mutation-free method for targeted expression regulation [123,300,301]. This

silencing can be achieved by RNA-guided degradation of target RNAs as mediated by RNAi and Cas13d. Both systems have been employed with AAV vectors before and allow multiplexed targeting within a single-vector system [210,302]. Importantly, multiplexed targeting of either several RNAs or for several sites within the same RNA target can offer crucial improvements in vector applicability and overall knockdown efficiency. Especially for antiviral applications, multiplexing can be essential for preventing mutational escape as demonstrated above and discussed below (sections 3.1.2.2 and 4.1.4, respectively). To find the most efficient multiplexing strategy for a single-AAV-vector setting, my goal was to compare and evaluate RNA-targeting using the shRNA/RNAi and Cas13d systems.

### 4.1.1 Targeted RNA degradation through Cas13d and RNAi

In the first part of this study, I evaluated *Rluc* reporter and *CD44* mRNA knockdown with CasRx (*Ruminococcus flavefaciens* Cas13d) and shRNA effectors. To establish optimal AAV-CasRx vectors, I created an all-in-one vector system using U6 promoter-controlled expression of the gRNA array and a secondary RNA polymerase II expression cassette for the CasRx cDNA. To fit all components into a single vector without exceeding the packaging limit, a trade-off between the size of the promoter and the polyadenylation signal was necessary. This involved the combination of either a strong CMV promoter with a weak but small synthetic polyadenylation signal (mini-pA), or of a small, weak EFS promoter with a larger but stronger bgh polyadenylation sequence (bgh-pA). These modulators have previously been compared by Choi *et al*., who demonstrated superior expression of AAV cassettes with bgh- over mini-pA [159], and Toktay *et al*., who clearly demonstrated a much higher expression with CMV as compared to the EFS promoter in Hek293T cells [303]. My results demonstrated a stronger knockdown efficiency for the combination of CMV and mini-pA, indicating a higher expression of CasRx for this setting. It must be noted, however, that CMV may not be the ideal promoter in all cell types, especially for applications *in vivo*. Its strong, ubiquitous expression may drive knockdown in a wide range of transduced cells and would thus require combination with either a highly specific AAV capsid or a secondary expression modulator to minimize unwanted knockdown in off-target cells.

Such a secondary modulator was introduced in section 3.1.1.2, using miR-122 binding sites in the 3'-UTR of CasRx mRNA. As expected, this strategy worked well when the aim was to inhibit knockdown through CasRx, by relying on miR-122-RISC-induced mRNA

cleavage and resulting post-transcriptional silencing in miR-122-expressing cells (*i.e*., hepatocytes or, in this case, Huh7 hepatoma cells). This data confirms and expands findings when the same approach was applied for other transgenes as well, including Cas9 [174,185,203]. miRNA-switches are especially useful in the AAV field as many AAV capsids transduce a broad range of target cells. Introducing binding sites for other miRNAs enables detargeting from defined cell types or tissues, *e.g*., by using mir-142-3p binding sites for deactivation in hematopoietic cells, or a combination of miR-122 and miR-1 binding sites for silencing of an AAV9-delivered transgene in liver and muscle (the main target tissues of AAV9 outside of the CNS) [304,305]. A deactivation in antigen-presenting cells by miR-142-3p has also been shown to increase tolerability to AAV administration by reducing transgene-directed immunity [306]. One concern of the miRNA-switch strategy is the sequestration of miRNAs that regulate endogenous gene expression, in turn causing de-regulation of cellular expression. Although this was not observed in the aforementioned case of miR-142-3p-based silencing [306], others have shown that this can indeed occur depending on the number of available miRNA target sites and the rate of endogenous miRNA expression [307]. Therefore, the application of such miRNA-switches *in vivo* should be paired with monitoring of endogenous gene expression in transduced cells, to evaluate the safety of such vectors.

Cas13d is capable of processing crRNA arrays into functional gRNAs by cleavage mediated *via* its HEPN domains [308]. I utilized this feature to create crRNA arrays that allow multiplexing of different target mRNAs, as was suggested by Konermann *et al*. [210]. Similar to my overhang-based cloning approach, Liao *et al*. also demonstrated efficient cloning of such crRNA arrays [309]. By combining a single targeting gRNA with one or more non-targeting gRNAs within two- to four-fold CasRx arrays, I could demonstrate that knockdown is possible and efficient for up to three-fold multiplexing. Using a four-fold array structure, however, I could not observe knockdown of the *Rluc* target. This is not due to the packaging limitation of AAV vectors since the poor performance of the four-fold array was observed for both plasmid transfection and AAV transduction. Intriguingly, Konermann *et al*. had demonstrated that a four-fold array structure can indeed induce efficient target knockdown [210]. Of note, in their setting each spacer targeted a different site in the same target RNA instead of including non-targeting spacers. This might indicate that in my setting, the four-fold array is functional *per se*, but oversaturated with non-targeting spacers, since 3/4 of the expressed gRNAs in the array were non-targeting and could block knockdown induced through the single targeting spacer in the array

structure. This may be investigated further by including two to four targeting spacers in combination with the non-targeting controls, to determine whether and when CasRx saturation occurs.

I found CasRx-based *Rluc* reporter knockdown to be functional across different experiments, both only when directly targeting *Rluc* coding sequences or a *CD44v6* sequence placed in the 3'-UTR of *Rluc*. In contrast, targeting of endogenous *CD44* mRNA did not produce any significant knockdown, neither for a single gRNA nor a triple-spacer array containing three *CD44*-targeting gRNAs. This discrepancy between reporter assays and endogenous mRNA knockdown was unexpected and warrants further investigation. As knockdown was functional with shRNAs, efficient AAV transduction of Hepa1-6 cells was most likely achieved. Moreover, as determined by the promoter evaluation in section 3.2.1.2, CMV promoter-based expression from ssAAVs in Hepa1-6 is efficient at three days post-transduction. Finally, as the shRNAs and the CasRx gRNAs use the same U6 promoter for expression, the availability of effector RNAs should be given in both settings. Notably, though, the shRNAs were expressed from scAAV vectors while CasRx and gRNAs were expressed from ssAAV vectors. Therefore, an earlier onset of transgene expression is expected for the shRNAs [98,99]. In the future, knockdown for shRNA- and CasRx-based effectors should be investigated over a larger time course to test whether the differential knockdown is based on kinetics. To determine whether this is a target-specific effect, knockdown for other mRNA targets should be assessed as well. In the literature, knockdown between the two systems (shRNAs and CasRx) is described as stronger for Cas13d when addressing exactly the same target sites [210]. Cas13d, and CasRx in particular, was used multiple times in AAV vectors highly similar to the setting here, both *in vitro* and *in vivo* [210,235,310,311]. Published results demonstrate a highly functional knockdown through Cas13d for multiple different endogenous target RNAs, including not only coding RNAs but also circular RNAs [312]. Therefore, CasRx-induced silencing of *CD44* should be possible in principle. A systematic dissection of CasRx knockdown with different endogenous target RNAs in Hepa1-6 cells may be conducted to better understand the observed lack in knockdown efficiency. This could include different assay time points as well as a detailed comparison to shRNAs or other silencing effectors. More CasRx gRNA sequences may be tested as well, as a lack of efficiency of the specific sequences tested here cannot be excluded.

Apart from the observed lack in knockdown efficiency for the cellular CD44 RNA, another emerging concern for Cas13d systems is specificity. While Konermann *et al*. initially

described a superior specificity of Cas13d as compared to shRNAs [210], other reports indicate the risk of off-targeting for Cas13d as well. In bacteria, Cas13d (and other Cas13 systems) induce bystander cleavage of non-target RNAs after binding of the gRNA-Cas13d ribonucleoprotein to a complementary target mRNA. This collateral bystander cleavage can induce bacterial dormancy to shut down ongoing infections with phages [313]. This effect was initially not observed in mammalian cells [210], but later reports indicate that this may be a problem after all [310,314-316]. Collateral cleavage can induce dysregulation of a broad range of off-target genes, and thus requires careful monitoring. In a mouse model with a CasRx knock-in, lethal neurotoxicity was observed due to collateral RNA cleavage [317]. This demonstrates the dangers of long-term expression of CasRx. To overcome this collateral activity, new orthologs of Cas13d were screened or generated with the aim to identify high-fidelity versions of the effector [318,319].

Finally, yet another issue with the application of Cas13d for human gene therapy is host immunity towards the Cas13d protein [320]. As Cas13d variants are expressed by multiple bacterial species of the human gut microbiome, an adaptive immunity towards CasRx is expected. This effect has also been observed for other CRISPR effectors including SpCas9 and SaCas9 [321], and must be monitored when considering CRISPR systems for application in human gene therapy.

In summary, although its use for the induction of targeted RNA degradation has been demonstrated by others, the Cas13d system is not yet ready for clinical applications until key efficiency and safety measures are addressed, including host-immune response to Cas13d and unspecific collateral RNA cleavage. As for other CRISPR and RNAi effectors, a close monitoring of both on-target efficiency and specificity is imperative for future applications.

### 4.1.2 Silencing through RNAi: shRNAs and beyond

My results demonstrated robust knockdown of reporter, cellular and viral RNAs through shRNA effectors. This confirmed the strong silencing efficiency described for AAV-delivered shRNAs [203,322]. One problem for AAV-shRNA vectors is unfaithful packaging due to the strong secondary structure of shRNA-encoding sequences [291,323]. These can induce packaging of truncated, sub-genomic vector species, as I could observe by agarose gel electrophoresis for packaged genomes containing either no, one or three shRNA expression cassettes. A more sophisticated technique for analyzing these sub-genomic fragments is AAV-genome population sequencing (AAV-GPseq), an approach

that employs single-molecule, real-time (SMRT) sequencing [93,323]. This method has been applied by Tai *et al*. for AAV vectors containing a single shRNA expression cassette and demonstrated high rates of mispackaging [323]. The same issue is observed for Cas9 sgRNA vectors, which contain a secondary structure in the sgRNA scaffold sequence [93]. Therefore, sub-genomic packaging may also be a problem for Cas13d gRNA arrays, although this has not been reported yet. Packaging of sub-genomic fragments renders part of the vector product non-functional. This would drive the necessity for higher doses to achieve functional knockdown and could thus increase the risk of toxic side effects. A possible solution has been described by Xie *et al*., who showed that placement of the shRNA sequence into a miR-33 scaffold can improve the integrity of packaged AAV genomes and increase on-target specificity [324].

My experiments with AAV-shRNA vectors primarily focused on testing knockdown efficiency. To complement these efforts, a transcriptome-wide analysis of both on- and off-target knockdown effects may be achieved by comparative RNA sequencing [325]. Apart from off-targeting, shRNA-induced toxicity is an additional concern for RNAi-mediated knockdown tools. In 2006, Grimm *et al*. employed AAV-shRNA vectors for inhibition of hepatitis B virus (among other targets), and detected severe and lethal hepatotoxicity for multiple shRNAs [207]. This toxic effect was shown to at least partially stem from competition between shRNAs and endogenous miRNAs including miR-122. By exploiting the same RNAi processing machinery including Exportin 5 and Dicer, highly expressed shRNAs prevent essential gene regulation through miRNAs and therefore exhibit broad side effects and resulting toxicity. Dose and sequence of shRNAs play major roles for this competition, with some shRNAs being less toxic and lower doses being tolerated better. Toxicity induced by AAV-shRNA vectors was found by others as well and is neither limited to the liver nor to mice [326-328]. Aside from lowering vector dose and using weaker RNA polymerase III promoters for shRNA expression to reduce competition, safer knockdown conditions may be achieved by modifying the encoded RNAi trigger [203,208,329]. One strategy to avoid competition with miRNAs focuses on circumventing the need for Dicer-based processing by modifying the shRNA sequence so that it is channeled into a Dicer-independent, AGO2-dependent processing pathway [208,330]. This modification also minimizes activity from the shRNA passenger strand, which is sometimes retained for shRNAs and can induce off-targeting as well [331]. An even further increase in specificity and safety has been achieved by placing agshRNAs into a miR-451 backbone for selective expression from a tissue-specific RNA polymerase II promoter instead of using

the highly and ubiquitously active U6 promoter (RNA polymerase III-dependent) [325,331]. Importantly, Suhy *et al*. demonstrated that employing weak RNA polymerase III promoters allows safe application of a triple-shRNA vector in non-human primates and in humans [332,333]. This confirms the applicability of triple-shRNA AAV vectors, and demonstrates that a sufficient safety level can be achieved with only minor changes in vector design.

### 4.1.3 Downregulation of *CD44* to counter metabolic steatohepatitis

I pursued shRNA- and CasRx-based downregulation of CD44, a surface glycoprotein involved in the progression of (non-alcoholic) metabolic dysfunction-associated fatty liver disease (MAFLD or NAFLD) towards the inflammatory (non-alcoholic) metabolic dysfunction-associated steatohepatitis (MASH or NASH) [253,260]. CD44 is known to be involved in other diseases as well, including adipose tissue inflammation, insulin resistance (type 2 diabetes), and progression in different types of cancer [334-336]. Different isoforms of CD44 are generated through alternative splicing, and CD44v6, for instance, is known as a marker in the progression of colorectal cancer [337]. Due to its involvement in different disease progressions, silencing of *CD44* may enable a therapeutic intervention for colorectal cancer, MASH, and other diseases [260,335,338]. Antibody-based targeting of CD44 protein has been applied for such modalities and may present an alternative option to AAV-induced silencing [262,338]. However, CD44 is ubiquitously expressed and interacts with several different ligands, driving multiple physiological functions including cell adhesion, activation of lymphocytes and migration [338]. Therefore, therapeutic downregulation of CD44 should ideally be limited only to disease-relevant cell types, which may indeed be possible with cell-type-specific AAV vectors. For MASH, the relevant cell-types have not yet been identified, but most likely involve sub-types of non-parenchymal liver cells [260]. Studies in mouse models of MASH progression, such as mice fed with a methionine-choline-deficient diet, may aid in understanding the exact involvement of CD44 in disease progression, as well as highlight therapeutic effects of AAV-induced CD44 downregulation [339].

Different approaches of AAV-based intervention in liver fibrosis, a process also relevant for advanced NASH, have been discussed by Bu *et al*. [340]. Importantly, cell-type-specific targeting of AAV vectors to hepatocytes, macrophages, hepatic stellate cells or endothelial cells is crucial for many of these intervention strategies. As for all AAV therapies, challenges include questions about translatability between animal models and

clinical application. Since AAV vectors exhibit different tropisms in different species, pre-clinical testing is not always indicative of vector performance in humans [156]. Another issue for AAV gene therapy for MASH and related diseases is potential hepatotoxicity [120]. For many gene therapy applications, patients with preceding liver injury are excluded to avoid hepatotoxic side effects [341]. In mice, preceding liver injury or feeding with a diet to induce MAFLD was also shown to increase the risk of hepatocellular carcinoma after AAV treatment [299]. For this study, however, the employed AAV vectors were designed for homology-based integration into the oncogenic *Rian* locus, while AAV integration of typical gene therapy vectors is rare and undirected. The risk for hepatocellular carcinoma is already elevated for MASH patients [342]. Therefore, AAV-based therapy for treatment of MASH may require careful dosing, meticulous monitoring of patients and vector products with high safety and specificity levels.

### 4.1.4 Targeting of SARS-CoV-2 genomic RNA – inhibition and escape

Next to silencing *CD44* expression, the second application of RNAi that I focused on was targeting SARS-CoV-2 genomic RNA to counteract or prevent virus infections. These experiments demonstrated that (i) SARS-CoV-2 is susceptible to inhibition through shRNA-triggered RNAi, (ii) escape mutations in shRNA-binding sites evolve quickly and render this inhibition ineffective, and (iii) multiplexed targeting with three shRNAs can block the emergence of escape mutants and maintain prolonged inhibition. Similar shRNA-based strategies have been pursued for other viruses as well, and were mostly intended for treatment of chronic infections with hepatitis B/C/E viruses or human immunodeficiency virus (HIV) [207,302,332,343]. In these cases, prolonged expression of shRNAs is necessary to counter lasting chronic infections. Viral escape by mutation within shRNA-binding sites was observed on multiple occasions, especially for quickly mutating RNA viruses such as hepatitis C virus (HCV) and HIV [344,345]. Combinatorial RNAi was soon suggested and applied to overcome this limitation by addressing multiple target sites simultaneously [312,332,343,346]. Nonetheless, even for dual shRNAs mutational escape can still render RNAi inefficient, as was demonstrated by Shah *et al*. when targeting HIV [347]. Therefore, most combinatorial approaches apply at least three RNAi triggers. Previously, a triple-shRNA AAV8-vector targeting HCV has been developed for inhibition of HCV infection. After lowering shRNA expression by using weaker RNA polymerase III promoters, the TT-034 vector showed efficient HCV inhibition *in vitro* and good tolerability in non-human primates [332]. In a first vector-delivered RNAi clinical trial in humans, TT-

034 elicited no serious adverse events, suggesting that safe AAV-delivered combinatorial RNAi is possible [333]. However, the TT-034 program was terminated due to the increasing availability of other treatment options for HCV.

Prior hints at the functionality of RNAi-based inhibition of SARS-CoV-2 were provided by the application of shRNAs and siRNAs to inhibit SARS-CoV(-1) [348,349]. My experiments were designed to confirm this susceptibility for SARS-CoV-2 as well, and to investigate the applicability of single- and multiplexed RNAi to induce or circumvent escape mutagenesis. The three strongest shRNAs I tested demonstrated highly efficient SARS-CoV-2 inhibition by targeting RNA-dependent RNA polymerase (*RdRP*) and nucleocapsid (*N*) genes. These target genes were suggested early on by Abbott *et al*. on due to their high conservation across virus isolates [251]. In contrast, more mutagenic sequences such as antigenic regions of the Spike (*S*) gene, which underlie selective pressure by host neutralizing antibodies, present less ideal targets for shRNAs [350].

Since the completion and publication of our work, multiple other RNAi approaches with siRNAs have shown successful inhibition of SARS-CoV-2 infection *in vitro* and *in vivo*, providing additional evidence for the effectiveness of RNAi to counteract the virus [250,351-355]. Prominent target sequences within the SARS-CoV-2 genomic RNA included conserved sites within the 5'-UTR and ORF1ab (including *RdRP* and other non-structural protein genes). While some studies assumed that targeting of highly conserved regions would prevent the evolution of escape mutant virus variants [250,351,354], my data suggests that even highly conserved sequences within *RdRP* can undergo single-point mutagenesis when exposed to selective pressure with RNAi triggers. Idris *et al*. and Yogev *et al*. have also applied cocktails of three or more effectors to prevent mutational escape and allow broad applicability to different strains, although these authors did not specifically screen for mutated target sequences [353,355]. While I used Sanger sequencing of amplicons containing the shRNA-binding sites to identify mutational escape variants over consecutive passages, high-throughput sequencing approaches (as applied for AAV peptide variant evolution in section 3.2.3.2) would be much more appropriate in the future, as these would allow for a detailed insight into changes in quasi-species variant composition prior to and after application of RNAi triggers [356].

Whereas delivery of shRNAs and siRNAs is easily achieved in cell culture by transfection, *in vivo* applications for pre-clinical and clinical testing usually require a more sophisticated delivery approach. Although not described in this thesis, my anti-SARS-CoV-2 TRISPR A vector was tested in collaboration with the Ralph Baric laboratory in a mouse infection

model (details can be found in Becker *et al*. [290]). The mouse model used wild-type mice in combination with a mouse-adapted SARS-CoV-2 variant (MA10) and demonstrated functional intranasal delivery of an AAV9-packaged TRISPR A as well as reduction in SARS-CoV-2 viral load in the lung [357]. Successful intranasal or intratracheal delivery of AAV vectors to pulmonary cells has also been demonstrated by others [358-360]. For administration to NHPs and humans, application of aerosolized AAVs through nebulizers offers potential for sufficient pulmonary delivery [361,362]. Directed capsid evolution may aid in generating AAV vectors with improved intrapulmonary transduction profiles, allowing a broad transduction of cells susceptible to infection by SARS-CoV-2 to achieve an optimal prophylactic treatment regimen. Interestingly, for treatment of cystic fibrosis lung disease, an engineered AAV capsid has been generated for nebulizer-based administration and is currently being tested in a phase 1/2 clinical trial (ClinicalTrials.gov Identifier: NCT05248230) [363]. For siRNA approaches, similar *in vivo* administration can be achieved using aerosol inhalation or intranasal installation of naked siRNAs, siRNA dendrimers or lipid nanoparticles [352-354]. This may be conducted before or after exposure to SARS-CoV-2, in order to investigate either prophylactic (as in my experiments) or therapeutic application. SARS-CoV-2 infection is primarily acute. Therefore, transient RNAi *via* administration of siRNAs can be sufficient to counter SARS-CoV-2 infection and usually does not require long-term expression of shRNAs as would be necessary in chronic infections with HIV or HCV. Still, rare cases of persistent SARS-CoV-2 infection have been described, and such patients may benefit from prolonged RNAi mediated through AAV-shRNA vectors [364,365].

Apart from the application of RNAi, other AAV-based approaches have been pursued to study, prevent or counter infections with SARS-CoV-2. To study SARS-CoV-2 in mouse models, AAV-based overexpression of its entry receptor human ACE2 has been conducted and successfully rendered the animals permissive to infection [359,360]. To prevent or counter infections, AAV-based prophylactic strategies have either focused on the development of AAV-vectored vaccine platforms or on the overexpression of receptor decoys. Vaccination strategies with AAV vectors expressing full or partial SARS-CoV-2 spike protein has shown multiple potential benefits, including long-lasting immunity after single-dose injection and vaccine stability [366-368]. Instead of building host immunity to counter SARS-CoV-2 infections, other AAV-based prophylactic approaches used vector-based overexpression of SARS-CoV-2 receptor decoys [369,370]. These modified versions of human ACE2 were expressed intranasally or intramuscularly and allowed prevention

or inhibition of SARS-CoV-2 infection through competitive binding of SARS-CoV-2 to ACE2-decoys instead of the cellular ACE2 entry receptors. Therefore, AAV vectors offer multiple pathways for interfering with SARS-CoV-2 infection that can complement my RNAi-based approach for future antiviral therapies. Further aspects of AAV-mediated, RNAi-based SARS-CoV-2 repression are discussed in the published manuscript comprising my experimental results of this project [290].

## 4.2 Screening of promoters for enhanced and selective transgene expression

To achieve therapeutic levels of transgene expression with AAV vectors, both capsid-defined transduction efficiencies and promoter-defined expression rates often require optimization. After successful transduction and second-strand DNA synthesis, the rate of expression is determined by tissue-specific interactions of transcription factors with sequences in the enhancer and promoter elements of an expression cassette [371]. A wide range of natural and synthetic promoter variants have been applied in AAV vectors, with the goal to drive either ubiquitous or tissue-specific transgene expression [174]. To compare a panel of 53 commonly employed promoter variants (“promoter library”) in parallel, former lab member Claire Domenger utilized a barcoding technique that is commonly used to track capsid transduction efficiencies [142]. In this strategy, a variant-specific barcode sequence is placed into the 3’-UTR of a reporter *eYFP* gene, thus allowing barcode readout by high-throughput sequencing on the DNA and RNA (cDNA) level. Barcoding has been widely and successfully applied to track functional transduction with different AAV capsid variants [142,143,372]. The approach is methodically related to Massively Parallel Reporter Assays (MPRAs), which were developed for parallel analysis of genetic modulators (*e.g.*, *cis*-regulatory elements/enhancers) [373,374]. In MPRAs, putative enhancer variants are paired with a minimal promoter sequence, a reporter gene, and a variant-specific tag (*i.e.*, a DNA barcode sequence) in the 3’-UTR. Active enhancer elements that drive efficient expression in combination with the minimal promoter can then be identified by sequencing of barcode pools in DNA and RNA. MPRA has been applied in the AAV context to facilitate *in vivo* screening of tissue-specific enhancer elements [375]. Advancing this technology further, Hrvatin *et al*. combined MPRA with single-cell RNA-sequencing, yielding the Paralleled Enhancer Single Cell Assay (PESCA) approach that allows screening of cell-type-specific enhancer variants [376]. Instead of testing putative enhancer variants in combination with a minimal promoter sequence,

Carrell *et al*. utilized barcoding to compare a set of putative promoters in an AAV context [377]. This enabled the identification of the VWA3A promoter for specific transgene expression in ventricular epithelial cells. Similarly, Westhaus *et al*. used a barcoded reporter setting to compare up to four different promoters in scAAVs *in vitro* and *in vivo*, showing a high activity of the CAG promoter and low but detectable ubiquitous activity of the AAV p40 promoter outside of its endogenous context [74]. Notably, the specific approach pursued and validated in this doctoral work is distinct from these prior reports as it is the first where barcoding was applied for systematic *in vivo* comparisons of a large panel of known promoter variants across different tissues.

Barcode sequencing yields variant ratios for each tissue. These variant ratios detected in RNA (cDNA) are then commonly normalized to the variant ratios found in gDNA, giving the rate with which each enhancer/promoter variant drives expression [378]. A cross-tissue comparison of this rate is difficult, as different tissues yield different bulk expression rates. Therefore, a secondary normalization step is required. My analyses of promoter expression profiles for four individual variants (CMV, LP1, SPc5-12 and GFAP) offered a dataset for direct comparison to the promoter library data. This comparison demonstrated that (i) good normalization is achieved using relative reporter expression values to compare promoters across tissues, and (ii), barcode-based parallel analysis correlates well with results obtained by single-promoter evaluations. Both library and individual promoter screens were conducted with AAV9, a variant with a broad transduction profile and prominent transgene delivery to liver, muscle and brain [294]. Due to its unselective tropism, AAV9 is uniquely qualified to assess promoter activity across multiple tissues. Intriguingly, it was reported that the AAV capsid identity can epigenetically influence the transgene expression profile [379,380]. Therefore, cross-tissue expression profiles of different promoters should be assessed with capsids other than AAV9 as well.

It is known that promoters and rAAV genomes in general can undergo epigenetic silencing [381,382]. For the CMV promoter, for instance, differential silencing rates across different tissues have been reported [176,180-182]. This silencing may also explain the superior expression profile I observed with the CMV promoter in hepatocytes *ex vivo* in comparison to its lower expression compared to GFAP and LP1 *in vivo* (section 3.2.1.2). To identify and investigate such silencing events, AAV-mediated promoter activity should ideally be measured over a longer time period. To complement the currently established promoter expression profiles, future iterations of barcode-based promoter screens may include different time points of analysis in order to comprehensively study transgene

expression dynamics. This could help to understand and compare expression kinetics mediated by different promoters.

The four promoters I investigated *in vivo* confirmed and validated findings of the barcoded promoter screen. As described in literature, CMV enabled strong expression across multiple tissues and showed a bias towards expression in pancreas [383]. Expression with LP1 was highly specific for liver tissue, for which it was originally developed and described [187-189]. SPc5-12 was active mostly in heart, skeletal muscle, and liver, and thereby also demonstrated its reported expression profile [384]. The one promoter that showed an unexpected activity was GFAP. In the literature, the full-length GFAP promoter (also known as gfa2) is renowned for its astrocyte-specific expression profile [164,292,385]. As such, it has been harnessed to focus transgene expression to astrocytes instead of neurons in AAV-based transduction of the CNS [386-388]. Several AAV capsids have been engineered for efficient crossing of the blood-brain barrier and transduction of different cell types within the CNS after systemic administration (primarily in mice) [135,145]. Therefore, an astrocyte-specific promoter is highly useful for applications requiring focused transgene expression within the CNS. Strikingly, as opposed to the astrocyte-specific expression reported in literature, my experiments demonstrated highly efficient GFAP promoter-driven transgene expression in the liver. As CNS-targeted AAV vectors still show peripheral off-targeting to the liver and other tissues after systemic administration, this finding demonstrates a major concern for off-targeting of CNS-directed AAV gene therapies employing the GFAP promoter. At the same time, its superior hepatic activity characterizes the GFAP promoter as a powerful new tool for liver-directed gene therapy.

### 4.2.1 Driving hepatocyte-specific transgene expression with the GFAP promoter

With a liver-directed expression profile in the AAV context, the GFAP promoter may aid hepatic gene therapy by enabling highly efficient transgene expression. Liver-specific expression mediated by the LP1 promoter has proved useful in a clinical setting with etranacogene dezaparvovec (Hemgenix), an FDA-approved gene therapy using AAV5 for overexpression of Factor IX in hemophilia B [389]. My results demonstrated that the GFAP promoter induces an even stronger expression than LP1 in liver with a similarly advanced specificity profile. Therefore, it could allow efficient gene therapy applications with even lower vector doses. Apart from overexpressing Factor IX to treat hemophilia B, liver-directed gene therapy offers potential treatment for a number of genetic or acquired disorders [390,391]. Some of these are hepatic diseases such as Wilson disease and liver

fibrosis, while others such as hemophilia A and B require secretory transgene expression from the liver for correcting systemic disorders [112,113,340,392]. In many of these therapeutic interventions, transduction of hepatocytes is preferred or required to reach optimal outcomes. This includes hemophilia B, as the missing blood clotting factor IX is naturally produced in hepatocytes and displays lower biological activity when artificially expressed *e.g*. in the musculature [393-395].

Although mostly expressed in astrocytes, GFAP protein has been identified as a marker in hepatic stellate cells (HSCs) [396,397]. To investigate whether AAV-mediated GFAP promoter-driven expression within the liver is limited to these or other liver cell type(s), I collaborated with the Willenbring lab who facilitated isolation of different liver cell populations using a custom FACS pipeline. My subsequent analyses of transgene expression in these isolated cell types demonstrated superior activity of the GFAP promoter in hepatocytes, but only minimal expression in HSCs, cholangiocytes, endothelial cells and macrophages. Its strong expression in mouse and human hepatocytes and hepatoma cells was later confirmed both *ex vivo* and *in vivo*, surpassing even expression mediated through the liver-specific LP1 promoter. This hepatocyte-specific expression driven by the GFAP promoter has not been reported so far. Of note, its general ability to express within the liver is confirmed by observations of Touahri *et al*., who found off-targeting of GFAP expression intended for Müller glia cells when using AAV8-based systemic delivery [398].

To better understand the hepatocyte-directed expression of the GFAP promoter, I mapped putative binding sites for human and mouse transcription factors to its sequence and identified several hits. These binding sites can differ significantly between the two species, therefore expression has to be tested in both [399]. Of note, the expression profiles of GFAP and other promoters differed only marginally when tested in parallel in human and murine hepatocytes within the FRGN mouse model. Deletion of major regions within GFAP and comparison to gfaABC1D showed decreased expression as compared to the full-length promoter. As no particular deletion variant demonstrated consistent superior or inferior expression compared to others, several of the mapped (or other) transcription factor binding sites likely contribute to hepatocyte-directed expression. Detailed insights into the relevance of different transcription factors may be obtained by selectively mutating individual transcription factor binding sites instead of implementing larger deletions, and it may then be additionally confirmed by overexpression or silencing of specific transcription factors. By determining and retaining the exact binding sites of

relevant transcription factors, functional truncations may be introduced in future studies of the GFAP promoter. Such modifications may also be aided by replacing partial regions of the promoter with short synthetic cassettes known to drive liver-specific expression [400,401]. This reduction in size may then advance its applicability even further.

Overall, as opposed to its endogenous expression profile, the GFAP promoter demonstrated superior expression in hepatocytes when used in the AAV vector context. This confirms and adds to similar observations of promoter or enhancer elements differing in transcriptional activity between endogenous and vector settings as described by others as well [402]. Therefore, future AAV-delivered promoter studies may revisit endogenous promoters with selective expression profiles to study their behavior in an exogenous vector setting.

## 4.3 Directed evolution of novel AAV capsids

Directed evolution is a powerful strategy that aims at creating novel AAV capsids with desired tropisms or other desired vector properties. As opposed to rational design, directed evolution requires no mechanistic understanding of the modifications implemented in the capsid, but rather uses randomized diversification paired with iterative selection. This strategy has proved highly successful over the last two decades [83,125], but is highly sensitive to the employed selection parameters. Focusing only on on-target transduction can lead to the enrichment of unspecific vector candidates, while focusing only on enrichment in DNA can yield variants with a suboptimal functional transduction and inadequate transgene expression. To overcome these two pitfalls, I implemented enhanced selection strategies to advance directed AAV capsid evolution.

### 4.3.1 DEPOOL – negative selection

The DEPOOL strategy (depletion of off-targeting AAVs from on-target libraries) focuses on improving capsid evolution by removing off-targeting variants. To validate the preceding iteration of this approach [275], I conducted a barcoded capsid library screen comprising 39 peptide display candidates, which had been selected for either lung or pancreas transduction, and eight benchmark capsids. I could demonstrate successful and robust assay readout by studying the transduction profiles of these benchmark capsids. Confirming to reports in literature, my screen showed pulmonary targeting for benchmark capsids AAV4 and AAV2_L1 [149,403], highly efficient transduction of liver for AAV8 and AAVrh10 [142,404], and a broad transduction profile with superior expression in the pancreas

for AAV9 [142]. Albeit the AAV8 peptide display candidates selected from the preceding *in vivo* screens did not outperform these benchmarks in terms of efficiency, it was interesting and encouraging to note that most of them demonstrated enhanced pulmonary targeting as compared to the parental AAV8. Targeting of pancreas proved more difficult, although two candidates of the depleted pancreas screen could surpass the parental AAV8 for pancreatic transgene expression, outperformed only by AAV9. Although demonstrating this enhanced transduction in pancreas, these candidates did not evolve a selective transduction of the on-target as compared to the liver. These suboptimal candidate expression profiles are likely rooted in issues during primary selection conditions, which may include (a) a suboptimal diversification approach and (b) insufficient primary selection prior to candidate picking. Initial diversification was achieved by inserting a partially randomized NXXRXXX motif (X=any amino acid) into variable region VIII of AAV8. This motif has previously been exploited for multiple peptide display screens and may drive efficient but unselective target tissue transduction by, *e.g*., improving capsid trafficking instead of selective attachment [405]. As shown later for the RNA-selection screen, a DNA-based selection as conducted for the initial DEPOOL screen can suffer from unselective enrichment and lack of functional transduction of enriched candidates. Thereby, the results of this screen reaffirm the necessity to establish functional selection as discussed below (section 4.3.2).

The barcode screen did not demonstrate improved liver-detargeting for variants enriched in the depleted (DEPOOL) setting as compared to the conventional enrichment setting. Therefore, I implemented three improvements for the Cas9-based depletion conditions. First, (i) I targeted a different peptide structure (randomized 7mer flanked by prolines), thereby achieving complete coverage of the randomized sequence by the sgRNA. This allowed for highly specific variant depletion as validated with the cap6-536 library. (ii) Instead of using Cas9 for variant removal after PCR amplification (as done for the preceding screen), I used depletion as an intermediate step of amplification. This protocol is derived from Hardigan *et al*., who employed a similar setting to remove overabundant variants from RNA-sequencing libraries [280]. The approach of employing Cas9 for *in vitro* depletion of unwanted sequences has been applied by others as well [406,407], mostly in the context high-throughput sequencing libraries. My DEPOOL strategy demonstrates, for the first time, that this can also be useful to remove unwanted *cap* variants from AAV capsid libraries. (iii) To expand the applications of DEPOOL beyond AAV peptide display libraries, I implemented the depletion protocol for a barcode region in the 3'-UTR of *cap*.

Barcoded *cap* constructs have been employed for capsid screening in multiple different settings [19,140,141]. They are highly useful to monitor *cap* libraries that include mutations along large stretches of *cap*, as opposed to peptide display. Barcodes can be associated either with a *cap* variant specified during cloning [19], which allows the creation of lookup tables for variant analysis. Alternatively, another approach from Pekrun *et al*. uses saturated barcoding where shuffled *cap* variants and barcodes are randomly paired [141]. Here, barcode enrichment is tracked by high-throughput sequencing, and enriched barcodes are then used as primer-binding sites to identify the associated *cap* variant. As demonstrated by my barcode-DEPOOL application, Cas9 can be used for *in vitro* removal of unwanted *cap* variants within such libraries based on their barcode identity. Variants may be selected for depletion either due to their high efficiency in off-target tissues, or due to their overabundance by which they may overshadow other promising but rare candidates. Apart from variant-specific barcoding, where each variant is linked to a specific barcode, grouped barcoding is also possible. Here, a group of variants shares a common barcode, and a library is created as a mix of these groups. This is useful for monitoring enrichment of libraries in parallel that utilize distinct diversification strategies, *i.e*., combining a peptide display library marked with barcode A and a shuffled library marked with barcode B. For such approaches, barcode-based deselection as offered by DEPOOL can provide advantages as well, such as removing one of the groups due to sub-optimal performance or due to dominant enrichment that limits the detectability of other groups.

Other negative selection strategies with the same premise of removing unwanted variants have employed de-selection based on capsid surface interactions instead of sequence identity. For instance, this can be achieved by incubating the AAV capsid library with off-target cells *in vitro* in order to remove variants with high affinity for attachment to these [408], by using affinity chromatography to remove variants that bind to specific attachment receptors such as HSPG [409], or by incubation with intravenous immunoglobulins (IVIGs) to remove variants with strong antibody-binding epitopes [133,153]. A sequence identity-based removal as offered by DEPOOL can be preferable for removing defined variants with undesired properties as identified by high-throughput sequencing. For instance, a highly enriched but unspecific variant A transducing both on- and off-target tissues overshadows the performance of less enriched but more specific candidates. If screening is continued with a library that contains mostly variant A and only minor fractions of other diverse variants, then detection of these potential candidates by high-throughput

sequencing can be hampered through overabundant reads of variant A. Hence, as discussed also by Hardigan *et al*. for RNA sequencing libraries, a targeted removal of such unwanted variants can prove highly beneficial [280].

### 4.3.2 RNA-based capsid selection

DNA-based selection of AAV capsid libraries has long been the standard in the field yet it does not guarantee enrichment of variants with enhanced functional transduction properties [154]. Alternative approaches are necessary to facilitate a selective enrichment of variants that have successfully undergone intracellular trafficking and uncoating in the nucleus. Therefore, my goal was to design a selection strategy for AAV capsid libraries that allows readout of functionally transducing variants. To be able to monitor specificity and efficiency of variants during selection, I wanted to implement variant tracing across both on- and off-target cells. For combination with FACS-based cell-type-isolation, I additionally aimed to express a fluorescent reporter that would enable isolation of efficiently transduced cells. As tested in comparison to other constructs and validated by *in vivo* screening, this design was realized with the MYECc2S construct. This construct employs CMV promoter-driven ubiquitous expression of the *cap* gene, and therefore allows tracing of *cap* variants on both DNA and RNA level across a broad spectrum of tissues. The *cap* mRNA is additionally stabilized by employing a SV40-late polyadenylation signal instead of the minimal AAV polyadenylation site [161]. Within the same vector, expression of an *eYFP* reporter is facilitated by the small ubiquitous EFS promoter and a minimal polyadenylation signal.

I selected the MYECc2S construct from a panel of construct designs for functional selection platforms that were created with the intention to optimize functional RNA-based selection. To identify the optimal design, I performed a systematic comparison of vector packaging and *cap* expression for these constructs, using the AAV2 wild-type as a control (rep-c2). Although decreased packaging was observed for all *rep*-depleted constructs, each gave a sufficient yield to enable *in vivo* screening. Such decreases in vector yield have been reported by others as well, and are to be expected when disconnecting *rep* and *cap* expression during vector production [146]. My analyses of vector capsids and packaged genomes by dot blot, silver stain and agarose gel electrophoresis showed similar profiles as the wild-type reference rep-c2 (wild-type), demonstrating the functionality of each vector design.

After confirming vector packaging, the next goal was to achieve boosted cap expression compared to rep-c2 in the absence of a helper virus. Replacing *rep* with a reverse RSV promoter-driven *eYFP* reporter and retaining p40 (BYRPc2; similar to the FT platform but using RSV instead of SFFV) resulted in a weak increase in *cap* expression after plasmid transfection but a decreased *cap* expression after vector transduction. Increased *cap* expression yields compared to wild-type as seen after plasmid transfection match observations by Westhaus *et al.* [74]. I found this boost in *cap* expression to be even stronger for constructs that harnessed CMV or CAG promoters for *cap* expression instead of using the endogenous AAV p40 promoter. The reasons for the surprising lack in a boost in *cap* expression after vector transduction remain unknown at this point, but this finding implies an interesting aspect of AAV biology that warrants further investigation. So far, based on my data, a *trans*-effect of *rep* can be excluded, as co-infection with a *rep*-expressing vector did not boost *cap* expression for BYRPc2. A combined *cis*- and *trans*-effect of Rep proteins binding to *rep* sequences in rep-cap2 and inducing *cap* expression is unlikely, as p5 and p19 promoters are known to be repressed in the absence of a helper virus co-infection [64]. An interference of the reporter cassette of BYRPc2 can also be excluded, since replacing either the RSV promoter or the entire reporter cassette in BYRPc2 did not increase *cap* levels either. Another possible explanation for the observed reduced *cap* expression after transduction for constructs devoid of *rep* compared to the AAV2 wild-type (rep-c2) may be a differential functionality of the assembled capsids. This would be in line with a report by Zeltner *et al*. who made similar observations of increased levels of empty capsids for recombinant AAVs lacking *rep* as compared to wild-type AAV [410]. An electron microscopy-based comparison of rep-c2 to BYRPc2 (or other constructs) may aid in identifying differential ratios of empty to full capsids.

Functional selection strategies have been proposed by others as well, but were faced with many challenges. One way to rescue functionally transducing variants with a rep-cap genome structure is to superinfect transduced cells with helper adenovirus. This will trigger AAV replication and packaging, but only allows selection in cells susceptible to infection by adenovirus [155]. Moreover, the inherent pathogenicity of adenovirus prevents the application of this strategy in larger animals. Selection of functionally transducing variants by Cre-based recombination can enrich variants that enter the nucleus, but is limited by the expression of Cre recombinase and therefore offers only limited insight into off-targeting [145,411]. Importantly, three publications reported within the last three years

demonstrated the usefulness of RNA-based selection for enrichment of functional variants of highly diverse AAV peptide display libraries [74,146,147]. Two of these used tissue-specific promoters to boost *cap* expression in on-target tissues [146,147]. Although this allows positive selection, off-target monitoring is again limited for these approaches due to their tissue-specific expression. The third strategy, termed functional transduction (FT) platform by Westhaus *et al*., utilized a ubiquitous *cap* expression driven by the AAV p40 promoter supported by an inverse SFFV promoter-driven reporter cassette [74]. Although showing RNA-based selection in humanized mouse livers and ubiquitous expression of the promoter design, Westhaus *et al*. did not include off-target monitoring in the selection process. Moreover, although their SFFVr-p40-combination enabled RNA-based selection, my *in vitro* comparisons of construct designs indicate that CMV may offer an even stronger *cap* expression. This strategy has recently been employed by Zinn *et al*., who used the CMV promoter to drive expression of barcoded *cap* variants, in order to study a set of capsids derived from ancestral reconstruction [151]. Instead of mere variant monitoring, my goal was to use CMV promoter-based ubiquitous expression for directed evolution of functional and cell-type-specific *cap* variants, an approach that has not been reported so far.

### 4.3.3 AAV-based targeting of distinct liver cell-types *in vivo*

After I could demonstrate functional *cap* expression and the potential for direct mRNA rescue for my RNA selection constructs *in vitro*, I selected a promising construct for further validation *in vivo* (MYECc2S). Next to CMV promoter-driven *cap* expression, this construct includes an *eYFP* cassette to enable FACS-based enrichment of transduced cells. Nonetheless, the reporter expression observed *in vivo* did not suffice to allow FACS-isolation of *eYFP*-positive cells. This may be explained by a low expression efficiency mediated by the EFS promoter, which I employed for its small size and ubiquitous expression profile. A low expression efficiency of EFS was also observed in the promoter library screen by Claire Domenger (data not shown) and may warrant adaptation of the current construct for future *in vivo* screens. Possible alternatives for EFS with small promoter size and ubiquitous expression may be the SFFV promoter as demonstrated by Westhaus *et al*. [74], or the recently reported micro-promoters µP-84 and µP-135 [412].

My results showed that the CMV promoter-based *cap* expression allows for RNA-based *in vivo* screening on the DNA and RNA level across different liver cell types. A comparison of the enrichment for these two modalities demonstrated a good correlation between

vector yield and variant abundance in DNA but not RNA. For RNA, a clear distinction was visible between highly expressing and non-expressing variants. This can be explained by differential rates of intracellular capsid trafficking, which has been observed for changes as little as single amino acid variations in the capsid [413]. DNA-based screening does not detect these biological differences, as it picks up all vectors regardless of whether they are present in the nucleus (functional transduction), arrested during intracellular trafficking or merely attached to the cell surface. Therefore, a pure DNA-based screening, while technically simple and thus appealing, offers limited insights into the functionality of enriched capsids. Overall, I could observe a higher barrier for capsid variants to achieve functional expression as compared to genome delivery. The expected and desired increase in stringency for RNA-based selection is useful for optimized selection of functional capsid variants. Furthermore, while many variants could deliver DNA to both hepatocytes and NPCs, I only observed a minor fraction of overlap between the two bulk populations in the primary selection round or across multiple sub-populations in the secondary selection round. This again indicates a high barrier for functional transduction across cell types. This is in line with comparative analyses of variant enrichment between DNA and RNA conducted by Tabebordbar *et al*. and Westhaus *et al*. [74,147].

Instead of generating the secondary selection library by rescue-PCR from on-target cells, I generated a synthetic library from an oligonucleotide pool in order to test the functional variants that were most enriched in NPCs. As opposed to PCR-based subcloning, this offered several advantages including reduced rates of mutagenesis, the ability to hand-pick variants with interesting transduction profiles, and near-uniform contributions of variants in the cloned plasmid library. While the size of this library was limited for the current study design to test its applicability, future iterations may increase library size by including additional controls such as variants encoded by different amino-acid codons for each peptide or by including parental viruses (AAV6 here) and other benchmark variants for direct comparison [19,411].

I cloned the oligonucleotide pool-based secondary library into both rep-cap and CMV-cap constructs and observed similar variant contributions after packaging for the two systems. This indicates that disconnecting *rep* and *cap* expression during library production does not alter the relative rate of packaging for individual variants and suggests translatability of capsid behavior between rep-cap and CMV-cap constructs. Adding additional AAP in *trans* during vector production to improve packaging of the *rep*-missing CMV library yielded a higher titer and did not change the relative contribution of each variant. This

indicates that this process can improve library yields without inducing a packaging bias. Zinn *et al*. have shown that some capsid variants may require *trans*-complementation with additional AAP and could trace this dependence to a single point mutation [151]. With an increase of titer observed after *trans*-supplementation with AAP, future vector productions of CMV libraries may benefit from improved packaging by addition of AAP.

*In vivo* screening of the secondary AAV6 peptide library was facilitated by FACS-based isolation of different liver cell populations after transduction. This strategy is highly useful to gain insights into bulk transduction in each isolated cell type, but is limited to the availability of useful cell type markers needed for isolation by FACS. Since screening with CMV promoter-driven *cap* expression proved functional, future screens with this and similar libraries may benefit from conducting additional single-cell RNA sequencing to study detailed expression profiles across more cell types and sub-types [414,415]. Monitoring of the second-round library in DNA and RNA showed low recovery of most variants on the RNA level but good recovery on the DNA level across all tested cell types. This is again indicative of the high selection stringency achieved with the RNA-based screening. Recovery of variants with low abundance in RNA samples may be improved by increasing the number of reads per sample, for instance, by switching from MiSeq- to NextSeq-based high-throughput sequencing. Injecting a higher virus titer may also enable better tracing, which would improve the accuracy of off-target assessment and candidate identification. Most importantly, the CMV library should be tested in additional animals to account for stochastic variance in transduction. As seen by the low overlap in functionality between variants found in the first and second library screens, variability between animals may be a major factor in variant enrichment.

RNA-based *in vivo* screening over two selection rounds allowed me to identify AAV6 peptide display capsid candidates with the potential for cell-type-specific transduction in mouse liver. Ultimately, candidates identified that exhibit functional transduction in the second-round library should be tested again using either individual vector administration or low-complexity libraries with barcoded reporter constructs. This would allow assessment of complete transduction profiles after systemic delivery, in order (i) to evaluate targeting across multiple tissues to identify off-targeting to organs other than the liver, and (ii) to compare the selected candidates to benchmark capsids such as AAV6 for non-parenchymal liver cells or AAV8 for hepatocyte transduction [297,404].

Having shown that CMV promoter-based *cap* expression allows readout from multiple different cell types, my findings combined with recent advances in RNA-based AAV library

screening including my own demonstrate the benefits of this approach for enrichment of functional variants. While the CMV promoter-based construct may be further advanced by improving the reporter cassette or using stronger ubiquitous promoters, the general approach is highly advantageous for variant assessment during the selection process. Comparing variant composition in the plasmid library, the input virus library, as well as DNA and RNA of transduced cells offers powerful insights into the functionality of each variant. Such datasets are uniquely compatible with machine-guided capsid design and may benefit future data-driven capsid engineering [19,416].

AAV-based targeting of specific liver cell types is desired for multiple applications. One use is the cell-type-selective downregulation of CD44 for dissection and treatment of (non-alcoholic) metabolic dysfunction-associated steatohepatitis (MASH or NASH) as described in section 4.1.3. Here, target cells are likely to be Kupffer cells, which contribute to CD44-mediated inflammation [260]. Other possible applications of these capsids include targeting of hepatic stellate cells (HSCs) for countering liver fibrosis [277], targeting of liver sinusoidal endothelial cells (LSECs) for replicating physiological factor VIII expression or inducing immunotolerance to AAV-delivered transgene products [417], and targeting of cholangiocytes for treatment of cystic fibrosis liver disease [418]. With these applications in mind, the novel liver cell-type-specific AAV capsids presented here may assist multiple future uses in biomedical research and gene therapy of hepatic diseases.

## 4.4 Conclusions and outlook

In this thesis, I developed and characterized a series of highly compatible advancements of the AAV vector system on the levels of targeted RNA degradation, optimized transgene expression and *in vivo* selection stratification. The new vector designs and protocols established and validated here provide a powerful and versatile basis for further improvement towards the ultimate aim, to reach the high safety and efficiency levels required for successful gene therapy applications.

Advances in AAV capsid library screening have allowed me to exert more control over the selection process and gain deeper insights into capsid functionality. To improve selection parameters of directed AAV capsid evolution, I investigated a Cas9-based negative selection approach (DEPOOL) that enables the targeted removal of unwanted variants. This strategy can improve future screens by employing a negative selection pressure based solely on *cap* or barcode sequence identity. To improve the functionality of enriched variants, I implemented an RNA-based selection platform. Using a ubiquitous CMV promoter, I could trace variants of an AAV6 peptide display library both in DNA and RNA (cDNA) of transduced liver cells after systemic administration to mice. The ubiquitous expression allowed *cap* tracing across different cell types to study enrichment in both on- and off-target cells. This forced rapid evolution of a small set of variants with distinct transduction properties tested across multiple liver cell types. Specific lead candidates can now be studied in detail by single-vector administration or barcoded reporter assays. This will enable the establishment of their complete systemic transduction profiles and comparisons to their parental AAV6 capsid and other benchmark variants. Among other applications, capsid variants with confirmed cell-type-specificity in liver may then be employed for targeted downregulation of *CD44,* a key regulator of non-alcoholic (metabolic) steatohepatitis. Using the shRNA-based knockdown effectors I have validated *in vitro*, cell-type-specific *CD44* knockdown may aid in understanding and countering the progression of this disease and thus benefit the development of future gene therapy products.

To counteract cellular infections with SARS-CoV-2, I applied single- or multiplexed shRNA vectors in a prophylactic setting prior to viral infection. I could demonstrate that SARS-CoV-2 is susceptible to RNAi, but quickly evolves escape mutations unless challenged by a cocktail of shRNAs. This is an essential finding that highlights the necessity for future possible RNAi-based interventions against SARS-CoV-2 (or other viruses) to consider multiplexed targeting and continued variant monitoring.

Following up on a parallel cross-tissue evaluation of multiple promoters, I could demonstrate that the human GFAP promoter mediates highly efficient transgene expression in hepatocytes after AAV-based delivery. Although this promoter is relatively large, GFAP promoter-based transgene expression may be used in future AAV vectors for liver-directed gene therapy applications that require high levels of transgene expression with low input vector doses. Possible follow-up studies may focus on identifying transcription factor binding sites within the GFAP promoter that facilitate its strong hepatocyte-directed expression. This information may then be used to create truncated promoter variants with maintained expression profiles but better compatibility with larger transgenes.

My advancements in different areas of AAV vector engineering demonstrate how the current limitations in transduction efficiency and specificity can be overcome in the future. By combining novel synthetic AAV capsids with improved expression modulators, highly efficient and cell-type-specific transgene expression is possible for facilitating therapeutic overexpression or targeted silencing of disease-associated genes. These strategies will complement other progresses in AAV vector development towards a new generation of safer gene therapy vectors.

# 5 References


1. Atchison, R.W., Casto, B.C., and Hammon, W.M. (1965). Adenovirus-Associated Defective Virus Particles. Science *149*, 754-755. 10.1126/science.149.3685.754.
2. Hoggan, M.D., Blacklow, N.R., and Rowe, W.P. (1966). Studies of small DNA viruses found in various adenovirus preparations: physical, biological, and immunological characteristics. Proc Natl Acad Sci U S A *55*, 1467-1474. 10.1073/pnas.55.6.1467.
3. Ogston, P., Raj, K., and Beard, P. (2000). Productive replication of adeno-associated virus can occur in human papillomavirus type 16 (HPV-16) episome-containing keratinocytes and is augmented by the HPV-16 E2 protein. J Virol *74*, 3494-3504. 10.1128/jvi.74.8.3494-3504.2000.
4. Georg-Fries, B., Biederlack, S., Wolf, J., and zur Hausen, H. (1984). Analysis of proteins, helper dependence, and seroepidemiology of a new human parvovirus. Virology *134*, 64-71. 10.1016/0042-6822(84)90272-1.
5. Meier, A.F., Fraefel, C., and Seyffert, M. (2020). The Interplay between Adeno-Associated Virus and its Helper Viruses. Viruses *12*. 10.3390/v12060662.
6. Yalkinoglu, A.O., Heilbronn, R., Bürkle, A., Schlehofer, J.R., and zur Hausen, H. (1988). DNA amplification of adeno-associated virus as a response to cellular genotoxic stress. Cancer Res *48*, 3123-3129.
7. Yakobson, B., Koch, T., and Winocour, E. (1987). Replication of adeno-associated virus in synchronized cells without the addition of a helper virus. J Virol *61*, 972-981. 10.1128/jvi.61.4.972-981.1987.
8. Berns, K.I., and Adler, S. (1972). Separation of two types of adeno-associated virus particles containing complementary polynucleotide chains. J Virol *9*, 394-396. 10.1128/jvi.9.2.394-396.1972.
9. Srivastava, A., Lusby, E.W., and Berns, K.I. (1983). Nucleotide sequence and organization of the adeno-associated virus 2 genome. J Virol *45*, 555-564. 10.1128/jvi.45.2.555-564.1983.
10. Green, M.R., and Roeder, R.G. (1980). Transcripts of the adeno-associated virus genome: mapping of the major RNAs. J Virol *36*, 79-92. 10.1128/jvi.36.1.79-92.1980.
11. Mendelson, E., Trempe, J.P., and Carter, B.J. (1986). Identification of the trans-acting Rep proteins of adeno-associated virus by antibodies to a synthetic oligopeptide. J Virol *60*, 823-832. 10.1128/jvi.60.3.823-832.1986.
12. McCarty, D.M., Pereira, D.J., Zolotukhin, I., Zhou, X., Ryan, J.H., and Muzyczka, N. (1994). Identification of linear DNA sequences that specifically bind the adeno-associated virus Rep protein. J Virol *68*, 4988-4997. 10.1128/jvi.68.8.4988-4997.1994.
13. Im, D.S., and Muzyczka, N. (1992). Partial purification of adeno-associated virus Rep78, Rep52, and Rep40 and their biochemical characterization. J Virol *66*, 1119-1128. 10.1128/jvi.66.2.1119-1128.1992.
14. Dong, J.Y., Fan, P.D., and Frizzell, R.A. (1996). Quantitative analysis of the packaging capacity of recombinant adeno-associated virus. Hum Gene Ther *7*, 2101-2112. 10.1089/hum.1996.7.17-2101.
15. Becerra, S.P., Koczot, F., Fabisch, P., and Rose, J.A. (1988). Synthesis of adeno-associated virus structural proteins requires both alternative mRNA splicing and alternative initiations from a single transcript. J Virol *62*, 2745-2754. 10.1128/jvi.62.8.2745-2754.1988.
16. Trempe, J.P., and Carter, B.J. (1988). Alternate mRNA splicing is required for synthesis of adeno-associated virus VP1 capsid protein. J Virol *62*, 3356-3363. 10.1128/jvi.62.9.3356-3363.1988.
17. Buller, R.M., and Rose, J.A. (1978). Characterization of adenovirus-associated virus-induced polypeptides in KB cells. J Virol *25*, 331-338. 10.1128/jvi.25.1.331-338.1978.
18. Snijder, J., van de Waterbeemd, M., Damoc, E., Denisov, E., Grinfeld, D., Bennett, A., Agbandje-McKenna, M., Makarov, A., and Heck, A.J. (2014). Defining the stoichiometry and cargo load of viral and bacterial nanoparticles by Orbitrap mass spectrometry. J Am Chem Soc *136*, 7295-7299. 10.1021/ja502616y.
19. Ogden, P.J., Kelsic, E.D., Sinai, S., and Church, G.M. (2019). Comprehensive AAV capsid fitness landscape reveals a viral gene and enables machine-guided design. Science *366*, 1139-1143. 10.1126/science.aaw2900.
20. Sonntag, F., Schmidt, K., and Kleinschmidt, J.A. (2010). A viral assembly factor promotes AAV2 capsid formation in the nucleolus. Proc Natl Acad Sci U S A *107*, 10220-10225. 10.1073/pnas.1001673107.
21. Galibert, L., Hyvönen, A., Eriksson, R.A.E., Mattola, S., Aho, V., Salminen, S., Albers, J.D., Peltola, S.K., Weman, S., Nieminen, T., Ylä-Herttuala, S., et al. (2021). Functional roles of the membrane-associated AAV protein MAAP. Sci Rep *11*, 21698. 10.1038/s41598-021-01220-7.
22. Hull, J.A., Mietzsch, M., Chipman, P., Strugatsky, D., and McKenna, R. (2022). Structural characterization of an envelope-associated adeno-associated virus type 2 capsid. Virology *565*, 22-28. 10.1016/j.virol.2021.09.010.
23. Earley, L.F., Powers, J.M., Adachi, K., Baumgart, J.T., Meyer, N.L., Xie, Q., Chapman, M.S., and Nakai, H. (2017). Adeno-associated Virus (AAV) Assembly-Activating Protein Is Not an Essential Requirement for Capsid Assembly of AAV Serotypes 4, 5, and 11. J Virol *91*. 10.1128/jvi.01980-16.
24. Mietzsch, M., Pénzes, J.J., and Agbandje-McKenna, M. (2019). Twenty-Five Years of Structural Parvovirology. Viruses *11*. 10.3390/v11040362.
25. Xie, Q., Bu, W., Bhatia, S., Hare, J., Somasundaram, T., Azzi, A., and Chapman, M.S. (2002). The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. Proc Natl Acad Sci U S A *99*, 10405-10410. 10.1073/pnas.162250899.
26. Mietzsch, M., Jose, A., Chipman, P., Bhattacharya, N., Daneshparvar, N., McKenna, R., and Agbandje-McKenna, M. (2021). Completion of the AAV Structural Atlas: Serotype Capsid Structures Reveals Clade-Specific Features. Viruses *13*. 10.3390/v13010101.
27. Issa, S.S., Shaimardanova, A.A., Solovyeva, V.V., and Rizvanov, A.A. (2023). Various AAV Serotypes and Their Applications in Gene Therapy: An Overview. Cells *12*. 10.3390/cells12050785.
28. Gao, G., Vandenberghe, L.H., Alvira, M.R., Lu, Y., Calcedo, R., Zhou, X., and Wilson, J.M. (2004). Clades of Adeno-associated viruses are widely disseminated in human tissues. J Virol *78*, 6381-6388. 10.1128/jvi.78.12.6381-6388.2004.

29. Louis Jeune, V., Joergensen, J.A., Hajjar, R.J., and Weber, T. (2013). Pre-existing anti-adeno-associated virus antibodies as a challenge in AAV gene therapy. Hum Gene Ther Methods *24*, 59-67. 10.1089/hgtb.2012.243.
30. Weber, T. (2021). Anti-AAV Antibodies in AAV Gene Therapy: Current Challenges and Possible Solutions. Front Immunol *12*, 658399. 10.3389/fimmu.2021.658399.
31. Boutin, S., Monteilhet, V., Veron, P., Leborgne, C., Benveniste, O., Montus, M.F., and Masurier, C. (2010). Prevalence of serum IgG and neutralizing factors against adeno-associated virus (AAV) types 1, 2, 5, 6, 8, and 9 in the healthy population: implications for gene therapy using AAV vectors. Hum Gene Ther *21*, 704-712. 10.1089/hum.2009.182.
32. Smith, R.H. (2008). Adeno-associated virus integration: virus versus vector. Gene Ther *15*, 817-822. 10.1038/gt.2008.55.
33. Ho, A., Orton, R., Tayler, R., Asamaphan, P., Herder, V., Davis, C., Tong, L., Smollett, K., Manali, M., Allan, J., Rawlik, K., et al. (2023). Adeno-associated virus 2 infection in children with non-A–E hepatitis. Nature *617*, 555-563. 10.1038/s41586-023-05948-2.
34. Morfopoulou, S., Buddle, S., Torres Montaguth, O.E., Atkinson, L., Guerra-Assunção, J.A., Moradi Marjaneh, M., Zennezini Chiozzi, R., Storey, N., Campos, L., Hutchinson, J.C., Counsell, J.R., et al. (2023). Genomic investigations of unexplained acute hepatitis in children. Nature *617*, 564-573. 10.1038/s41586-023-06003-w.
35. Servellita, V., Sotomayor Gonzalez, A., Lamson, D.M., Foresythe, A., Huh, H.J., Bazinet, A.L., Bergman, N.H., Bull, R.L., Garcia, K.Y., Goodrich, J.S., Lovett, S.P., et al. (2023). Adeno-associated virus type 2 in US children with acute severe hepatitis. Nature *617*, 574-580. 10.1038/s41586-023-05949-1.
36. Calcedo, R., and Wilson, J.M. (2013). Humoral Immune Response to AAV. Front Immunol *4*, 341. 10.3389/fimmu.2013.00341.
37. Hildebrandt, E., Penzes, J.J., Gifford, R.J., Agbandje-Mckenna, M., and Kotin, R.M. (2020). Evolution of dependoparvoviruses across geological timescales—implications for design of AAV-based gene therapy vectors. Virus Evol *6*, veaa043. 10.1093/ve/veaa043.
38. Zhu, C., Wang, C., Wu, J., Ye, F., Lv, R., Hu, D., Ai, L., Yang, L., Wu, T., Li, B., Ding, C., et al. (2020). Distribution and genetic diversity of adeno-associated viruses in bats from coastal areas of Southeast China. Sci Rep *10*, 3725. 10.1038/s41598-020-60721-z.
39. Srivastava, A. (2016). In vivo tissue-tropism of adeno-associated viral vectors. Curr Opin Virol *21*, 75-80. 10.1016/j.coviro.2016.08.003.
40. Maurer, A.C., and Weitzman, M.D. (2020). Adeno-Associated Virus Genome Interactions Important for Vector Production and Transduction. Hum Gene Ther *31*, 499-511. 10.1089/hum.2020.069.
41. Kern, A., Schmidt, K., Leder, C., Müller, O.J., Wobus, C.E., Bettinger, K., Von der Lieth, C.W., King, J.A., and Kleinschmidt, J.A. (2003). Identification of a Heparin-Binding Motif on Adeno-Associated Virus Type 2 Capsids. J Virol *77*, 11072-11081. 10.1128/JVI.77.20.11072-11081.2003.
42. Opie, S.R., Warrington, K.H., Jr., Agbandje-McKenna, M., Zolotukhin, S., and Muzyczka, N. (2003). Identification of amino acid residues in the capsid proteins of adeno-associated virus type 2 that contribute to heparan sulfate proteoglycan binding. J Virol *77*, 6995-7006. 10.1128/jvi.77.12.6995-7006.2003.
43. Summerford, C., and Samulski Richard, J. (1998). Membrane-Associated Heparan Sulfate Proteoglycan Is a Receptor for Adeno-Associated Virus Type 2 Virions. J Virol *72*, 1438-1445. 10.1128/jvi.72.2.1438-1445.1998.
44. Summerford, C., Bartlett, J.S., and Samulski, R.J. (1999). αVβ5 integrin: a co-receptor for adeno-associated virus type 2 infection. Nat Med *5*, 78-82. 10.1038/4768.
45. Asokan, A., Hamra Julie, B., Govindasamy, L., Agbandje-McKenna, M., and Samulski Richard, J. (2006). Adeno-Associated Virus Type 2 Contains an Integrin α5β1 Binding Domain Essential for Viral Cell Entry. J Virol *80*, 8961-8969. 10.1128/jvi.00843-06.
46. Qing, K., Mah, C., Hansen, J., Zhou, S., Dwarki, V., and Srivastava, A. (1999). Human fibroblast growth factor receptor 1 is a co-receptor for infection by adeno-associated virus 2. Nat Med *5*, 71-77. 10.1038/4758.
47. Huang, L.Y., Halder, S., and Agbandje-McKenna, M. (2014). Parvovirus glycan interactions. Curr Opin Virol *7*, 108-118. 10.1016/j.coviro.2014.05.007.
48. Di Pasquale, G., Davidson, B.L., Stein, C.S., Martins, I., Scudiero, D., Monks, A., and Chiorini, J.A. (2003). Identification of PDGFR as a receptor for AAV-5 transduction. Nat Med *9*, 1306-1312. 10.1038/nm929.
49. Weller, M.L., Amornphimoltham, P., Schmidt, M., Wilson, P.A., Gutkind, J.S., and Chiorini, J.A. (2010). Epidermal growth factor receptor is a co-receptor for adeno-associated virus serotype 6. Nat Med *16*, 662-664. 10.1038/nm.2145.
50. Dudek, A.M., Pillay, S., Puschnik, A.S., Nagamine, C.M., Cheng, F., Qiu, J., Carette, J.E., and Vandenberghe, L.H. (2018). An Alternate Route for Adeno-associated Virus (AAV) Entry Independent of AAV Receptor. J Virol *92*. 10.1128/jvi.02213-17.
51. Pillay, S., Meyer, N.L., Puschnik, A.S., Davulcu, O., Diep, J., Ishikawa, Y., Jae, L.T., Wosen, J.E., Nagamine, C.M., Chapman, M.S., and Carette, J.E. (2016). An essential receptor for adeno-associated virus infection. Nature *530*, 108-112. 10.1038/nature16465.
52. Duan, D., Li, Q., Kao, A.W., Yue, Y., Pessin, J.E., and Engelhardt, J.F. (1999). Dynamin is required for recombinant adeno-associated virus type 2 infection. J Virol *73*, 10371-10376. 10.1128/jvi.73.12.10371-10376.1999.
53. Nonnenmacher, M., and Weber, T. (2011). Adeno-associated virus 2 infection requires endocytosis through the CLIC/GEEC pathway. Cell Host Microbe *10*, 563-576. 10.1016/j.chom.2011.10.014.
54. Sanlioglu, S., Benson, P.K., Yang, J., Atkinson, E.M., Reynolds, T., and Engelhardt, J.F. (2000). Endocytosis and nuclear trafficking of adeno-associated virus type 2 are controlled by rac1 and phosphatidylinositol-3 kinase activation. J Virol *74*, 9184-9196. 10.1128/jvi.74.19.9184-9196.2000.
55. Nam, H.J., Gurda, B.L., McKenna, R., Potter, M., Byrne, B., Salganik, M., Muzyczka, N., and Agbandje-McKenna, M. (2011). Structural studies of adeno-associated virus serotype 8 capsid transitions associated with endosomal trafficking. J Virol *85*, 11791-11799. 10.1128/jvi.05305-11.

56. Nonnenmacher, M.E., Cintrat, J.C., Gillet, D., and Weber, T. (2015). Syntaxin 5-dependent retrograde transport to the trans-Golgi network is required for adeno-associated virus transduction. J Virol *89*, 1673-1687. 10.1128/jvi.02520-14.
57. Girod, A., Wobus, C.E., Zádori, Z., Ried, M., Leike, K., Tijssen, P., Kleinschmidt, J.A., and Hallek, M. (2002). The VP1 capsid protein of adeno-associated virus type 2 is carrying a phospholipase A2 domain required for virus infectivity. J Gen Virol *83*, 973-978. 10.1099/0022-1317-83-5-973.
58. Yan, Z., Zak, R., Luxton, G.W., Ritchie, T.C., Bantel-Schaal, U., and Engelhardt, J.F. (2002). Ubiquitination of both adeno-associated virus type 2 and 5 capsid proteins affects the transduction efficiency of recombinant vectors. J Virol *76*, 2043-2053. 10.1128/jvi.76.5.2043-2053.2002.
59. Xiao, P.J., and Samulski, R.J. (2012). Cytoplasmic trafficking, endosomal escape, and perinuclear accumulation of adeno-associated virus type 2 particles are facilitated by microtubule network. J Virol *86*, 10462-10473. 10.1128/jvi.00935-12.
60. Junod, S.L., Saredy, J., and Yang, W. (2021). Nuclear Import of Adeno-Associated Viruses Imaged by High-Speed Single-Molecule Microscopy. Viruses *13*. 10.3390/v13020167.
61. Sonntag, F., Bleker, S., Leuchs, B., Fischer, R., and Kleinschmidt, J.A. (2006). Adeno-associated virus type 2 capsids with externalized VP1/VP2 trafficking domains are generated prior to passage through the cytoplasm and are maintained until uncoating occurs in the nucleus. J Virol *80*, 11040-11054. 10.1128/jvi.01056-06.
62. Johnson, J.S., and Samulski, R.J. (2009). Enhancement of adeno-associated virus infection by mobilizing capsids into and out of the nucleolus. J Virol *83*, 2632-2644. 10.1128/jvi.02309-08.
63. Hong, G., Ward, P., and Berns, K.I. (1992). In vitro replication of adeno-associated virus DNA. Proc Natl Acad Sci U S A *89*, 4673-4677. 10.1073/pnas.89.10.4673.
64. Hörer, M., Weger, S., Butz, K., Hoppe-Seyler, F., Geisen, C., and Kleinschmidt, J.A. (1995). Mutational analysis of adeno-associated virus Rep protein-mediated inhibition of heterologous and homologous promoters. J Virol *69*, 5485-5496. 10.1128/jvi.69.9.5485-5496.1995.
65. McCarty, D.M., Young, S.M., Jr., and Samulski, R.J. (2004). Integration of adeno-associated virus (AAV) and recombinant AAV vectors. Annu Rev Genet *38*, 819-845. 10.1146/annurev.genet.37.110801.143717.
66. Samulski, R.J., Zhu, X., Xiao, X., Brook, J.D., Housman, D.E., Epstein, N., and Hunter, L.A. (1991). Targeted integration of adeno-associated virus (AAV) into human chromosome 19. Embo J *10*, 3941-3950. 10.1002/j.1460-2075.1991.tb04964.x.
67. Weitzman, M.D., Kyöstiö, S.R., Kotin, R.M., and Owens, R.A. (1994). Adeno-associated virus (AAV) Rep proteins mediate complex formation between AAV DNA and its integration site in human DNA. Proc Natl Acad Sci U S A *91*, 5808-5812. 10.1073/pnas.91.13.5808.
68. Pereira, D.J., McCarty, D.M., and Muzyczka, N. (1997). The adeno-associated virus (AAV) Rep protein acts as both a repressor and an activator to regulate AAV transcription during a productive infection. J Virol *71*, 1079-1088. 10.1128/jvi.71.2.1079-1088.1997.
69. Kyöstiö, S.R., Owens, R.A., Weitzman, M.D., Antoni, B.A., Chejanovsky, N., and Carter, B.J. (1994). Analysis of adeno-associated virus (AAV) wild-type and mutant Rep proteins for their abilities to negatively regulate AAV p5 and p19 mRNA levels. J Virol *68*, 2947-2957. 10.1128/jvi.68.5.2947-2957.1994.
70. Shi, Y., Seto, E., Chang, L.S., and Shenk, T. (1991). Transcriptional repression by YY1, a human GLI-Krüppel-related protein, and relief of repression by adenovirus E1A protein. Cell *67*, 377-388. 10.1016/0092-8674(91)90189-6.
71. Pereira, D.J., and Muzyczka, N. (1997). The cellular transcription factor SP1 and an unknown cellular protein are required to mediate Rep protein activation of the adeno-associated virus p19 promoter. J Virol *71*, 1747-1756. 10.1128/jvi.71.3.1747-1756.1997.
72. Mouw, M.B., and Pintel, D.J. (2000). Adeno-associated virus RNAs appear in a temporal order and their splicing is stimulated during coinfection with adenovirus. J Virol *74*, 9878-9888. 10.1128/jvi.74.21.9878-9888.2000.
73. McCarty, D.M., Christensen, M., and Muzyczka, N. (1991). Sequences required for coordinate induction of adeno-associated virus p19 and p40 promoters by Rep protein. J Virol *65*, 2936-2945. 10.1128/jvi.65.6.2936-2945.1991.
74. Westhaus, A., Cabanes-Creus, M., Jonker, T., Sallard, E., Navarro, R.G., Zhu, E., Baltazar Torres, G., Lee, S., Wilmott, P., Gonzalez-Cordero, A., Santilli, G., et al. (2022). AAV-p40 Bioengineering Platform for Variant Selection Based on Transgene Expression. Hum Gene Ther. 10.1089/hum.2021.278.
75. Im, D.S., and Muzyczka, N. (1990). The AAV origin binding protein Rep68 is an ATP-dependent site-specific endonuclease with DNA helicase activity. Cell *61*, 447-457. 10.1016/0092-8674(90)90526-k.
76. Snyder, R.O., and Flotte, T.R. (2002). Production of clinical-grade recombinant adeno-associated virus vectors. Curr Opin Biotechnol *13*, 418-423. 10.1016/S0958-1669(02)00369-5.
77. Grosse, S., Penaud-Budloo, M., Herrmann, A.K., Börner, K., Fakhiri, J., Laketa, V., Krämer, C., Wiedtke, E., Gunkel, M., Ménard, L., Ayuso, E., et al. (2017). Relevance of Assembly-Activating Protein for Adeno-associated Virus Vector Production and Capsid Protein Stability in Mammalian and Insect Cells. J Virol *91*. 10.1128/jvi.01198-17.
78. King, J.A., Dubielzig, R., Grimm, D., and Kleinschmidt, J.A. (2001). DNA helicase-mediated packaging of adeno-associated virus type 2 genomes into preformed capsids. Embo J *20*, 3282-3291. 10.1093/emboj/20.12.3282.
79. Bleker, S., Sonntag, F., and Kleinschmidt, J.A. (2005). Mutational analysis of narrow pores at the fivefold symmetry axes of adeno-associated virus type 2 capsids reveals a dual role in genome packaging and activation of phospholipase A2 activity. J Virol *79*, 2528-2540. 10.1128/jvi.79.4.2528-2540.2005.
80. Senapathy, P., Tratschin, J.D., and Carter, B.J. (1984). Replication of adeno-associated virus DNA. Complementation of naturally occurring rep- mutants by a wild-type genome or an ori- mutant and correction of terminal palindrome deletions. J Mol Biol *179*, 1-20. 10.1016/0022-2836(84)90303-6.
81. Samulski, R.J., Berns, K.I., Tan, M., and Muzyczka, N. (1982). Cloning of adeno-associated virus into pBR322: rescue of intact virus from the recombinant plasmid in human cells. Proc Natl Acad Sci U S A *79*, 2077-2081. 10.1073/pnas.79.6.2077.

82. Hermonat, P.L., and Muzyczka, N. (1984). Use of adeno-associated virus as a mammalian DNA cloning vector: transduction of neomycin resistance into mammalian tissue culture cells. Proc Natl Acad Sci U S A *81*, 6466-6470. 10.1073/pnas.81.20.6466.
83. Wang, D., Tai, P.W.L., and Gao, G. (2019). Adeno-associated virus vector as a platform for gene therapy delivery. Nat Rev Drug Discov *18*, 358-378. 10.1038/s41573-019-0012-9.
84. McLaughlin, S.K., Collis, P., Hermonat, P.L., and Muzyczka, N. (1988). Adeno-associated virus general transduction vectors: analysis of proviral structures. J Virol *62*, 1963-1973. 10.1128/jvi.62.6.1963-1973.1988.
85. Samulski, R.J., Chang, L.S., and Shenk, T. (1989). Helper-free stocks of recombinant adeno-associated viruses: normal integration does not require viral gene expression. J Virol *63*, 3822-3828. 10.1128/jvi.63.9.3822-3828.1989.
86. Xiao, X., Li, J., and Samulski, R.J. (1998). Production of high-titer recombinant adeno-associated virus vectors in the absence of helper adenovirus. J Virol *72*, 2224-2232. 10.1128/jvi.72.3.2224-2232.1998.
87. Matsushita, T., Elliger, S., Elliger, C., Podsakoff, G., Villarreal, L., Kurtzman, G.J., Iwaki, Y., and Colosi, P. (1998). Adeno-associated virus vectors can be efficiently produced without helper virus. Gene Ther *5*, 938-945. 10.1038/sj.gt.3300680.
88. Grimm, D., Kern, A., Rittner, K., and Kleinschmidt, J.A. (1998). Novel tools for production and purification of recombinant adenoassociated virus vectors. Hum Gene Ther *9*, 2745-2760. 10.1089/hum.1998.9.18-2745.
89. Gao, G.-P., Qu, G., Faust, L.Z., Engdahl, R.K., Xiao, W., Hughes, J.V., Zoltick, P.W., and Wilson, J.M. (1998). High-Titer Adeno-Associated Viral Vectors from a Rep/Cap Cell Line and Hybrid Shuttle Virus. Hum Gene Ther *9*, 2353-2362. 10.1089/hum.1998.9.16-2353.
90. Wu, Y., Mei, T., Jiang, L., Han, Z., Dong, R., Yang, T., and Xu, F. (2019). Development of Versatile and Flexible Sf9 Packaging Cell Line-Dependent OneBac System for Large-Scale Recombinant Adeno-Associated Virus Production. Hum Gene Ther Methods *30*, 172-183. 10.1089/hgtb.2019.123.
91. El Andari, J., and Grimm, D. (2021). Production, Processing, and Characterization of Synthetic AAV Gene Therapy Vectors. Biotechnol J *16*, 2000025. 10.1002/biot.202000025.
92. Wright, J.F. (2014). AAV empty capsids: for better or for worse? Mol Ther *22*, 1-2. 10.1038/mt.2013.268.
93. Tran, N.T., Heiner, C., Weber, K., Weiand, M., Wilmot, D., Xie, J., Wang, D., Brown, A., Manokaran, S., Su, Q., Zapp, M.L., et al. (2020). AAV-Genome Population Sequencing of Vectors Packaging CRISPR Components Reveals Design-Influenced Heterogeneity. Mol Ther Methods Clin Dev *18*, 639-651. 10.1016/j.omtm.2020.07.007.
94. Duan, D., Sharma, P., Yang, J., Yue, Y., Dudus, L., Zhang, Y., Fisher, K.J., and Engelhardt, J.F. (1998). Circular intermediates of recombinant adeno-associated virus have defined structural characteristics responsible for long-term episomal persistence in muscle tissue. J Virol *72*, 8568-8577. 10.1128/jvi.72.11.8568-8577.1998.
95. Sabatino, D.E., Bushman, F.D., Chandler, R.J., Crystal, R.G., Davidson, B.L., Dolmetsch, R., Eggan, K.C., Gao, G., Gil-Farina, I., Kay, M.A., McCarty, D.M., et al. (2022). Evaluating the state of the science for adeno-associated virus integration: An integrated perspective. Mol Ther *30*, 2646-2663. 10.1016/j.ymthe.2022.06.004.
96. Nguyen, G.N., Everett, J.K., Kafle, S., Roche, A.M., Raymond, H.E., Leiby, J., Wood, C., Assenmacher, C.A., Merricks, E.P., Long, C.T., Kazazian, H.H., et al. (2021). A long-term study of AAV gene therapy in dogs with hemophilia A identifies clonal expansions of transduced liver cells. Nat Biotechnol *39*, 47-55. 10.1038/s41587-020-0741-7.
97. Ferrari, F.K., Samulski, T., Shenk, T., and Samulski, R.J. (1996). Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. J Virol *70*, 3227-3234. 10.1128/jvi.70.5.3227-3234.1996.
98. McCarty, D.M., Fu, H., Monahan, P.E., Toulson, C.E., Naik, P., and Samulski, R.J. (2003). Adeno-associated virus terminal repeat (TR) mutant generates self-complementary vectors to overcome the rate-limiting step to transduction in vivo. Gene Ther *10*, 2112-2118. 10.1038/sj.gt.3302134.
99. Wang, Z., Ma, H.I., Li, J., Sun, L., Zhang, J., and Xiao, X. (2003). Rapid and highly efficient transduction by double-stranded adeno-associated virus vectors in vitro and in vivo. Gene Ther *10*, 2105-2111. 10.1038/sj.gt.3302133.
100. Wu, J., Zhao, W., Zhong, L., Han, Z., Li, B., Ma, W., Weigel-Kelley, K.A., Warrington, K.H., and Srivastava, A. (2007). Self-complementary recombinant adeno-associated viral vectors: packaging capacity and the role of rep proteins in vector purity. Hum Gene Ther *18*, 171-182. 10.1089/hum.2006.088.
101. Xiao, W., Chirmule, N., Berta, S.C., McCullough, B., Gao, G., and Wilson, J.M. (1999). Gene therapy vectors based on adeno-associated virus type 1. J Virol *73*, 3994-4003. 10.1128/jvi.73.5.3994-4003.1999.
102. Gao, G.P., Alvira, M.R., Wang, L., Calcedo, R., Johnston, J., and Wilson, J.M. (2002). Novel adeno-associated viruses from rhesus monkeys as vectors for human gene therapy. Proc Natl Acad Sci U S A *99*, 11854-11859. 10.1073/pnas.182412299.
103. Burger, C., Gorbatyuk, O.S., Velardo, M.J., Peden, C.S., Williams, P., Zolotukhin, S., Reier, P.J., Mandel, R.J., and Muzyczka, N. (2004). Recombinant AAV Viral Vectors Pseudotyped with Viral Capsids from Serotypes 1, 2, and 5 Display Differential Efficiency and Cell Tropism after Delivery to Different Regions of the Central Nervous System. Mol Ther *10*, 302-317. 10.1016/j.ymthe.2004.05.024.
104. Flotte, T.R., Afione, S.A., Conrad, C., McGrath, S.A., Solow, R., Oka, H., Zeitlin, P.L., Guggino, W.B., and Carter, B.J. (1993). Stable in vivo expression of the cystic fibrosis transmembrane conductance regulator with an adeno-associated virus vector. Proc Natl Acad Sci U S A *90*, 10613-10617. 10.1073/pnas.90.22.10613.
105. Flotte, T., Carter, B., Conrad, C., Guggino, W., Reynolds, T., Rosenstein, B., Taylor, G., Walden, S., and Wetzel, R. (1996). A phase I study of an adeno-associated virus-CFTR gene vector in adult CF patients with mild lung disease. Hum Gene Ther *7*, 1145-1159. 10.1089/hum.1996.7.9-1145.
106. Flotte, T.R., Zeitlin, P.L., Reynolds, T.C., Heald, A.E., Pedersen, P., Beck, S., Conrad, C.K., Brass-Ernst, L., Humphries, M., Sullivan, K., Wetzel, R., et al. (2003). Phase I trial of intranasal and endobronchial administration of a recombinant adeno-associated virus serotype 2 (rAAV2)-CFTR vector in adult cystic fibrosis patients: a two-part clinical study. Hum Gene Ther *14*, 1079-1088. 10.1089/104303403322124792.

107. Wiley Database: "Gene Therapy Clinical Trials Worldwide". (Last accessed: 20.08.2023). https://a873679.fmphost.com/fmi/webd/GTCT. The Journal of Gene Medicine.
108. Scott, L.J. (2015). Alipogene Tiparvovec: A Review of Its Use in Adults with Familial Lipoprotein Lipase Deficiency. Drugs *75*, 175-182. 10.1007/s40265-014-0339-9.
109. Russell, S., Bennett, J., Wellman, J.A., Chung, D.C., Yu, Z.-F., Tillman, A., Wittes, J., Pappas, J., Elci, O., McCague, S., Cross, D., et al. (2017). Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. Lancet *390*, 849-860. 10.1016/S0140-6736(17)31868-8.
110. Hoy, S.M. (2019). Onasemnogene Abeparvovec: First Global Approval. Drugs *79*, 1255-1262. 10.1007/s40265-019-01162-5.
111. Keam, S.J. (2022). Eladocagene Exuparvovec: First Approval. Drugs *82*, 1427-1432. 10.1007/s40265-022-01775-3.
112. Blair, H.A. (2022). Valoctocogene Roxaparvovec: First Approval. Drugs *82*, 1505-1510. 10.1007/s40265-022-01788-y.
113. Heo, Y.A. (2023). Etranacogene Dezaparvovec: First Approval. Drugs *83*, 347-352. 10.1007/s40265-023-01845-0.
114. Hoy, S.M. (2023). Delandistrogene Moxeparvovec: First Approval. Drugs. 10.1007/s40265-023-01929-x.
115. Moran, N. (2012). First gene therapy approved. Nat Biotechnol *30*, 1153-1153. 10.1038/nbt1212-1153.
116. Burnett, J.R., and Hooper, A.J. (2009). Alipogene tiparvovec, an adeno-associated virus encoding the Ser(447)X variant of the human lipoprotein lipase gene for the treatment of patients with lipoprotein lipase deficiency. Curr Opin Mol Ther *11*, 681-691.
117. Senior, M. (2017). After Glybera's withdrawal, what's next for gene therapy? Nat Biotechnol *35*, 491-492. 10.1038/nbt0617-491.
118. Keeler, A.M., and Flotte, T.R. (2019). Recombinant Adeno-Associated Virus Gene Therapy in Light of Luxturna (and Zolgensma and Glybera): Where Are We, and How Did We Get Here? Annu Rev Virol *6*, 601-621. 10.1146/annurev-virology-092818-015530.
119. Shen, W., Liu, S., and Ou, L. (2022). rAAV immunogenicity, toxicity, and durability in 255 clinical trials: A meta-analysis. Front Immunol *13*, 1001263. 10.3389/fimmu.2022.1001263.
120. Ertl, H.C.J. (2022). Immunogenicity and toxicity of AAV gene therapy. Front Immunol *13*, 975803. 10.3389/fimmu.2022.975803.
121. Shieh, P.B., Bönnemann, C.G., Müller-Felber, W., Blaschek, A., Dowling, J.J., Kuntz, N.L., and Seferian, A.M. (2020). Re: "Moving Forward After Two Deaths in a Gene Therapy Trial of Myotubular Myopathy" by Wilson and Flotte. Hum Gene Ther *31*, 787. 10.1089/hum.2020.217.
122. Zolotukhin, S., and Vandenberghe, L.H. (2022). AAV capsid design: A Goldilocks challenge. Trends Mol Med *28*, 183-193. 10.1016/j.molmed.2022.01.003.
123. Au, H.K.E., Isalan, M., and Mielcarek, M. (2021). Gene Therapy Advances: A Meta-Analysis of AAV Usage in Clinical Settings. Front Med *8*, 809118. 10.3389/fmed.2021.809118.
124. Becker, J., Fakhiri, J., and Grimm, D. (2022). Fantastic AAV Gene Therapy Vectors and How to Find Them - Random Diversification, Rational Design and Machine Learning. Pathogens *11*. 10.3390/pathogens11070756.
125. Grimm, D., and Zolotukhin, S. (2015). E Pluribus Unum: 50 Years of Research, Millions of Viruses, and One Goal—Tailored Acceleration of AAV Evolution. Mol Ther *23*, 1819-1831. 10.1038/mt.2015.173.
126. Lam, A.K., Frabutt, D., Li, L., and Xiao, W. (2021). Chemical Modifications of the Capsid for Redirecting and Improving the Efficacy of Adeno-Associated Virus Vectors. Hum Gene Ther *32*, 1433-1438. 10.1089/hum.2021.124.
127. Wu, Z., Asokan, A., and Samulski, R.J. (2006). Adeno-associated Virus Serotypes: Vector Toolkit for Human Gene Therapy. Mol Ther *14*, 316-327. 10.1016/j.ymthe.2006.05.009.
128. Jackson, C.B., Richard, A.S., Ojha, A., Conkright, K.A., Trimarchi, J.M., Bailey, C.C., Alpert, M.D., Kay, M.A., Farzan, M., and Choe, H. (2020). AAV vectors engineered to target insulin receptor greatly enhance intramuscular gene delivery. Mol Ther Methods Clin Dev *19*, 496-506. 10.1016/j.omtm.2020.11.004.
129. Eichhoff, A.M., Borner, K., Albrecht, B., Schafer, W., Baum, N., Haag, F., Korbelin, J., Trepel, M., Braren, I., Grimm, D., Adriouch, S., et al. (2019). Nanobody-Enhanced Targeting of AAV Gene Therapy Vectors. Mol Ther Methods Clin Dev *15*, 211-220. 10.1016/j.omtm.2019.09.003.
130. Michels, A., Frank, A.M., Gunther, D.M., Mataei, M., Borner, K., Grimm, D., Hartmann, J., and Buchholz, C.J. (2021). Lentiviral and adeno-associated vectors efficiently transduce mouse T lymphocytes when targeted to murine CD8. Mol Ther Methods Clin Dev *23*, 334-347. 10.1016/j.omtm.2021.09.014.
131. Perabo, L., Endell, J., King, S., Lux, K., Goldnau, D., Hallek, M., and Büning, H. (2006). Combinatorial engineering of a gene therapy vector: directed evolution of adeno-associated virus. J Gene Med *8*, 155-162. 10.1002/jgm.849.
132. Müller, O.J., Kaul, F., Weitzman, M.D., Pasqualini, R., Arap, W., Kleinschmidt, J.A., and Trepel, M. (2003). Random peptide libraries displayed on adeno-associated virus to select for targeted gene therapy vectors. Nat Biotechnol *21*, 1040. 10.1038/nbt856.
133. Grimm, D., Lee, J.S., Wang, L., Desai, T., Akache, B., Storm, T.A., and Kay, M.A. (2008). In vitro and in vivo gene therapy vector evolution via multispecies interbreeding and retargeting of adeno-associated viruses. J Virol *82*, 5887-5911. 10.1128/jvi.00254-08.
134. Govindasamy, L., Padron, E., McKenna, R., Muzyczka, N., Kaludov, N., Chiorini John, A., and Agbandje-McKenna, M. (2006). Structurally Mapping the Diverse Phenotype of Adeno-Associated Virus Serotype 4. J Virol *80*, 11556-11570. 10.1128/JVI.01536-06.
135. Goertsen, D., Flytzanis, N.C., Goeden, N., Chuapoco, M.R., Cummins, A., Chen, Y., Fan, Y., Zhang, Q., Sharma, J., Duan, Y., Wang, L., et al. (2022). AAV capsid variants with brain-wide transgene expression and

decreased liver targeting after intravenous delivery in mouse and marmoset. Nat Neurosci *25*, 106-115. 10.1038/s41593-021-00969-4.
136. Shen, X., Storm, T., and Kay, M.A. (2007). Characterization of the Relationship of AAV Capsid Domain Swapping to Liver Transduction Efficiency. Mol Ther *15*, 1955-1962. 10.1038/sj.mt.6300293.
137. Ojala, D.S., Sun, S., Santiago-Ortiz, J.L., Shapiro, M.G., Romero, P.A., and Schaffer, D.V. (2018). In Vivo Selection of a Computationally Designed SCHEMA AAV Library Yields a Novel Variant for Infection of Adult Neural Stem Cells in the SVZ. Mol Ther *26*, 304-319. 10.1016/j.ymthe.2017.09.006.
138. Marsic, D., Govindasamy, L., Currlin, S., Markusic, D.M., Tseng, Y.-S., Herzog, R.W., Agbandje-McKenna, M., and Zolotukhin, S. (2014). Vector Design Tour de Force: Integrating Combinatorial and Rational Approaches to Derive Novel Adeno-associated Virus Variants. Mol Ther *22*, 1900-1909. 10.1038/mt.2014.139.
139. Körbelin, J., and Trepel, M. (2017). How to Successfully Screen Random Adeno-Associated Virus Display Peptide Libraries In Vivo. Hum Gene Ther Methods *28*, 109-123. 10.1089/hgtb.2016.177.
140. Adachi, K., Enoki, T., Kawano, Y., Veraz, M., and Nakai, H. (2014). Drawing a high-resolution functional map of adeno-associated virus capsid by massively parallel sequencing. Nat Commun *5*, 3075. 10.1038/ncomms4075.
141. Pekrun, K., De Alencastro, G., Luo, Q.-J., Liu, J., Kim, Y., Nygaard, S., Galivo, F., Zhang, F., Song, R., Tiffany, M.R., Xu, J., et al. (2019). Using a barcoded AAV capsid library to select for clinically relevant gene therapy vectors. JCI Insight *4*. 10.1172/jci.insight.131610.
142. Weinmann, J., Weis, S., Sippel, J., Tulalamba, W., Remes, A., El Andari, J., Herrmann, A.K., Pham, Q.H., Borowski, C., Hille, S., Schönberger, T., et al. (2020). Identification of a myotropic AAV by massively parallel in vivo evaluation of barcoded capsid variants. Nat Commun *11*, 5432. 10.1038/s41467-020-19230-w.
143. Westhaus, A., Cabanes-Creus, M., Rybicki, A., Baltazar, G., Navarro, R.G., Zhu, E., Drouyer, M., Knight, M., Albu, R.F., Ng, B.H., Kalajdzic, P., et al. (2020). High-Throughput In Vitro, Ex Vivo, and In Vivo Screen of Adeno-Associated Virus Vectors Based on Physical and Functional Transduction. Hum Gene Ther *31*, 575-589. 10.1089/hum.2019.264.
144. Davidsson, M., Wang, G., Aldrin-Kirk, P., Cardoso, T., Nolbrant, S., Hartnor, M., Mudannayake, J., Parmar, M., and Björklund, T. (2019). A systematic capsid evolution approach performed in vivo for the design of AAV vectors with tailored properties and tropism. Proc Natl Acad Sci U S A *116*, 27053-27062. 10.1073/pnas.1910061116.
145. Deverman, B.E., Pravdo, P.L., Simpson, B.P., Kumar, S.R., Chan, K.Y., Banerjee, A., Wu, W.-L., Yang, B., Huber, N., Pasca, S.P., and Gradinaru, V. (2016). Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. Nat Biotechnol *34*, 204-209. 10.1038/nbt.3440.
146. Nonnenmacher, M., Wang, W., Child, M.A., Ren, X.-Q., Huang, C., Ren, A.Z., Tocci, J., Chen, Q., Bittner, K., Tyson, K., Pande, N., et al. (2021). Rapid evolution of blood-brain-barrier-penetrating AAV capsids by RNA-driven biopanning. Mol Ther Methods Clin Dev *20*, 366-378. 10.1016/j.omtm.2020.12.006.
147. Tabebordbar, M., Lagerborg, K.A., Stanton, A., King, E.M., Ye, S., Tellez, L., Krunnfusz, A., Tavakoli, S., Widrick, J.J., Messemer, K.A., Troiano, E.C., et al. (2021). Directed evolution of a family of AAV capsid variants enabling potent muscle-directed gene delivery across species. Cell *184*, 4919-4938.e4922. 10.1016/j.cell.2021.08.028.
148. Börner, K., Kienle, E., Huang, L.-Y., Weinmann, J., Sacher, A., Bayer, P., Stüllein, C., Fakhiri, J., Zimmermann, L., Westhaus, A., Beneke, J., et al. (2020). Pre-arrayed Pan-AAV Peptide Display Libraries for Rapid Single-Round Screening. Mol Ther *28*, 1016-1032. 10.1016/j.ymthe.2020.02.009.
149. Körbelin, J., Sieber, T., Michelfelder, S., Lunding, L., Spies, E., Hunger, A., Alawi, M., Rapti, K., Indenbirken, D., Müller, O.J., Pasqualini, R., et al. (2016). Pulmonary Targeting of Adeno-associated Viral Vectors by Next-generation Sequencing-guided Screening of Random Capsid Displayed Peptide Libraries. Mol Ther *24*, 1050-1061. 10.1038/mt.2016.62.
150. Zinn, E., Pacouret, S., Khaychuk, V., Turunen, Heikki T., Carvalho, Livia S., Andres-Mateos, E., Shah, S., Shelke, R., Maurer, Anna C., Plovie, E., Xiao, R., et al. (2015). In Silico Reconstruction of the Viral Evolutionary Lineage Yields a Potent Gene Therapy Vector. Cell Rep *12*, 1056-1068. 10.1016/j.celrep.2015.07.019.
151. Zinn, E., Unzu, C., Schmit, P.F., Turunen, H.T., Zabaleta, N., Sanmiguel, J., Fieldsend, A., Bhatt, U., Diop, C., Merkel, E., Gurrala, R., et al. (2022). Ancestral library identifies conserved reprogrammable liver motif on AAV capsid. Cell Rep Med *3*, 100803. 10.1016/j.xcrm.2022.100803.
152. Tse Longping, V., Klinc Kelli, A., Madigan Victoria, J., Castellanos Rivera Ruth, M., Wells Lindsey, F., Havlik, L.P., Smith, J.K., Agbandje-McKenna, M., and Asokan, A. (2017). Structure-guided evolution of antigenically distinct adeno-associated virus variants for immune evasion. Proc Natl Acad Sci U S A *114*, E4812-E4821. 10.1073/pnas.1704766114.
153. Paulk, N.K., Pekrun, K., Zhu, E., Nygaard, S., Li, B., Xu, J., Chu, K., Leborgne, C., Dane, A.P., Haft, A., Zhang, Y., et al. (2018). Bioengineered AAV Capsids with Combined High Human Liver Transduction In Vivo and Unique Humoral Seroreactivity. Mol Ther *26*, 289-303. 10.1016/j.ymthe.2017.09.021.
154. Szumska, J., and Grimm, D. (2023). Boosters for adeno-associated virus (AAV) vector (r)evolution. Cytotherapy *25*, 254-260. 10.1016/j.jcyt.2022.07.005.
155. Lisowski, L., Dane, A.P., Chu, K., Zhang, Y., Cunningham, S.C., Wilson, E.M., Nygaard, S., Grompe, M., Alexander, I.E., and Kay, M.A. (2014). Selection and evaluation of clinically relevant AAV variants in a xenograft liver model. Nature *506*, 382-386. 10.1038/nature12875.
156. Gonzalez, T.J., Simon, K.E., Blondel, L.O., Fanous, M.M., Roger, A.L., Maysonet, M.S., Devlin, G.W., Smith, T.J., Oh, D.K., Havlik, L.P., Castellanos Rivera, R.M., et al. (2022). Cross-species evolution of a highly potent AAV variant for therapeutic gene transfer and genome editing. Nat Commun *13*, 5947. 10.1038/s41467-022-33745-4.
157. Ronzitti, G., Bortolussi, G., van Dijk, R., Collaud, F., Charles, S., Leborgne, C., Vidal, P., Martin, S., Gjata, B., Sola, M.S., van Wittenberghe, L., et al. (2016). A translationally optimized AAV-UGT1A1 vector drives safe and long-lasting correction of Crigler-Najjar syndrome. Mol Ther Methods Clin Dev *3*, 16049. 10.1038/mtm.2016.49.

158. Wang, L., Wang, Z., Zhang, F., Zhu, R., Bi, J., Wu, J., Zhang, H., Wu, H., Kong, W., Yu, B., and Yu, X. (2016). Enhancing Transgene Expression from Recombinant AAV8 Vectors in Different Tissues Using Woodchuck Hepatitis Virus Post-Transcriptional Regulatory Element. Int J Med Sci *13*, 286-291. 10.7150/ijms.14152.
159. Choi, J.-H., Yu, N.-K., Baek, G.-C., Bakes, J., Seo, D., Nam, H.J., Baek, S.H., Lim, C.-S., Lee, Y.-S., and Kaang, B.-K. (2014). Optimization of AAV expression cassettes to improve packaging capacity and transgene expression in neurons. Mol Brain *7*, 17. 10.1186/1756-6606-7-17.
160. Chamberlain, K., Riyad, J.M., and Weber, T. (2016). Expressing Transgenes That Exceed the Packaging Capacity of Adeno-Associated Virus Capsids. Hum Gene Ther Methods *27*, 1-12. 10.1089/hgtb.2015.140.
161. Buck, T.M., and Wijnholds, J. (2020). Recombinant Adeno-Associated Viral Vectors (rAAV)-Vector Elements in Ocular Gene Therapy Clinical Trials and Transgene Expression and Bioactivity Assays. Int J Mol Sci *21*, 4197.
162. Pellissier, L.P., Hoek, R.M., Vos, R.M., Aartsen, W.M., Klimczak, R.R., Hoyng, S.A., Flannery, J.G., and Wijnholds, J. (2014). Specific tools for targeting and expression in Müller glial cells. Mol Ther Methods Clin Dev *1*, 14009. 10.1038/mtm.2014.9.
163. Debnath, S., Jaako, P., Siva, K., Rothe, M., Chen, J., Dahl, M., Gaspar, H.B., Flygare, J., Schambach, A., and Karlsson, S. (2017). Lentiviral Vectors with Cellular Promoters Correct Anemia and Lethal Bone Marrow Failure in a Mouse Model for Diamond-Blackfan Anemia. Mol Ther *25*, 1805-1814. 10.1016/j.ymthe.2017.04.002.
164. Lee, Y., Messing, A., Su, M., and Brenner, M. (2008). GFAP promoter elements required for region-specific and astrocyte-specific expression. Glia *56*, 481-493. 10.1002/glia.20622.
165. Duan, D. (2018). Systemic AAV Micro-dystrophin Gene Therapy for Duchenne Muscular Dystrophy. Mol Ther *26*, 2337-2356. 10.1016/j.ymthe.2018.07.011.
166. Harper, S.Q., Hauser, M.A., DelloRusso, C., Duan, D., Crawford, R.W., Phelps, S.F., Harper, H.A., Robinson, A.S., Engelhardt, J.F., Brooks, S.V., and Chamberlain, J.S. (2002). Modular flexibility of dystrophin: Implications for gene therapy of Duchenne muscular dystrophy. Nat Med *8*, 253-261. 10.1038/nm0302-253.
167. Casazza, J.P., Cale, E.M., Narpala, S., Yamshchikov, G.V., Coates, E.E., Hendel, C.S., Novik, L., Holman, L.A., Widge, A.T., Apte, P., Gordon, I., et al. (2022). Safety and tolerability of AAV8 delivery of a broadly neutralizing antibody in adults living with HIV: a phase 1, dose-escalation trial. Nat Med *28*, 1022-1030. 10.1038/s41591-022-01762-x.
168. Liu, Z., Chen, O., Wall, J.B.J., Zheng, M., Zhou, Y., Wang, L., Vaseghi, H.R., Qian, L., and Liu, J. (2017). Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. Sci Rep *7*, 2193. 10.1038/s41598-017-02460-2.
169. Halbert, C.L., Allen, J.M., and Miller, A.D. (2002). Efficient mouse airway transduction following recombination between AAV vectors carrying parts of a larger gene. Nat Biotechnol *20*, 697-701. 10.1038/nbt0702-697.
170. Yan, Z., Zhang, Y., Duan, D., and Engelhardt, J.F. (2000). Trans-splicing vectors expand the utility of adeno-associated virus for gene therapy. Proc Natl Acad Sci U S A *97*, 6716-6721. 10.1073/pnas.97.12.6716.
171. Aranko, A.S., and Volkmann, G. (2011). Protein trans-splicing as a protein ligation tool to study protein structure and function. Biomol Concepts *2*, 183-198. 10.1515/bmc.2011.014.
172. Tornabene, P., Trapani, I., Minopoli, R., Centrulo, M., Lupo, M., de Simone, S., Tiberi, P., Dell'Aquila, F., Marrocco, E., Iodice, C., Iuliano, A., et al. (2019). Intein-mediated protein trans-splicing expands adeno-associated virus transfer capacity in the retina. Sci Transl Med *11*. 10.1126/scitranslmed.aav4523.
173. Schmelas, C., and Grimm, D. (2018). Split Cas9, Not Hairs - Advancing the Therapeutic Index of CRISPR Technology. Biotechnol J *13*, e1700432. 10.1002/biot.201700432.
174. Domenger, C., and Grimm, D. (2019). Next-generation AAV vectors-do not judge a virus (only) by its cover. Hum Mol Genet *28*, R3-r14. 10.1093/hmg/ddz148.
175. Isomura, H., Tsurumi, T., and Stinski, M.F. (2004). Role of the proximal enhancer of the major immediate-early promoter in human cytomegalovirus replication. J Virol *78*, 12788-12799. 10.1128/jvi.78.23.12788-12799.2004.
176. Gray, S.J., Foti, S.B., Schwartz, J.W., Bachaboina, L., Taylor-Blake, B., Coleman, J., Ehlers, M.D., Zylka, M.J., McCown, T.J., and Samulski, R.J. (2011). Optimizing promoters for recombinant adeno-associated virus-mediated gene expression in the peripheral and central nervous system using self-complementary vectors. Hum Gene Ther *22*, 1143-1153. 10.1089/hum.2010.245.
177. Johari, Y.B., Scarrott, J.M., Pohle, T.H., Liu, P., Mayer, A., Brown, A.J., and James, D.C. (2022). Engineering of the CMV promoter for controlled expression of recombinant genes in HEK293 cells. Biotechnol J *17*, e2200062. 10.1002/biot.202200062.
178. Niwa, H., Yamamura, K., and Miyazaki, J. (1991). Efficient selection for high-expression transfectants with a novel eukaryotic vector. Gene *108*, 193-199. 10.1016/0378-1119(91)90434-d.
179. Klein, R.L., Hamby, M.E., Gong, Y., Hirko, A.C., Wang, S., Hughes, J.A., King, M.A., and Meyer, E.M. (2002). Dose and Promoter Effects of Adeno-Associated Viral Vector for Green Fluorescent Protein Expression in the Rat Brain. Exp Neurol *176*, 66-74. 10.1006/exnr.2002.7942.
180. Prösch, S., Stein, J., Staak, K., Liebenthal, C., Volk, H.D., and Krüger, D.H. (1996). Inactivation of the very strong HCMV immediate early promoter by DNA CpG methylation in vitro. Biol Chem Hoppe Seyler *377*, 195-201. 10.1515/bchm3.1996.377.3.195.
181. Hsu, C.C., Li, H.P., Hung, Y.H., Leu, Y.W., Wu, W.H., Wang, F.S., Lee, K.D., Chang, P.J., Wu, C.S., Lu, Y.J., Huang, T.H., et al. (2010). Targeted methylation of CMV and E1A viral promoters. Biochem Biophys Res Commun *402*, 228-234. 10.1016/j.bbrc.2010.09.131.
182. Brooks, A.R., Harkins, R.N., Wang, P., Qian, H.S., Liu, P., and Rubanyi, G.M. (2004). Transcriptional silencing is associated with extensive methylation of the CMV promoter following adenoviral gene delivery to muscle. J Gene Med *6*, 395-404. 10.1002/jgm.516.

183. Buss, N., Lanigan, L., Zeller, J., Cissell, D., Metea, M., Adams, E., Higgins, M., Kim, K.H., Budzynski, E., Yang, L., Liu, Y., et al. (2022). Characterization of AAV-mediated dorsal root ganglionopathy. Mol Ther Methods Clin Dev *24*, 342-354. 10.1016/j.omtm.2022.01.013.
184. Ronzitti, G., Gross, D.A., and Mingozzi, F. (2020). Human Immune Responses to Adeno-Associated Virus (AAV) Vectors. Front Immunol *11*, 670. 10.3389/fimmu.2020.00670.
185. Qiao, C., Yuan, Z., Li, J., He, B., Zheng, H., Mayer, C., Li, J., and Xiao, X. (2011). Liver-specific microRNA-122 target sequences incorporated in AAV vectors efficiently inhibits transgene expression in the liver. Gene Ther *18*, 403-410. 10.1038/gt.2010.157.
186. Schoch, S., Cibelli, G., and Thiel, G. (1996). Neuron-specific gene expression of synapsin I. Major role of a negative regulatory mechanism. J Biol Chem *271*, 3317-3323. 10.1074/jbc.271.6.3317.
187. Mendell, J.R., Sahenk, Z., Lehman, K., Nease, C., Lowes, L.P., Miller, N.F., Iammarino, M.A., Alfano, L.N., Nicholl, A., Al-Zaidy, S., Lewis, S., et al. (2020). Assessment of Systemic Delivery of rAAVrh74.MHCK7.micro-dystrophin in Children With Duchenne Muscular Dystrophy: A Nonrandomized Controlled Trial. JAMA Neurol *77*, 1122-1131. 10.1001/jamaneurol.2020.1484.
188. Li, X., Eastman, E.M., Schwartz, R.J., and Draghia-Akli, R. (1999). Synthetic muscle promoters: activities exceeding naturally occurring regulatory sequences. Nat Biotechnol *17*, 241-245. 10.1038/6981.
189. Nathwani, A.C., Gray, J.T., Ng, C.Y., Zhou, J., Spence, Y., Waddington, S.N., Tuddenham, E.G., Kemball-Cook, G., McIntosh, J., Boon-Spijker, M., Mertens, K., et al. (2006). Self-complementary adeno-associated virus vectors containing a novel liver-specific human factor IX expression cassette enable highly efficient transduction of murine and nonhuman primate liver. Blood *107*, 2653-2661. 10.1182/blood-2005-10-4035.
190. O'Brien, J., Hayder, H., Zayed, Y., and Peng, C. (2018). Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. Front Endocrinol *9*, 402. 10.3389/fendo.2018.00402.
191. Bandiera, S., Pfeffer, S., Baumert, T.F., and Zeisel, M.B. (2015). miR-122 – A key factor and therapeutic target in liver disease. J Hepatol *62*, 448-457. 10.1016/j.jhep.2014.10.004.
192. Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H., and Kim, V.N. (2004). MicroRNA genes are transcribed by RNA polymerase II. Embo J *23*, 4051-4060. 10.1038/sj.emboj.7600385.
193. Han, J., Lee, Y., Yeom, K.H., Kim, Y.K., Jin, H., and Kim, V.N. (2004). The Drosha-DGCR8 complex in primary microRNA processing. Genes Dev *18*, 3016-3027. 10.1101/gad.1262504.
194. Yi, R., Qin, Y., Macara, I.G., and Cullen, B.R. (2003). Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. Genes Dev *17*, 3011-3016. 10.1101/gad.1158803.
195. Fareh, M., Yeom, K.-H., Haagsma, A.C., Chauhan, S., Heo, I., and Joo, C. (2016). TRBP ensures efficient Dicer processing of precursor microRNA in RNA-crowded environments. Nat Commun *7*, 13694. 10.1038/ncomms13694.
196. Wang, H.-W., Noland, C., Siridechadilok, B., Taylor, D.W., Ma, E., Felderer, K., Doudna, J.A., and Nogales, E. (2009). Structural insights into RNA processing by the human RISC-loading complex. Nat Struct Mol Biol *16*, 1148-1153. 10.1038/nsmb.1673.
197. Ender, C., and Meister, G. (2010). Argonaute proteins at a glance. J Cell Sci *123*, 1819-1823. 10.1242/jcs.055210.
198. Braun, J.E., Huntzinger, E., and Izaurralde, E. (2013). The role of GW182 proteins in miRNA-mediated gene silencing. Adv Exp Med Biol *768*, 147-163. 10.1007/978-1-4614-5107-5_9.
199. Liu, J., Carmell, M.A., Rivas, F.V., Marsden, C.G., Thomson, J.M., Song, J.J., Hammond, S.M., Joshua-Tor, L., and Hannon, G.J. (2004). Argonaute2 is the catalytic engine of mammalian RNAi. Science *305*, 1437-1441. 10.1126/science.1102513.
200. Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature *391*, 806-811. 10.1038/35888.
201. Agrawal, N., Dasaradhi, P.V., Mohmmed, A., Malhotra, P., Bhatnagar, R.K., and Mukherjee, S.K. (2003). RNA interference: biology, mechanism, and applications. Microbiol Mol Biol Rev *67*, 657-685. 10.1128/mmbr.67.4.657-685.2003.
202. Setten, R.L., Rossi, J.J., and Han, S.-p. (2019). The current state and future directions of RNAi-based therapeutics. Nat Rev Drug Discov *18*, 421-446. 10.1038/s41573-019-0017-4.
203. Borel, F., Kay, M.A., and Mueller, C. (2014). Recombinant AAV as a Platform for Translating the Therapeutic Potential of RNA Interference. Mol Ther *22*, 692-701. 10.1038/mt.2013.285.
204. Boudreau, R.L., and Davidson, B.L. (2012). Chapter fourteen - Generation of Hairpin-Based RNAi Vectors for Biological and Therapeutic Application. In Methods Enzymol, T. Friedmann, ed. (Academic Press), pp. 275-296. 10.1016/B978-0-12-386509-0.00014-4.
205. Ma, H., Wu, Y., Dang, Y., Choi, J.G., Zhang, J., and Wu, H. (2014). Pol III Promoters to Express Small RNAs: Delineation of Transcription Initiation. Mol Ther Nucleic Acids *3*, e161. 10.1038/mtna.2014.12.
206. Goguen, R.P., Del Corpo, O., Malard, C.M.G., Daher, A., Alpuche-Lazcano, S.P., Chen, M.J., Scarborough, R.J., and Gatignol, A. (2021). Efficacy, accumulation, and transcriptional profile of anti-HIV shRNAs expressed from human U6, 7SK, and H1 promoters. Mol Ther Nucleic Acids *23*, 1020-1034. 10.1016/j.omtn.2020.12.022.
207. Grimm, D., Streetz, K.L., Jopling, C.L., Storm, T.A., Pandey, K., Davis, C.R., Marion, P., Salazar, F., and Kay, M.A. (2006). Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. Nature *441*, 537-541. 10.1038/nature04791.
208. Herrera-Carrillo, E., and Berkhout, B. (2017). Dicer-independent processing of small RNA duplexes: mechanistic insights and applications. Nucleic Acids Res *45*, 10369-10379. 10.1093/nar/gkx779.
209. Boudreau, R.L., Martins, I., and Davidson, B.L. (2009). Artificial microRNAs as siRNA shuttles: improved safety as compared to shRNAs in vitro and in vivo. Mol Ther *17*, 169-175. 10.1038/mt.2008.231.
210. Konermann, S., Lotfy, P., Brideau, N.J., Oki, J., Shokhirev, M.N., and Hsu, P.D. (2018). Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. Cell *173*, 665-676.e614. 10.1016/j.cell.2018.02.033.

211. Brouns, S.J., Jore, M.M., Lundgren, M., Westra, E.R., Slijkhuis, R.J., Snijders, A.P., Dickman, M.J., Makarova, K.S., Koonin, E.V., and van der Oost, J. (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. Science *321*, 960-964. 10.1126/science.1159689.
212. Mohanraju, P., Makarova, K.S., Zetsche, B., Zhang, F., Koonin, E.V., and van der Oost, J. (2016). Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. Science *353*, aad5147. 10.1126/science.aad5147.
213. Amitai, G., and Sorek, R. (2016). CRISPR–Cas adaptation: insights into the mechanism of action. Nat Rev Microbiol *14*, 67-76. 10.1038/nrmicro.2015.14.
214. Li, H. (2015). Structural Principles of CRISPR RNA Processing. Structure *23*, 13-20. 10.1016/j.str.2014.10.006.
215. Makarova, K.S., Wolf, Y.I., Iranzo, J., Shmakov, S.A., Alkhnbashi, O.S., Brouns, S.J.J., Charpentier, E., Cheng, D., Haft, D.H., Horvath, P., Moineau, S., et al. (2020). Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. Nat Rev Microbiol *18*, 67-83. 10.1038/s41579-019-0299-x.
216. Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J.A., and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science *337*, 816-821. 10.1126/science.1225829.
217. Gleditzsch, D., Pausch, P., Müller-Esparza, H., Özcan, A., Guo, X., Bange, G., and Randau, L. (2019). PAM identification by CRISPR-Cas effector complexes: diversified mechanisms and structures. RNA Biol *16*, 504-517. 10.1080/15476286.2018.1504546.
218. Sander, J.D., and Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotechnol *32*, 347-355. 10.1038/nbt.2842.
219. Senís, E., Fatouros, C., Große, S., Wiedtke, E., Niopek, D., Mueller, A.K., Börner, K., and Grimm, D. (2014). CRISPR/Cas9-mediated genome engineering: an adeno-associated viral (AAV) vector toolbox. Biotechnol J *9*, 1402-1412. 10.1002/biot.201400046.
220. Maeder, M.L., Stefanidakis, M., Wilson, C.J., Baral, R., Barrera, L.A., Bounoutas, G.S., Bumcrot, D., Chao, H., Ciulla, D.M., DaSilva, J.A., Dass, A., et al. (2019). Development of a gene-editing approach to restore vision loss in Leber congenital amaurosis type 10. Nat Med *25*, 229-233. 10.1038/s41591-018-0327-9.
221. Anzalone, A.V., Koblan, L.W., and Liu, D.R. (2020). Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. Nat Biotechnol *38*, 824-844. 10.1038/s41587-020-0561-9.
222. Dominguez, A.A., Lim, W.A., and Qi, L.S. (2016). Beyond editing: repurposing CRISPR–Cas9 for precision genome regulation and interrogation. Nat Rev Mol Cell Biol *17*, 5-15. 10.1038/nrm.2015.2.
223. Cullot, G., Boutin, J., Toutain, J., Prat, F., Pennamen, P., Rooryck, C., Teichmann, M., Rousseau, E., Lamrissi-Garcia, I., Guyonnet-Duperat, V., Bibeyran, A., et al. (2019). CRISPR-Cas9 genome editing induces megabase-scale chromosomal truncations. Nat Commun *10*, 1136. 10.1038/s41467-019-09006-2.
224. Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173-1183. 10.1016/j.cell.2013.02.022.
225. Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., Lim, W.A., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell *154*, 442-451. 10.1016/j.cell.2013.06.044.
226. Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., and Joung, J.K. (2013). CRISPR RNA–guided activation of endogenous human genes. Nat Methods *10*, 977-979. 10.1038/nmeth.2598.
227. Black, Joshua B., Adler, Andrew F., Wang, H.-G., D'Ippolito, Anthony M., Hutchinson, Hunter A., Reddy, Timothy E., Pitt, Geoffrey S., Leong, Kam W., and Gersbach, Charles A. (2016). Targeted Epigenetic Remodeling of Endogenous Loci by CRISPR/Cas9-Based Transcriptional Activators Directly Converts Fibroblasts to Neuronal Cells. Cell Stem Cell *19*, 406-414. 10.1016/j.stem.2016.07.001.
228. Thakore, P.I., Kwon, J.B., Nelson, C.E., Rouse, D.C., Gemberling, M.P., Oliver, M.L., and Gersbach, C.A. (2018). RNA-guided transcriptional silencing in vivo with S. aureus CRISPR-Cas9 repressors. Nat Commun *9*, 1674. 10.1038/s41467-018-04048-4.
229. Moreno, A.M., Fu, X., Zhu, J., Katrekar, D., Shih, Y.-R.V., Marlett, J., Cabotaje, J., Tat, J., Naughton, J., Lisowski, L., Varghese, S., et al. (2018). In Situ Gene Therapy via AAV-CRISPR-Cas9-Mediated Targeted Gene Regulation. Mol Ther *26*, 1818-1827. 10.1016/j.ymthe.2018.04.017.
230. Chen, Y., Zhi, S., Liu, W., Wen, J., Hu, S., Cao, T., Sun, H., Li, Y., Huang, L., Liu, Y., Liang, P., et al. (2020). Development of Highly Efficient Dual-AAV Split Adenosine Base Editor for In Vivo Gene Therapy. Small Methods *4*, 2000309. 10.1002/smtd.202000309.
231. Chew, W.L., Tabebordbar, M., Cheng, J.K., Mali, P., Wu, E.Y., Ng, A.H., Zhu, K., Wagers, A.J., and Church, G.M. (2016). A multifunctional AAV-CRISPR-Cas9 and its host response. Nat Methods *13*, 868-874. 10.1038/nmeth.3993.
232. Zhang, C., Konermann, S., Brideau, N.J., Lotfy, P., Wu, X., Novick, S.J., Strutzenberg, T., Griffin, P.R., Hsu, P.D., and Lyumkis, D. (2018). Structural Basis for the RNA-Guided Ribonuclease Activity of CRISPR-Cas13d. Cell *175*, 212-223.e217. 10.1016/j.cell.2018.09.001.
233. Keng, C.T., Yogarajah, T., Lee, R.C.H., Muhammad, I.B.H., Chia, B.S., Vasandani, S.R., Lim, D.S., Guo, K., Wong, Y.H., Mok, C.K., Chu, J.J.H., et al. (2023). AAV-CRISPR-Cas13 eliminates human enterovirus and prevents death of infected mice. EBioMedicine *93*, 104682. 10.1016/j.ebiom.2023.104682.
234. Morelli, K.H., Wu, Q., Gosztyla, M.L., Liu, H., Yao, M., Zhang, C., Chen, J., Marina, R.J., Lee, K., Jones, K.L., Huang, M.Y., et al. (2023). An RNA-targeting CRISPR–Cas13d system alleviates disease-related phenotypes in Huntington's disease models. Nat Neurosci *26*, 27-38. 10.1038/s41593-022-01207-1.

235. He, B., Peng, W., Huang, J., Zhang, H., Zhou, Y., Yang, X., Liu, J., Li, Z., Xu, C., Xue, M., Yang, H., et al. (2020). Modulation of metabolic functions through Cas13d-mediated gene knockdown in liver. Protein Cell *11*, 518-524. 10.1007/s13238-020-00700-2.
236. Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., Zhao, X., Huang, B., Shi, W., Lu, R., Niu, P., et al. (2020). A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med *382*, 727-733. 10.1056/NEJMoa2001017.
237. Siddiqui, S., Alhamdi, H.W.S., and Alghamdi, H.A. (2022). Recent Chronology of COVID-19 Pandemic. Front Public Health *10*, 778037. 10.3389/fpubh.2022.778037.
238. Aslan, A., Aslan, C., Zolbanin, N.M., and Jafari, R. (2021). Acute respiratory distress syndrome in COVID-19: possible mechanisms and therapeutic management. Pneumonia *13*, 14. 10.1186/s41479-021-00092-9.
239. Bowe, B., Xie, Y., and Al-Aly, Z. (2023). Postacute sequelae of COVID-19 at 2 years. Nat Med. 10.1038/s41591-023-02521-2.
240. Markov, P.V., Ghafari, M., Beer, M., Lythgoe, K., Simmonds, P., Stilianakis, N.I., and Katzourakis, A. (2023). The evolution of SARS-CoV-2. Nat Rev Microbiol *21*, 361-379. 10.1038/s41579-023-00878-2.
241. Kumari, M., Lu, R.-M., Li, M.-C., Huang, J.-L., Hsu, F.-F., Ko, S.-H., Ke, F.-Y., Su, S.-C., Liang, K.-H., Yuan, J.P.-Y., Chiang, H.-L., et al. (2022). A critical overview of current progress for COVID-19: development of vaccines, antiviral drugs, and therapeutic antibodies. J Biomed Sci *29*, 68. 10.1186/s12929-022-00852-9.
242. Gorbalenya, A.E., Baker, S.C., Baric, R.S., de Groot, R.J., Drosten, C., Gulyaeva, A.A., Haagmans, B.L., Lauber, C., Leontovich, A.M., Neuman, B.W., Penzar, D., et al. (2020). The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. Nat Microbiol *5*, 536-544. 10.1038/s41564-020-0695-z.
243. Wang, C.C., Prather, K.A., Sznitman, J., Jimenez, J.L., Lakdawala, S.S., Tufekci, Z., and Marr, L.C. (2021). Airborne transmission of respiratory viruses. Science *373*. 10.1126/science.abd9149.
244. Lamers, M.M., and Haagmans, B.L. (2022). SARS-CoV-2 pathogenesis. Nat Rev Microbiol *20*, 270-284. 10.1038/s41579-022-00713-0.
245. Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., and Zhou, Q. (2020). Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science *367*, 1444-1448. 10.1126/science.abb2762.
246. Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.H., Nitsche, A., Müller, M.A., et al. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell *181*, 271-280.e278. 10.1016/j.cell.2020.02.052.
247. Yan, W., Zheng, Y., Zeng, X., He, B., and Cheng, W. (2022). Structural biology of SARS-CoV-2: open the door for novel therapies. Signal Transduct Target Ther *7*, 26. 10.1038/s41392-022-00884-5.
248. Taylor, P.C., Adams, A.C., Hufford, M.M., de la Torre, I., Winthrop, K., and Gottlieb, R.L. (2021). Neutralizing monoclonal antibodies for treatment of COVID-19. Nat Rev Immunol *21*, 382-393. 10.1038/s41577-021-00542-x.
249. Lei, S., Chen, X., Wu, J., Duan, X., and Men, K. (2022). Small molecules in the treatment of COVID-19. Signal Transduct Target Ther *7*, 387. 10.1038/s41392-022-01249-8.
250. Ambike, S., Cheng, C.-C., Feuerherd, M., Velkov, S., Baldassi, D., Afridi, S.Q., Porras-Gonzalez, D., Wei, X., Hagen, P., Kneidinger, N., Stoleriu, Mircea G., et al. (2022). Targeting genomic SARS-CoV-2 RNA with siRNAs allows efficient inhibition of viral replication and spread. Nucleic Acids Res *50*, 333-349. 10.1093/nar/gkab1248.
251. Abbott, T.R., Dhamdhere, G., Liu, Y., Lin, X., Goudy, L., Zeng, L., Chemparathy, A., Chmura, S., Heaton, N.S., Debs, R., Pande, T., et al. (2020). Development of CRISPR as an Antiviral Strategy to Combat SARS-CoV-2 and Influenza. Cell *181*, 865-876.e812. 10.1016/j.cell.2020.04.020.
252. Haniff, H.S., Tong, Y., Liu, X., Chen, J.L., Suresh, B.M., Andrews, R.J., Peterson, J.M., O'Leary, C.A., Benhamou, R.I., Moss, W.N., and Disney, M.D. (2020). Targeting the SARS-CoV-2 RNA Genome with Small Molecule Binders and Ribonuclease Targeting Chimera (RIBOTAC) Degraders. ACS Cent Sci *6*, 1713-1721. 10.1021/acscentsci.0c00984.
253. Rinella, M.E., Lazarus, J.V., Ratziu, V., Francque, S.M., Sanyal, A.J., Kanwal, F., Romero, D., Abdelmalek, M.F., Anstee, Q.M., Arab, J.P., Arrese, M., et al. (2023). A multi-society Delphi consensus statement on new fatty liver disease nomenclature. Hepatology.
254. Benedict, M., and Zhang, X. (2017). Non-alcoholic fatty liver disease: An expanded review. World J Hepatol *9*, 715-732. 10.4254/wjh.v9.i16.715.
255. Sayiner, M., Koenig, A., Henry, L., and Younossi, Z.M. (2016). Epidemiology of Nonalcoholic Fatty Liver Disease and Nonalcoholic Steatohepatitis in the United States and the Rest of the World. Clin Liver Dis *20*, 205-214. 10.1016/j.cld.2015.10.001.
256. Younossi, Z.M., Koenig, A.B., Abdelatif, D., Fazel, Y., Henry, L., and Wymer, M. (2016). Global epidemiology of nonalcoholic fatty liver disease—Meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology *64*, 73-84. 10.1002/hep.28431.
257. Estes, C., Anstee, Q.M., Arias-Loste, M.T., Bantel, H., Bellentani, S., Caballeria, J., Colombo, M., Craxi, A., Crespo, J., Day, C.P., Eguchi, Y., et al. (2018). Modeling NAFLD disease burden in China, France, Germany, Italy, Japan, Spain, United Kingdom, and United States for the period 2016–2030. J Hepatol *69*, 896-904. 10.1016/j.jhep.2018.05.036.
258. Targher, G., Byrne, C.D., Lonardo, A., Zoppini, G., and Barbui, C. (2016). Non-alcoholic fatty liver disease and risk of incident cardiovascular disease: A meta-analysis. J Hepatol *65*, 589-600. 10.1016/j.jhep.2016.05.013.
259. Pouwels, S., Sakran, N., Graham, Y., Leal, A., Pintar, T., Yang, W., Kassir, R., Singhal, R., Mahawar, K., and Ramnarain, D. (2022). Non-alcoholic fatty liver disease (NAFLD): a review of pathophysiology, clinical management and effects of weight loss. BMC Endocr Disord *22*, 63. 10.1186/s12902-022-00980-1.
260. Patouraux, S., Rousseau, D., Bonnafous, S., Lebeaupin, C., Luci, C., Canivet, C.M., Schneck, A.S., Bertola, A., Saint-Paul, M.C., Iannelli, A., Gugenheim, J., et al. (2017). CD44 is a key player in non-alcoholic steatohepatitis. J Hepatol *67*, 328-338. 10.1016/j.jhep.2017.03.003.

261. Fang, M., Yao, M., Yang, J., Zheng, W.J., Wang, L., and Yao, D.F. (2020). Abnormal CD44 activation of hepatocytes with nonalcoholic fatty accumulation in rat hepatocarcinogenesis. World J Gastrointest Oncol *12*, 66-76. 10.4251/wjgo.v12.i1.66.
262. Kodama, K., Toda, K., Morinaga, S., Yamada, S., and Butte, A.J. (2014). Anti-CD44 Antibody Treatment Lowers Hyperglycemia and Improves Insulin Resistance, Adipose Inflammation, and Hepatic Steatosis in Diet-Induced Obese Mice. Diabetes *64*, 867-875. 10.2337/db14-0149.
263. Gibert-Ramos, A., Sanfeliu-Redondo, D., Aristu-Zabalza, P., Martínez-Alcocer, A., Gracia-Sancho, J., Guixé-Muntet, S., and Fernández-Iglesias, A. (2021). The Hepatic Sinusoid in Chronic Liver Disease: The Optimal Milieu for Cancer. Cancers.
264. Knolle, P.A., and Wohlleber, D. (2016). Immunological functions of liver sinusoidal endothelial cells. Cell Mol Immunol *13*, 347-353. 10.1038/cmi.2016.5.
265. Lee, U.E., and Friedman, S.L. (2011). Mechanisms of hepatic fibrogenesis. Best Pract Res Clin Gastroenterol *25*, 195-206. 10.1016/j.bpg.2011.02.005.
266. Si-Tayeb, K., Lemaigre, F.P., and Duncan, S.A. (2010). Organogenesis and Development of the Liver. Dev Cell *18*, 175-189. 10.1016/j.devcel.2010.01.011.
267. Banales, J.M., Huebert, R.C., Karlsen, T., Strazzabosco, M., LaRusso, N.F., and Gores, G.J. (2019). Cholangiocyte pathobiology. Nat Rev Gastroenterol Hepatol *16*, 269-281. 10.1038/s41575-019-0125-y.
268. Davis, M.W., and Jorgensen, E.M. (2022). ApE, A Plasmid Editor: A Freely Available DNA Manipulation and Visualization Program. Front Bioinform *2*, 818619. 10.3389/fbinf.2022.818619.
269. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat Methods *9*, 671-675. 10.1038/nmeth.2089.
270. Wistuba, A., Weger, S., Kern, A., and Kleinschmidt, J.A. (1995). Intermediates of adeno-associated virus type 2 assembly: identification of soluble complexes containing Rep and Cap proteins. J Virol *69*, 5311-5319. 10.1128/jvi.69.9.5311-5319.1995.
271. Engler, C., Kandzia, R., and Marillonnet, S. (2008). A One Pot, One Step, Precision Cloning Method with High Throughput Capability. PLoS ONE *3*, e3647. 10.1371/journal.pone.0003647.
272. Wessels, H.-H., Méndez-Mancilla, A., Guo, X., Legut, M., Daniloski, Z., and Sanjana, N.E. (2020). Massively parallel Cas13 screens reveal principles for guide RNA design. Nat Biotechnol *38*, 722-727. 10.1038/s41587-020-0456-9.
273. Guo, X., Rahman, J.A., Wessels, H.-H., Méndez-Mancilla, A., Haro, D., Chen, X., and Sanjana, N.E. (2021). Transcriptome-wide Cas13 guide RNA design for model organisms and viral RNA pathogens. Cell Genom *1*, 100001. 10.1016/j.xgen.2021.100001.
274. Amoasii, L., Long, C., Li, H., Mireault, A.A., Shelton, J.M., Sanchez-Ortiz, E., McAnally, J.R., Bhattacharyya, S., Schmidt, F., Grimm, D., Hauschka, S.D., et al. (2017). Single-cut genome editing restores dystrophin expression in a new mouse model of muscular dystrophy. Sci Transl Med *9*, eaan8081. 10.1126/scitranslmed.aan8081.
275. Becker, J. (2019). DEPOOL - Depletion of Off-Targeting AAVs from Organ Libraries. Faculty of Biosciences, Molecular Biotechnology. University of Heidelberg.
276. Xia, H., Mao, Q., Paulson, H.L., and Davidson, B.L. (2002). siRNA-mediated gene silencing in vitro and in vivo. Nat Biotechnol *20*, 1006-1010. 10.1038/nbt739.
277. Mederacke, I., Hsu, C.C., Troeger, J.S., Huebener, P., Mu, X., Dapito, D.H., Pradere, J.P., and Schwabe, R.F. (2013). Fate tracing reveals hepatic stellate cells as dominant contributors to liver fibrosis independent of its aetiology. Nat Commun *4*, 2823. 10.1038/ncomms3823.
278. Mederacke, I., Dapito, D.H., Affò, S., Uchinami, H., and Schwabe, R.F. (2015). High-yield and high-purity isolation of hepatic stellate cells from normal and fibrotic mouse livers. Nat Protoc *10*, 305-315. 10.1038/nprot.2015.017.
279. Wilson, E.M., Bial, J., Tarlow, B., Bial, G., Jensen, B., Greiner, D.L., Brehm, M.A., and Grompe, M. (2014). Extensive double humanization of both liver and hematopoiesis in FRGN mice. Stem Cell Res *13*, 404-412. 10.1016/j.scr.2014.08.006.
280. Hardigan, A.A., Roberts, B.S., Moore, D.E., Ramaker, R.C., Jones, A.L., and Myers, R.M. (2019). CRISPR/Cas9-targeted removal of unwanted sequences from small-RNA sequencing libraries. Nucleic Acids Res *47*, e84. 10.1093/nar/gkz425.
281. Rapti, K., Maiakovska, O., Becker, J., Szumska, J., Zayas, M., Bubeck, F., Liu, J., Gerstmann, E., Krämer, C., Wiedtke, E., and Grimm, D. (2022). Isolation of Next-Generation Gene Therapy Vectors through Engineering, Barcoding, and Screening of Adeno-Associated Virus (AAV) Capsid Variants. J Vis Exp. 10.3791/64389.
282. Smith, R.P., Taher, L., Patwardhan, R.P., Kim, M.J., Inoue, F., Shendure, J., Ovcharenko, I., and Ahituv, N. (2013). Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. Nat Genet *45*, 1021-1028. 10.1038/ng.2713.
283. Nair, N., Rincon, M.Y., Evens, H., Sarcar, S., Dastidar, S., Samara-Kuko, E., Ghandeharian, O., Man Viecelli, H., Thöny, B., De Bleser, P., VandenDriessche, T., et al. (2014). Computationally designed liver-specific transcriptional modules and hyperactive factor IX improve hepatic gene therapy. Blood *123*, 3195-3199. 10.1182/blood-2013-10-534032.
284. Nagaki, M., and Moriwaki, H. (2008). Transcription factor HNF and hepatocyte differentiation. Hepatol Res *38*, 961-969. 10.1111/j.1872-034X.2008.00367.x.
285. Zhou, Q., Liu, M., Xia, X., Gong, T., Feng, J., Liu, W., Liu, Y., Zhen, B., Wang, Y., Ding, C., and Qin, J. (2017). A mouse tissue transcription factor atlas. Nat Commun *8*, 15089. 10.1038/ncomms15089.
286. Kaestner, K.H. (2000). The hepatocyte nuclear factor 3 (HNF3 or FOXA) family in metabolism. Trends Endocrinol Metab *11*, 281-285. 10.1016/s1043-2760(00)00271-x.
287. Liu, G., Wang, L., Wess, J., and Dean, A. (2022). Enhancer looping protein LDB1 regulates hepatocyte gene expression by cooperating with liver transcription factors. Nucleic Acids Res *50*, 9195-9211. 10.1093/nar/gkac707.

288. Castro-Mondragon, J.A., Riudavets-Puig, R., Rauluseviciute, I., Lemma, R.B., Turchi, L., Blanc-Mathieu, R., Lucas, J., Boddie, P., Khan, A., Manosalva Pérez, N., Fornes, O., et al. (2022). JASPAR 2022: the 9th release of the open-access database of transcription factor binding profiles. Nucleic Acids Res *50*, D165-d173. 10.1093/nar/gkab1113.
289. Fukuhara, T., Kambara, H., Shiokawa, M., Ono, C., Katoh, H., Morita, E., Okuzaki, D., Maehara, Y., Koike, K., and Matsuura, Y. (2012). Expression of microRNA miR-122 facilitates an efficient replication in nonhepatic cells upon infection with hepatitis C virus. J Virol *86*, 7918-7933. 10.1128/jvi.00567-12.
290. Becker, J., Stanifer, M.L., Leist, S.R., Stolp, B., Maiakovska, O., West, A., Wiedtke, E., Börner, K., Ghanem, A., Ambiel, I., Tse, L.V., et al. (2022). Ex vivo and in vivo suppression of SARS-CoV-2 with combinatorial AAV/RNAi expression vectors. Mol Ther *30*, 2005-2023. 10.1016/j.ymthe.2022.01.024.
291. Xie, J., Mao, Q., Tai, P.W.L., He, R., Ai, J., Su, Q., Zhu, Y., Ma, H., Li, J., Gong, S., Wang, D., et al. (2017). Short DNA Hairpins Compromise Recombinant Adeno-Associated Virus Genome Homogeneity. Mol Ther *25*, 1363-1374. 10.1016/j.ymthe.2017.03.028.
292. Brenner, M., Kisseberth, W.C., Su, Y., Besnard, F., and Messing, A. (1994). GFAP promoter directs astrocyte-specific expression in transgenic mice. J Neurosci *14*, 1030-1037. 10.1523/jneurosci.14-03-01030.1994.
293. Nathwani, A.C., Rosales, C., McIntosh, J., Rastegarlari, G., Nathwani, D., Raj, D., Nawathe, S., Waddington, S.N., Bronson, R., Jackson, S., Donahue, R.E., et al. (2011). Long-term safety and efficacy following systemic administration of a self-complementary AAV vector encoding human FIX pseudotyped with serotype 5 and 8 capsid proteins. Mol Ther *19*, 876-885. 10.1038/mt.2010.274.
294. Zincarelli, C., Soltys, S., Rengo, G., and Rabinowitz, J.E. (2008). Analysis of AAV Serotypes 1–9 Mediated Gene Expression and Tropism in Mice After Systemic Injection. Mol Ther *16*, 1073-1080. 10.1038/mt.2008.76.
295. Joo, M.S., Koo, J.H., Kim, T.H., Kim, Y.S., and Kim, S.G. (2019). LRH1-driven transcription factor circuitry for hepatocyte identity: Super-enhancer cistromic analysis. EBioMedicine *40*, 488-503. 10.1016/j.ebiom.2018.12.056.
296. Nonnenmacher, M., van Bakel, H., Hajjar, R.J., and Weber, T. (2015). High capsid-genome correlation facilitates creation of AAV libraries for directed evolution. Mol Ther *23*, 675-682. 10.1038/mt.2015.3.
297. Weinmann, J. (2018). Massively parallel in vivo characterization of novel adeno associated viral (AAV) capsids using DNA/RNA barcoding and next generation sequencing. Combined Faculties for the Natural Sciences and Mathematics. University of Heidelberg.
298. Guillou, J., de Pellegars, A., Porcheret, F., Frémeaux-Bacchi, V., Allain-Launay, E., Debord, C., Denis, M., Péréon, Y., Barnérias, C., Desguerre, I., Roussey, G., et al. (2022). Fatal thrombotic microangiopathy case following adeno-associated viral SMN gene therapy. Blood Adv *6*, 4266-4270. 10.1182/bloodadvances.2021006419.
299. Dalwadi, D.A., Torrens, L., Abril-Fornaguera, J., Pinyol, R., Willoughby, C., Posey, J., Llovet, J.M., Lanciault, C., Russell, D.W., Grompe, M., and Naugler, W.E. (2021). Liver Injury Increases the Incidence of HCC following AAV Gene Therapy in Mice. Mol Ther *29*, 680-690. 10.1016/j.ymthe.2020.10.018.
300. Wang, D., Zhang, F., and Gao, G. (2020). CRISPR-Based Therapeutic Genome Editing: Strategies and In Vivo Delivery by AAV Vectors. Cell *181*, 136-150. 10.1016/j.cell.2020.03.023.
301. Goel, K., and Ploski, J.E. (2022). RISC-y Business: Limitations of Short Hairpin RNA-Mediated Gene Silencing in the Brain and a Discussion of CRISPR/Cas-Based Alternatives. Front Mol Neurosci *15*. 10.3389/fnmol.2022.914430.
302. Zhang, C., Freistaedter, A., Schmelas, C., Gunkel, M., Dao Thi, V.L., and Grimm, D. (2022). An RNA Interference/Adeno-Associated Virus Vector-Based Combinatorial Gene Therapy Approach Against Hepatitis E Virus. Hepatol Commun *6*, 878-888. 10.1002/hep4.1842.
303. Toktay, Y., Dayanc, B., and Senturk, S. (2022). Engineering and validation of a dual luciferase reporter system for quantitative and systematic assessment of regulatory sequences in Chinese hamster ovary cells. Sci Rep *12*, 6050. 10.1038/s41598-022-09887-2.
304. Brown, B.D., Venneri, M.A., Zingale, A., Sergi Sergi, L., and Naldini, L. (2006). Endogenous microRNA regulation suppresses transgene expression in hematopoietic lineages and enables stable gene transfer. Nat Med *12*, 585-591. 10.1038/nm1398.
305. Xie, J., Xie, Q., Zhang, H., Ameres, S.L., Hung, J.H., Su, Q., He, R., Mu, X., Seher Ahmed, S., Park, S., Kato, H., et al. (2011). MicroRNA-regulated, systemically delivered rAAV9: a step closer to CNS-restricted transgene expression. Mol Ther *19*, 526-535. 10.1038/mt.2010.279.
306. Xiao, Y., Muhuri, M., Li, S., Qin, W., Xu, G., Luo, L., Li, J., Letizia, A.J., Wang, S.K., Chan, Y.K., Wang, C., et al. (2019). Circumventing cellular immunity by miR142-mediated regulation sufficiently supports rAAV-delivered OVA expression without activating humoral immunity. JCI Insight *5*. 10.1172/jci.insight.99052.
307. Gentner, B., Schira, G., Giustacchini, A., Amendola, M., Brown, B.D., Ponzoni, M., and Naldini, L. (2009). Stable knockdown of microRNA in vivo by lentiviral vectors. Nat Methods *6*, 63-66. 10.1038/nmeth.1277.
308. Zhang, B., Ye, Y., Ye, W., Perčulija, V., Jiang, H., Chen, Y., Li, Y., Chen, J., Lin, J., Wang, S., Chen, Q., et al. (2019). Two HEPN domains dictate CRISPR RNA maturation and target cleavage in Cas13d. Nat Commun *10*, 2544. 10.1038/s41467-019-10507-3.
309. Liao, C., Ttofali, F., Slotkowski, R.A., Denny, S.R., Cecil, T.D., Leenay, R.T., Keung, A.J., and Beisel, C.L. (2019). Modular one-pot assembly of CRISPR arrays enables library generation and reveals factors influencing crRNA biogenesis. Nat Commun *10*, 2948. 10.1038/s41467-019-10747-3.
310. Powell, J.E., Lim, C.K.W., Krishnan, R., McCallister, T.X., Saporito-Magriña, C., Zeballos, M.A., McPheron, G.D., and Gaj, T. (2022). Targeted gene silencing in the nervous system with CRISPR-Cas13. Sci Adv *8*, eabk2485. 10.1126/sciadv.abk2485.
311. Gupta, R., Ghosh, A., Chakravarti, R., Singh, R., Ravichandiran, V., Swarnakar, S., and Ghosh, D. (2022). Cas13d: A New Molecular Scissor for Transcriptome Engineering. Front Cell Dev Biol *10*, 866800. 10.3389/fcell.2022.866800.

312. Zhang, Y., Nguyen, T.M., Zhang, X.-O., Wang, L., Phan, T., Clohessy, J.G., and Pandolfi, P.P. (2021). Optimized RNA-targeting CRISPR/Cas13d technology outperforms shRNA in identifying functional circRNAs. Genome Biol *22*, 41. 10.1186/s13059-021-02263-9.
313. Meeske, A.J., Nakandakari-Higa, S., and Marraffini, L.A. (2019). Cas13-induced cellular dormancy prevents the rise of CRISPR-resistant bacteriophage. Nature *570*, 241-245. 10.1038/s41586-019-1257-5.
314. Ai, Y., Liang, D., and Wilusz, J.E. (2022). CRISPR/Cas13 effectors have differing extents of off-target effects that limit their utility in eukaryotic cells. Nucleic Acids Res *50*, e65. 10.1093/nar/gkac159.
315. Kelley, C.P., Haerle, M.C., and Wang, E.T. (2022). Negative autoregulation mitigates collateral RNase activity of repeat-targeting CRISPR-Cas13d in mammalian cells. Cell Rep *40*, 111226. 10.1016/j.celrep.2022.111226.
316. Shi, P., Murphy, M.R., Aparicio, A.O., Kesner, J.S., Fang, Z., Chen, Z., Trehan, A., Guo, Y., and Wu, X. (2023). Collateral activity of the CRISPR/RfxCas13d system in human cells. Commun Biol *6*, 334. 10.1038/s42003-023-04708-2.
317. Li, Y., Xu, J., Guo, X., Li, Z., Cao, L., Liu, S., Guo, Y., Wang, G., Luo, Y., Zhang, Z., Wei, X., et al. (2023). The collateral activity of RfxCas13d can induce lethality in a RfxCas13d knock-in mouse model. Genome Biol *24*, 20. 10.1186/s13059-023-02860-w.
318. Tong, H., Huang, J., Xiao, Q., He, B., Dong, X., Liu, Y., Yang, X., Han, D., Wang, Z., Wang, X., Ying, W., et al. (2023). High-fidelity Cas13 variants for targeted RNA degradation with minimal collateral effects. Nat Biotechnol *41*, 108-119. 10.1038/s41587-022-01419-7.
319. Wei, J., Lotfy, P., Faizi, K., Wang, E., Slabodkin, H., Kinnaman, E., Chandrasekaran, S., Kitano, H., Durrant, M.G., Duffy, C.V., Hsu, P.D., et al. (2022). Deep learning and CRISPR-Cas13d ortholog discovery for optimized RNA targeting. bioRxiv, 2021.2009.2014.460134. 10.1101/2021.09.14.460134.
320. Tang, X.-Z.E., Tan, S.X., Hoon, S., and Yeo, G.W. (2022). Pre-existing adaptive immunity to the RNA-editing enzyme Cas13d in humans. Nat Med *28*, 1372-1376. 10.1038/s41591-022-01848-6.
321. Mehta, A., and Merkel, O.M. (2020). Immunogenicity of Cas9 Protein. J Pharm Sci *109*, 62-67. 10.1016/j.xphs.2019.10.003.
322. Grimm, D., Pandey, K., and Kay, M.A. (2005). Adeno-associated virus vectors for short hairpin RNA expression. Methods Enzymol *392*, 381-405. 10.1016/s0076-6879(04)92023-x.
323. Tai, P.W.L., Xie, J., Fong, K., Seetin, M., Heiner, C., Su, Q., Weiand, M., Wilmot, D., Zapp, M.L., and Gao, G. (2018). Adeno-associated Virus Genome Population Sequencing Achieves Full Vector Genome Resolution and Reveals Human-Vector Chimeras. Mol Ther Methods Clin Dev *9*, 130-141. 10.1016/j.omtm.2018.02.002.
324. Xie, J., Tai, P.W.L., Brown, A., Gong, S., Zhu, S., Wang, Y., Li, C., Colpan, C., Su, Q., He, R., Ma, H., et al. (2020). Effective and Accurate Gene Silencing by a Recombinant AAV-Compatible MicroRNA Scaffold. Mol Ther *28*, 422-430. 10.1016/j.ymthe.2019.11.018.
325. Alsing, S., Doktor, T.K., Askou, A.L., Jensen, E.G., Ahmadov, U., Kristensen, L.S., Andresen, B.S., Aagaard, L., and Corydon, T.J. (2022). VEGFA-targeting miR-agshRNAs combine efficacy with specificity and safety for retinal gene therapy. Mol Ther Nucleic Acids *28*, 58-76. 10.1016/j.omtn.2022.02.019.
326. Borel, F., van Logtenstein, R., Koornneef, A., Maczuga, P., Ritsema, T., Petry, H., van Deventer, S.J., Jansen, P.L., and Konstantinova, P. (2011). In vivo knock-down of multidrug resistance transporters ABCC1 and ABCC2 by AAV-delivered shRNAs and by artificial miRNAs. J RNAi Gene Silencing *7*, 434-442.
327. Sun, C.P., Wu, T.H., Chen, C.C., Wu, P.Y., Shih, Y.M., Tsuneyama, K., and Tao, M.H. (2013). Studies of efficacy and liver toxicity related to adeno-associated virus-mediated RNA interference. Hum Gene Ther *24*, 739-750. 10.1089/hum.2012.239.
328. de Solis, C.A., Holehonnur, R., Banerjee, A., Luong, J.A., Lella, S.K., Ho, A., Pahlavan, B., and Ploski, J.E. (2015). Viral delivery of shRNA to amygdala neurons leads to neurotoxicity and deficits in Pavlovian fear conditioning. Neurobiol Learn Mem *124*, 34-47. 10.1016/j.nlm.2015.07.005.
329. Becker, J., Fischer, N., and Grimm, D. (2022). Right on target: The next class of efficient, safe, and specific RNAi triggers. Mol Ther Nucleic Acids *28*, 363-365. 10.1016/j.omtn.2022.03.027.
330. Liu, Y.P., Schopman, N.C., and Berkhout, B. (2013). Dicer-independent processing of short hairpin RNAs. Nucleic Acids Res *41*, 3723-3733. 10.1093/nar/gkt036.
331. Kaadt, E., Alsing, S., Cecchi, C.R., Damgaard, C.K., Corydon, T.J., and Aagaard, L. (2019). Efficient Knockdown and Lack of Passenger Strand Activity by Dicer-Independent shRNAs Expressed from Pol II-Driven MicroRNA Scaffolds. Mol Ther Nucleic Acids *14*, 318-328. 10.1016/j.omtn.2018.11.013.
332. Suhy, D.A., Kao, S.C., Mao, T., Whiteley, L., Denise, H., Souberbielle, B., Burdick, A.D., Hayes, K., Wright, J.F., Lavender, H., Roelvink, P., et al. (2012). Safe, long-term hepatic expression of anti-HCV shRNA in a nonhuman primate model. Mol Ther *20*, 1737-1749. 10.1038/mt.2012.119.
333. Patel, K., Kilfoil, G., Wyles, D.L., Naggie, S., Lawitz, E., Bradley, S., Lindell, P., and Suhy, D. (2016). 258. Phase I/IIa Study of TT-034, a DNA-Directed RNA Interference (ddRNAi) Agent Delivered as a Single Administration for the Treatment of Subjects with Chronic Hepatitis C Virus (HCV). Mol Ther *24*, S102. 10.1016/S1525-0016(16)33067-2.
334. Senbanjo, L.T., and Chellaiah, M.A. (2017). CD44: A Multifunctional Cell Surface Adhesion Receptor Is a Regulator of Progression and Metastasis of Cancer Cells. Front Cell Dev Biol *5*. 10.3389/fcell.2017.00018.
335. Xu, H., Niu, M., Yuan, X., Wu, K., and Liu, A. (2020). CD44 as a tumor biomarker and therapeutic target. Exp Hematol Oncol *9*, 36. 10.1186/s40164-020-00192-0.
336. Kang, H.S., Liao, G., DeGraff, L.M., Gerrish, K., Bortner, C.D., Garantziotis, S., and Jetten, A.M. (2013). CD44 Plays a Critical Role in Regulating Diet-Induced Adipose Inflammation, Hepatic Steatosis, and Insulin Resistance. PLoS ONE *8*, e58417. 10.1371/journal.pone.0058417.
337. Ma, L., Dong, L., and Chang, P. (2019). CD44v6 engages in colorectal cancer progression. Cell Death Dis *10*, 30. 10.1038/s41419-018-1265-7.

338. Jordan, A.R., Racine, R.R., Hennig, M.J.P., and Lokeshwar, V.B. (2015). The Role of CD44 in Disease Pathophysiology and Targeted Treatment. Front Immunol *6*. 10.3389/fimmu.2015.00182.
339. Machado, M.V., Michelotti, G.A., Xie, G., Almeida Pereira, T., Boursier, J., Bohnic, B., Guy, C.D., and Diehl, A.M. (2015). Mouse models of diet-induced nonalcoholic steatohepatitis reproduce the heterogeneity of the human disease. PLoS ONE *10*, e0127991. 10.1371/journal.pone.0127991.
340. Bu, F.-t., Jia, P.-c., Zhu, Y., Yang, Y.-r., Meng, H.-w., Bi, Y.-h., Huang, C., and Li, J. (2022). Emerging therapeutic potential of adeno-associated virus-mediated gene therapy in liver fibrosis. Mol Ther Methods Clin Dev *26*, 191-206. 10.1016/j.omtm.2022.06.009.
341. de Jong, Y.P., and Herzog, R.W. (2021). Liver gene therapy and hepatocellular carcinoma: A complex web. Mol Ther *29*, 1353-1354. 10.1016/j.ymthe.2021.03.009.
342. Dhamija, E., Paul, S.B., and Kedia, S. (2019). Non-alcoholic fatty liver disease associated with hepatocellular carcinoma: An increasing concern. Indian J Med Res *149*, 9-17. 10.4103/ijmr.IJMR_1456_17.
343. Choi, J.G., Bharaj, P., Abraham, S., Ma, H., Yi, G., Ye, C., Dang, Y., Manjunath, N., Wu, H., and Shankar, P. (2015). Multiplexing seven miRNA-Based shRNAs to suppress HIV replication. Mol Ther *23*, 310-320. 10.1038/mt.2014.205.
344. Boden, D., Pusch, O., Lee, F., Tucker, L., and Ramratnam, B. (2003). Human Immunodeficiency Virus Type 1 Escape from RNA Interference. J Virol *77*, 11531-11535. 10.1128/jvi.77.21.11531-11535.2003.
345. Wilson, J.A., and Richardson, C.D. (2005). Hepatitis C virus replicons escape RNA interference induced by a short interfering RNA directed against the NS5b coding region. J Virol *79*, 7050-7058. 10.1128/jvi.79.11.7050-7058.2005.
346. Grimm, D., and Kay, M.A. (2007). Combinatorial RNAi: A Winning Strategy for the Race Against Evolving Targets? Mol Ther *15*, 878-888. 10.1038/sj.mt.6300116.
347. Shah, P.S., Pham, N.P., and Schaffer, D.V. (2012). HIV Develops Indirect Cross-resistance to Combinatorial RNAi Targeting Two Distinct and Spatially Distant Sites. Mol Ther *20*, 840-848. 10.1038/mt.2012.3.
348. Lu, A., Zhang, H., Zhang, X., Wang, H., Hu, Q., Shen, L., Schaffhausen, B.S., Hou, W., and Li, L. (2004). Attenuation of SARS coronavirus by a short hairpin RNA expression plasmid targeting RNA-dependent RNA polymerase. Virology *324*, 84-89. 10.1016/j.virol.2004.03.031.
349. Li, B.-j., Tang, Q., Cheng, D., Qin, C., Xie, F.Y., Wei, Q., Xu, J., Liu, Y., Zheng, B.-j., Woodle, M.C., Zhong, N., et al. (2005). Using siRNA in prophylactic and therapeutic regimens against SARS coronavirus in Rhesus macaque. Nat Med *11*, 944-951. 10.1038/nm1280.
350. Harvey, W.T., Carabelli, A.M., Jackson, B., Gupta, R.K., Thomson, E.C., Harrison, E.M., Ludden, C., Reeve, R., Rambaut, A., Peacock, S.J., Robertson, D.L., et al. (2021). SARS-CoV-2 variants, spike mutations and immune escape. Nat Rev Microbiol *19*, 409-424. 10.1038/s41579-021-00573-0.
351. Tolksdorf, B., Nie, C., Niemeyer, D., Röhrs, V., Berg, J., Lauster, D., Adler, J.M., Haag, R., Trimpert, J., Kaufer, B., Drosten, C., et al. (2021). Inhibition of SARS-CoV-2 Replication by a Small Interfering RNA Targeting the Leader Sequence. Viruses.
352. Khaitov, M., Nikonova, A., Shilovskiy, I., Kozhikhova, K., Kofiadi, I., Vishnyakova, L., Nikolskii, A., Gattinger, P., Kovchina, V., Barvinskaia, E., Yumashev, K., et al. (2021). Silencing of SARS-CoV-2 with modified siRNA-peptide dendrimer formulation. Allergy *76*, 2840-2854. 10.1111/all.14850.
353. Idris, A., Davis, A., Supramaniam, A., Acharya, D., Kelly, G., Tayyar, Y., West, N., Zhang, P., McMillan, C.L.D., Soemardy, C., Ray, R., et al. (2021). A SARS-CoV-2 targeted siRNA-nanoparticle therapy for COVID-19. Mol Ther *29*, 2219-2226. 10.1016/j.ymthe.2021.05.004.
354. Chang, Y.-C., Yang, C.-F., Chen, Y.-F., Yang, C.-C., Chou, Y.-L., Chou, H.-W., Chang, T.-Y., Chao, T.-L., Hsu, S.-C., Ieong, S.-M., Tsai, Y.-M., et al. (2022). A siRNA targets and inhibits a broad range of SARS-CoV-2 infections including Delta variant. EMBO Mol Med *14*, e15298. 10.15252/emmm.202115298.
355. Yogev, O., Weissbrod, O., Battistoni, G., Bressan, D., Naamati, A., Falciatori, I., Berkyurek, A.C., Rasnic, R., Izuagbe, R., Hosmillo, M., Ilan, S., et al. (2023). From a genome-wide screen of RNAi molecules against SARS-CoV-2 to a validated broad-spectrum and potent prophylaxis. Commun Biol *6*, 277. 10.1038/s42003-023-04589-5.
356. Bader, W., Delerce, J., Aherfi, S., La Scola, B., and Colson, P. (2022). Quasispecies Analysis of SARS-CoV-2 of 15 Different Lineages during the First Year of the Pandemic Prompts Scratching under the Surface of Consensus Genome Sequences. Int J Mol Sci *23*. 10.3390/ijms232415658.
357. Leist, S.R., Dinnon, K.H., 3rd, Schäfer, A., Tse, L.V., Okuda, K., Hou, Y.J., West, A., Edwards, C.E., Sanders, W., Fritch, E.J., Gully, K.L., et al. (2020). A Mouse-Adapted SARS-CoV-2 Induces Acute Lung Injury and Mortality in Standard Laboratory Mice. Cell *183*, 1070-1085.e1012. 10.1016/j.cell.2020.09.050.
358. Gadenstaetter, A.J., Schmutzler, L., Grimm, D., and Landegger, L.D. (2022). Intranasal application of adeno-associated viruses: a systematic review. Translational Research *248*, 87-110. 10.1016/j.trsl.2022.05.002.
359. Yang, M.S., Park, M.J., Lee, J., Oh, B., Kang, K.W., Kim, Y., Lee, S.M., Lim, J.O., Jung, T.Y., Park, J.H., Park, S.C., et al. (2022). Non-invasive administration of AAV to target lung parenchymal cells and develop SARS-CoV-2-susceptible mice. Mol Ther *30*, 1994-2004. 10.1016/j.ymthe.2022.01.010.
360. Tailor, N., Warner, B.M., Griffin, B.D., Tierney, K., Moffat, E., Frost, K., Vendramelli, R., Leung, A., Willman, M., Thomas, S.P., Pei, Y., et al. (2023). Generation and Characterization of a SARS-CoV-2-Susceptible Mouse Model Using Adeno-Associated Virus (AAV6.2FF)-Mediated Respiratory Delivery of the Human ACE2 Gene. Viruses.
361. Guggino, W.B., Benson, J., Seagrave, J., Yan, Z., Engelhardt, J., Gao, G., Conlon, T.J., and Cebotaru, L. (2017). A Preclinical Study in Rhesus Macaques for Cystic Fibrosis to Assess Gene Transfer and Transduction by AAV1 and AAV5 with a Dual-Luciferase Reporter System. Hum Gene Ther Clin Dev *28*, 145-156. 10.1089/humc.2017.067.

362. Watanabe, S., Ishikawa, K., Plataki, M., Bikou, O., Kohlbrenner, E., Aguero, J., Hadri, L., Zarragoikoetxea, I., Fish, K., Leopold, J.A., and Hajjar, R.J. (2018). Safety and long-term efficacy of AAV1.SERCA2a using nebulizer delivery in a pig model of pulmonary hypertension. Pulm Circ *8*, 2045894018799738. 10.1177/2045894018799738.
363. Calton, M.A., Kotterman, M., Schmitt, C., Klein, A., Burns, C., Croze, R., Nye, J., Beliakoff, G., Quezada, M., Tucker, S., Sullivan, T., et al. (2021). Identification and Characterization of a Novel AAV Capsid and Product for the Treatment of Cystic Fibrosis. In Novel Discoveries in Lung Biology, (American Thoracic Society), pp. A1043-A1043. doi:10.1164/ajrccm-conference.2021.203.1_MeetingAbstracts.A1043
10.1164/ajrccm-conference.2021.203.1_MeetingAbstracts.A1043.
364. Bussani, R., Zentilin, L., Correa, R., Colliva, A., Silvestri, F., Zacchigna, S., Collesi, C., and Giacca, M. (2023). Persistent SARS-CoV-2 infection in patients seemingly recovered from COVID-19. J Pathol *259*, 254-263. 10.1002/path.6035.
365. Harari, S., Tahor, M., Rutsinsky, N., Meijer, S., Miller, D., Henig, O., Halutz, O., Levytskyi, K., Ben-Ami, R., Adler, A., Paran, Y., et al. (2022). Drivers of adaptive evolution during chronic SARS-CoV-2 infections. Nat Med *28*, 1501-1508. 10.1038/s41591-022-01882-4.
366. Zabaleta, N., Dai, W., Bhatt, U., Hérate, C., Maisonnasse, P., Chichester, J.A., Sanmiguel, J., Estelien, R., Michalson, K.T., Diop, C., Maciorowski, D., et al. (2021). An AAV-based, room-temperature-stable, single-dose COVID-19 vaccine provides durable immunogenicity and protection in non-human primates. Cell Host Microbe *29*, 1437-1453.e1438. 10.1016/j.chom.2021.08.002.
367. Zabaleta, N., Bhatt, U., Hérate, C., Maisonnasse, P., Sanmiguel, J., Diop, C., Castore, S., Estelien, R., Li, D., Dereuddre-Bosquet, N., Cavarelli, M., et al. (2022). Durable immunogenicity, adaptation to emerging variants, and low-dose efficacy of an AAV-based COVID-19 vaccine platform in macaques. Mol Ther *30*, 2952-2967. 10.1016/j.ymthe.2022.05.007.
368. Zhao, S., Ke, J., Yang, B., Tan, F., Yang, J., Lin, C.-P., Wang, H., and Zhong, G. (2022). A protective AAV vaccine for SARS-CoV-2. Signal Transduct Target Ther *7*, 310. 10.1038/s41392-022-01158-w.
369. Sims, J.J., Greig, J.A., Michalson, K.T., Lian, S., Martino, R.A., Meggersee, R., Turner, K.B., Nambiar, K., Dyer, C., Hinderer, C., Horiuchi, M., et al. (2021). Intranasal gene therapy to prevent infection by SARS-CoV-2 variants. PLoS Pathog *17*, e1009544. 10.1371/journal.ppat.1009544.
370. Tada, T., Minnee, J., and Landau, N.R. (2023). Vectored immunoprophylaxis and treatment of SARS-CoV-2 infection in a preclinical model. Proc Natl Acad Sci U S A *120*, e2303509120. 10.1073/pnas.2303509120.
371. Coulon, A., Chow, C.C., Singer, R.H., and Larson, D.R. (2013). Eukaryotic transcriptional dynamics: from single molecules to cell populations. Nat Rev Genet *14*, 572-584. 10.1038/nrg3484.
372. Kondratov, O., Kondratova, L., Mandel, R.J., Coleman, K., Savage, M.A., Gray-Edwards, H.L., Ness, T.J., Rodriguez-Lebron, E., Bell, R.D., Rabinowitz, J., Gamlin, P.D., et al. (2021). A comprehensive study of a 29-capsid AAV library in a non-human primate central nervous system. Mol Ther *29*, 2806-2820. 10.1016/j.ymthe.2021.07.010.
373. Melnikov, A., Murugan, A., Zhang, X., Tesileanu, T., Wang, L., Rogov, P., Feizi, S., Gnirke, A., Callan, C.G., Jr., Kinney, J.B., Kellis, M., et al. (2012). Systematic dissection and optimization of inducible enhancers in human cells using a massively parallel reporter assay. Nat Biotechnol *30*, 271-277. 10.1038/nbt.2137.
374. Patwardhan, R.P., Hiatt, J.B., Witten, D.M., Kim, M.J., Smith, R.P., May, D., Lee, C., Andrie, J.M., Lee, S.I., Cooper, G.M., Ahituv, N., et al. (2012). Massively parallel functional dissection of mammalian enhancers in vivo. Nat Biotechnol *30*, 265-270. 10.1038/nbt.2136.
375. Shen, S.Q., Myers, C.A., Hughes, A.E., Byrne, L.C., Flannery, J.G., and Corbo, J.C. (2016). Massively parallel cis-regulatory analysis in the mammalian central nervous system. Genome Res *26*, 238-255. 10.1101/gr.193789.115.
376. Hrvatin, S., Tzeng, C.P., Nagy, M.A., Stroud, H., Koutsioumpa, C., Wilcox, O.F., Assad, E.G., Green, J., Harvey, C.D., Griffith, E.C., and Greenberg, M.E. (2019). A scalable platform for the development of cell-type-specific viral drivers. eLife *8*, e48089. 10.7554/eLife.48089.
377. Carrell, E.M., Chen, Y.H., Ranum, P.T., Coffin, S.L., Singh, L.N., Tecedor, L., Keiser, M.S., Hudry, E., Hyman, B.T., and Davidson, B.L. (2023). VWA3A-derived ependyma promoter drives increased therapeutic protein secretion into the CSF. Mol Ther Nucleic Acids *33*, 296-304. 10.1016/j.omtn.2023.07.016.
378. Mulvey, B., Lagunas, T., Jr., and Dougherty, J.D. (2021). Massively Parallel Reporter Assays: Defining Functional Psychiatric Genetic Variants Across Biological Contexts. Biol Psychiatry *89*, 76-89. 10.1016/j.biopsych.2020.06.011.
379. Bohlen, M.O., McCown, T.J., Powell, S.K., El-Nahal, H.G., Daw, T., Basso, M.A., Sommer, M.A., and Samulski, R.J. (2020). Adeno-Associated Virus Capsid-Promoter Interactions in the Brain Translate from Rat to the Nonhuman Primate. Hum Gene Ther *31*, 1155-1168. 10.1089/hum.2020.196.
380. Gonzalez-Sandoval, A., Pekrun, K., Tsuji, S., Zhang, F., Hung, K.L., Chang, H.Y., and Kay, M.A. (2023). The AAV capsid can influence the epigenetic marking of rAAV delivered episomal genomes in a species dependent manner. Nat Commun *14*, 2448. 10.1038/s41467-023-38106-3.
381. Cabrera, A., Edelstein, H.I., Glykofrydis, F., Love, K.S., Palacios, S., Tycko, J., Zhang, M., Lensch, S., Shields, C.E., Livingston, M., Weiss, R., et al. (2022). The sound of silence: Transgene silencing in mammalian cell engineering. Cell Syst *13*, 950-973. 10.1016/j.cels.2022.11.005.
382. Das, A., Vijayan, M., Walton, E.M., Stafford, V.G., Fiflis, D.N., and Asokan, A. (2022). Epigenetic Silencing of Recombinant Adeno-associated Virus Genomes by NP220 and the HUSH Complex. J Virol *96*, e0203921. 10.1128/jvi.02039-21.
383. Zhan, Y., Brady, J.L., Johnston, A.M., and Lew, A.M. (2000). Predominant transgene expression in exocrine pancreas directed by the CMV promoter. DNA Cell Biol *19*, 639-645. 10.1089/10445490050199045.

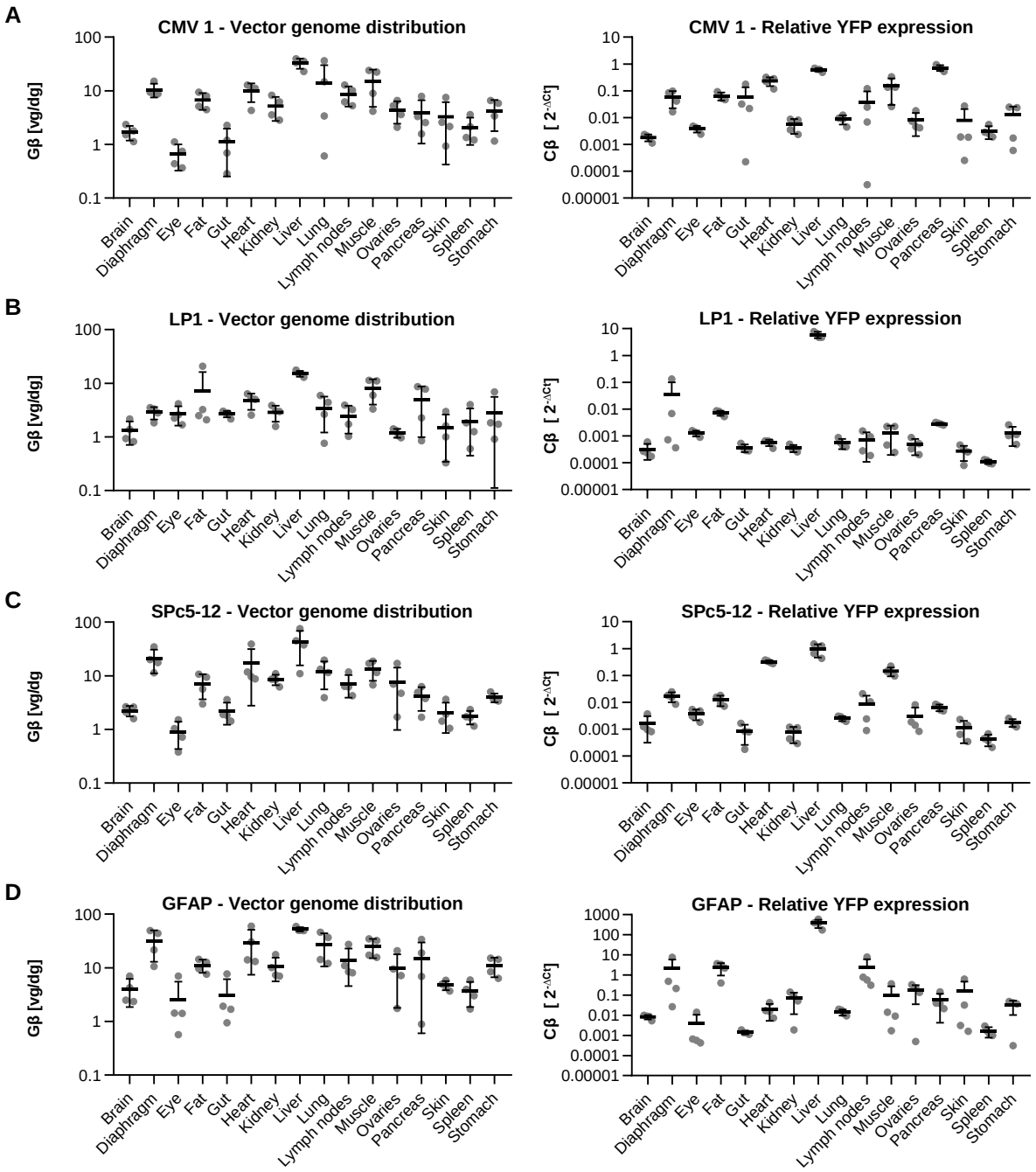
384. Ahlskog, N., Hayler, D., Krueger, A., Kubinski, S., Claus, P., Hammond, S.M., Wood, M.J.A., Yáñez-Muñoz, R.J., and Bowerman, M. (2019). Muscle overexpression of Klf15 via an AAV8-Spc5-12 construct does not provide benefits in spinal muscular atrophy mice. Gene Ther *27*, 505-515. 10.1038/s41434-020-0146-8.
385. Su, M., Hu, H., Lee, Y., d'Azzo, A., Messing, A., and Brenner, M. (2004). Expression specificity of GFAP transgenes. Neurochem Res *29*, 2075-2093. 10.1007/s11064-004-6881-1.
386. Taschenberger, G., Tereshchenko, J., and Kügler, S. (2017). A MicroRNA124 Target Sequence Restores Astrocyte Specificity of gfaABC<sub>1</sub>D-Driven Transgene Expression in AAV-Mediated Gene Transfer. Mol Ther Nucleic Acids *8*, 13-25. 10.1016/j.omtn.2017.03.009.
387. von Jonquieres, G., Mersmann, N., Klugmann, C.B., Harasta, A.E., Lutz, B., Teahan, O., Housley, G.D., Fröhlich, D., Krämer-Albers, E.-M., and Klugmann, M. (2013). Glial Promoter Selectivity following AAV-Delivery to the Immature Brain. PLoS ONE *8*, e65646. 10.1371/journal.pone.0065646.
388. Griffin, J.M., Fackelmeier, B., Fong, D.M., Mouravlev, A., Young, D., and O'Carroll, S.J. (2019). Astrocyte-selective AAV gene therapy through the endogenous GFAP promoter results in robust transduction in the rat spinal cord following injury. Gene Ther *26*, 198-210. 10.1038/s41434-019-0075-6.
389. Thornburg, C.D. (2021). Etranacogene dezaparvovec for hemophilia B gene therapy. Ther Adv Rare Dis *2*, 26330040211058896. 10.1177/26330040211058896.
390. Kattenhorn, L.M., Tipper, C.H., Stoica, L., Geraghty, D.S., Wright, T.L., Clark, K.R., and Wadsworth, S.C. (2016). Adeno-Associated Virus Gene Therapy for Liver Disease. Hum Gene Ther *27*, 947-961. 10.1089/hum.2016.160.
391. Jacobs, F., Gordts, S.C., Muthuramu, I., and De Geest, B. (2012). The Liver as a Target Organ for Gene Therapy: State of the Art, Challenges, and Future Perspectives. Pharmaceuticals.
392. Leng, Y., Li, P., Zhou, L., Xiao, L., Liu, Y., Zheng, Z., Qin, F., Hao, Q., Xu, H., Yao, S., and Dong, B. (2019). Long-Term Correction of Copper Metabolism in Wilson's Disease Mice with AAV8 Vector Delivering Truncated ATP7B. Hum Gene Ther *30*, 1494-1504. 10.1089/hum.2019.148.
393. Tatsumi, K., Ohashi, K., Mukobata, S., Kubo, A., Koyama, F., Nakajima, Y., Shima, M., and Okano, T. (2012). Hepatocyte Is a Sole Cell Type Responsible for the Production of Coagulation Factor IX In Vivo. Cell Med *3*, 25-31. 10.3727/215517912x639496.
394. George, L.A., Sullivan, S.K., Giermasz, A., Rasko, J.E.J., Samelson-Jones, B.J., Ducore, J., Cuker, A., Sullivan, L.M., Majumdar, S., Teitel, J., McGuinn, C.E., et al. (2017). Hemophilia B Gene Therapy with a High-Specific-Activity Factor IX Variant. N Engl J Med *377*, 2215-2227. 10.1056/NEJMoa1708538.
395. Manno, C.S., Chew, A.J., Hutchison, S., Larson, P.J., Herzog, R.W., Arruda, V.R., Tai, S.J., Ragni, M.V., Thompson, A., Ozelo, M., Couto, L.B., et al. (2003). AAV-mediated factor IX gene transfer to skeletal muscle in patients with severe hemophilia B. Blood *101*, 2963-2972. 10.1182/blood-2002-10-3296.
396. Morini, S., Carotti, S., Carpino, G., Franchitto, A., Corradini, S.G., Merli, M., and Gaudio, E. (2005). GFAP expression in the liver as an early marker of stellate cells activation. Ital J Anat Embryol *110*, 193-207.
397. Carotti, S., Morini, S., Corradini, S.G., Burza, M.A., Molinaro, A., Carpino, G., Merli, M., De Santis, A., Muda, A.O., Rossi, M., Attili, A.F., et al. (2008). Glial fibrillary acidic protein as an early marker of hepatic stellate cell activation in chronic and posttransplant recurrent hepatitis C. Liver Transpl *14*, 806-814. 10.1002/lt.21436.
398. Touahri, Y., Dixit, R., Kofoed, R.H., Miloska, K., Park, E., Raeisossadati, R., Markham-Coultes, K., David, L.A., Rijal, H., Zhao, J., Lynch, M., et al. (2020). Focused ultrasound as a novel strategy for noninvasive gene delivery to retinal Müller glia. Theranostics *10*, 2982-2999. 10.7150/thno.42611.
399. Odom, D.T., Dowell, R.D., Jacobsen, E.S., Gordon, W., Danford, T.W., MacIsaac, K.D., Rolfe, P.A., Conboy, C.M., Gifford, D.K., and Fraenkel, E. (2007). Tissue-specific transcriptional regulation has diverged significantly between human and mouse. Nature Genetics *39*, 730-732. 10.1038/ng2047.
400. Brown, H.C., Zakas, P.M., George, S.N., Parker, E.T., Spencer, H.T., and Doering, C.B. (2018). Target-Cell-Directed Bioengineering Approaches for Gene Therapy of Hemophilia A. Mol Ther Methods Clin Dev *9*, 57-69. 10.1016/j.omtm.2018.01.004.
401. Chuah, M.K., Petrus, I., De Bleser, P., Le Guiner, C., Gernoux, G., Adjali, O., Nair, N., Willems, J., Evens, H., Rincon, M.Y., Matrai, J., et al. (2014). Liver-specific transcriptional modules identified by genome-wide in silico analysis enable efficient gene therapy in mice and non-human primates. Mol Ther *22*, 1605-1613. 10.1038/mt.2014.114.
402. Chan, Y.-C., Kienle, E., Oti, M., Di Liddo, A., Mendez-Lago, M., Aschauer, D.F., Peter, M., Pagani, M., Arnold, C., Vonderheit, A., Schön, C., et al. (2023). An unbiased AAV-STARR-seq screen revealing the enhancer activity map of genomic regions in the mouse brain in vivo. Sci Rep *13*, 6745. 10.1038/s41598-023-33448-w.
403. Shen, S., Troupes Andrew, N., Pulicherla, N., and Asokan, A. (2013). Multiple Roles for Sialylated Glycans in Determining the Cardiopulmonary Tropism of Adeno-Associated Virus 4. J Virol *87*, 13206-13213. 10.1128/jvi.02109-13.
404. Nakai, H., Fuess, S., Storm, T.A., Muramatsu, S., Nara, Y., and Kay, M.A. (2005). Unrestricted hepatocyte transduction with adeno-associated virus serotype 8 vectors in mice. J Virol *79*, 214-224. 10.1128/jvi.79.1.214-224.2005.
405. Flitsch, L.J., Börner, K., Stüllein, C., Ziegler, S., Sonntag-Buck, V., Wiedtke, E., Semkova, V., Au Yeung, S.W.C., Schlee, J., Hajo, M., Mathews, M., et al. (2022). Identification of adeno-associated virus variants for gene transfer into human neural cell types by parallel capsid screening. Sci Rep *12*, 8356. 10.1038/s41598-022-12404-0.
406. Gu, W., Crawford, E.D., O'Donovan, B.D., Wilson, M.R., Chow, E.D., Retallack, H., and DeRisi, J.L. (2016). Depletion of Abundant Sequences by Hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. Genome Biol *17*, 41. 10.1186/s13059-016-0904-5.

407. Prezza, G., Heckel, T., Dietrich, S., Homberger, C., Westermann, A.J., and Vogel, J. (2020). Improved bacterial RNA-seq by Cas9-based depletion of ribosomal RNA reads. RNA *26*, 1069-1078. 10.1261/rna.075945.120.
408. Wooley, D.P., Sharma, P., Weinstein, J.R., Kotha Lakshmi Narayan, P., Schaffer, D.V., and Excoffon, K. (2017). A directed evolution approach to select for novel Adeno-associated virus capsids on an HIV-1 producer T cell line. J Virol Methods *250*, 47-54. 10.1016/j.jviromet.2017.09.008.
409. Sallach, J., Di Pasquale, G., Larcher, F., Niehoff, N., Rübsam, M., Huber, A., Chiorini, J., Almarza, D., Eming, S.A., Ulus, H., Nishimura, S., et al. (2014). Tropism-modified AAV Vectors Overcome Barriers to Successful Cutaneous Therapy. Mol Ther *22*, 929-939. 10.1038/mt.2014.14.
410. Zeltner, N., Kohlbrenner, E., Clément, N., Weber, T., and Linden, R.M. (2010). Near-perfect infectivity of wild-type AAV as benchmark for infectivity of recombinant AAV vectors. Gene Ther *17*, 872-879. 10.1038/gt.2010.27.
411. Ravindra Kumar, S., Miles, T.F., Chen, X., Brown, D., Dobreva, T., Huang, Q., Ding, X., Luo, Y., Einarsson, P.H., Greenbaum, A., Jang, M.J., et al. (2020). Multiplexed Cre-dependent selection yields systemic AAVs for targeting distinct brain cell types. Nat Methods *17*, 541-550. 10.1038/s41592-020-0799-7.
412. Chai, S., Wakefield, L., Norgard, M., Li, B., Enicks, D., Marks, D.L., and Grompe, M. (2023). Strong ubiquitous micro-promoters for recombinant adeno-associated viral vectors. Mol Ther Methods Clin Dev *29*, 504-512. 10.1016/j.omtm.2023.05.013.
413. Cabanes-Creus, M., Navarro, R.G., Liao, S.H.Y., Baltazar, G., Drouyer, M., Zhu, E., Scott, S., Luong, C., Wilson, L.O.W., Alexander, I.E., and Lisowski, L. (2021). Single amino acid insertion allows functional transduction of murine hepatocytes with human liver tropic AAV capsids. Mol Ther Methods Clin Dev *21*, 607-620. 10.1016/j.omtm.2021.04.010.
414. Öztürk, B.E., Johnson, M.E., Kleyman, M., Turunç, S., He, J., Jabalameli, S., Xi, Z., Visel, M., Dufour, V.L., Iwabe, S., Pompeo Marinho, L.F.L., et al. (2021). scAAVengr, a transcriptome-based pipeline for quantitative ranking of engineered AAVs with single-cell resolution. eLife *10*, e64175. 10.7554/eLife.64175.
415. Xi, Z., Öztürk, B.E., Johnson, M.E., Turunç, S., Stauffer, W.R., and Byrne, L.C. (2022). Quantitative single-cell transcriptome-based ranking of engineered AAVs in human retinal explants. Mol Ther Methods Clin Dev *25*, 476-489. 10.1016/j.omtm.2022.04.014.
416. Bryant, D.H., Bashir, A., Sinai, S., Jain, N.K., Ogden, P.J., Riley, P.F., Church, G.M., Colwell, L.J., and Kelsic, E.D. (2021). Deep diversification of an AAV capsid protein by machine learning. Nat Biotechnol *39*, 691-696. 10.1038/s41587-020-00793-4.
417. El-Maarri, O., Jamil, M.A., and Oldenburg, J. (2020). Molecular Profiling of Liver Sinusoidal Endothelial Cells in Comparison to Hepatocytes: Reflection on Which Cell Type Should Be the Target for Gene Therapy. Hamostaseologie *40*, S26-s31. 10.1055/a-1282-2286.
418. Feranchak, A.P., and Sokol, R.J. (2001). Cholangiocyte biology and cystic fibrosis liver disease. Semin Liver Dis *21*, 471-488. 10.1055/s-2001-19030.

# 6 Supplementary information

**Supplementary table 1: AAV8-peptide variants included in the DEPOOL validation library.** Peptides of the format "GQ(R/S)GNXXRXXXAQAA" are represented as 9-mer sequences missing the constant flanking sequences (GQ and AQAA).

| Variant Name | 9mer Peptide | Variant Name | 9mer Peptide |
|---|---|---|---|
| DEPOOL_Lu_ND_1 | SGNGTRFTS | DEPOOL_Pa_ND_1 | RGNSSRSVE |
| DEPOOL_Lu_ND_2 | SGNSVRGSM | DEPOOL_Pa_ND_2 | RGNEVRRDH |
| DEPOOL_Lu_ND_3 | SGNATRAFN | DEPOOL_Pa_ND_3 | RGNTIRDSM |
| DEPOOL_Lu_ND_4 | SGNNTRMHA | DEPOOL_Pa_ND_4 | SGNSTRAQI |
| DEPOOL_Lu_ND_5 | SGNHTRDRM | DEPOOL_Pa_ND_5 | SGNASRSNM |
| DEPOOL_Lu_ND_6 | SGNGTRASM | DEPOOL_Pa_ND_6 | SGNEIRVKN |
| DEPOOL_Lu_ND_7 | SGNHTRMSV | DEPOOL_Pa_ND_7 | SGNIIRSME |
| DEPOOL_Lu_ND_8 | SGNMTRDPR | DEPOOL_Pa_ND_8 | SGNHTRDPW |
| DEPOOL_Lu_ND_9 | RGNDVRSQA | DEPOOL_Pa_ND_9 | SGNHTRDSM |
| DEPOOL_Lu_ND_10 | SGNHVRSSV | DEPOOL_Pa_ND_10 | SGNSTRDSI |
| DEPOOL_Lu_D_1 | SGNNVRAQT | DEPOOL_Pa_D_1 | SGNDARKWD |
| DEPOOL_Lu_D_2 | RGNDVRRDT | DEPOOL_Pa_D_3 | SGNHARSVM |
| DEPOOL_Lu_D_3 | SGNNIRNMG | DEPOOL_Pa_D_4 | RGNPTRGID |
| DEPOOL_Lu_D_4 | RGNEVRSAT | DEPOOL_Pa_D_5 | SGNSIRDIK |
| DEPOOL_Lu_D_5 | SGNSVRPQN | DEPOOL_Pa_D_6 | SGNSVRMHN |
| DEPOOL_Lu_D_6 | SGNNVRSGA | DEPOOL_Pa_D_7 | RGNSVRDDH |
| DEPOOL_Lu_D_7 | RGNQVRDFG | DEPOOL_Pa_D_8 | SGNSARMGA |
| DEPOOL_Lu_D_8 | RGNTTRDYT | DEPOOL_Pa_D_9 | SGNNTRSFD |
| DEPOOL_Lu_D_9_PaD2 | SGNICRFCM | DEPOOL_Pa_D_10 | SGNVTRPMQ |
| DEPOOL_Lu_D_10 | SGNDSRRPL | | |




**Supplementary figure 1: Vector genome distribution (left) and relative *eYFP* expression (right) measured for CMV (A), LP1 (B), SPc5-12 (C) and GFAP (D) promoter constructs.**

# 7 Acknowledgements / Danksagung

Hiermit möchte ich mich bei Allen bedanken, die mich während meiner Doktorarbeit unterstützt haben. Dies gilt an erster Stelle Prof. Dr. Dirk Grimm, der mich in seine Forschungsgruppe aufnahm und mir damit die Möglichkeit gab, im dynamischen und äußerst spannenden Feld der AAV-basierten Gentherapie zu arbeiten. Seine Expertise, Ratschläge und Anregungen zu neuen Kollaborationen haben maßgeblich dazu beigetragen, meine Doktorarbeit in der jetzigen Form zu ermöglichen. Ich bin ebenfalls sehr dankbar für seine Unterstützung bei den Manuskripten, zu denen ich beitragen durfte, sowie für die Möglichkeit, wissenschaftliche Konferenzen zu besuchen, um meine Arbeit zu präsentieren.

Weiterhin möchte ich mich bei Prof. Dr. Marc Freichel für seine Unterstützung als Erstgutachter dieser Doktorarbeit bedanken, sowie bei Dr. Marco Binder und Dr. Marina Lusic für die Bereitschaft, an der Prüfungskommission teilzunehmen. Mein Dank gilt auch den drei Mitgliedern meines TAC (Prof. Dr. Dirk Grimm, Prof. Dr. Marc Freichel und Dr. Viet Loan Dao Thi), die während den TAC-Meetings stets konstruktiven Austausch und anregende wissenschaftliche Diskurse ermöglichten. Vielen Dank auch an das HBIGS-Team, besonders Dr. Rolf Lutz und Martina Galvan, für zuverlässige Unterstützung und schnelle Bearbeitungszeiten.

Ich möchte mich weiterhin bei Prof. Dr. Philippe Gual und allen Mitgliedern des MATRIXNASH-Konsortiums bedanken, sowohl für wissenschaftlichen Austausch als auch für die Strukturierung der übergeordneten Projekte. Ich bin weiterhin allen Kollaborationspartnern dankbar, die maßgeblich bei der Planung und Durchführung mehrerer meiner Projekte beteiligt waren. Dazu zählen Prof. Dr. Megan Stanifer und Prof. Dr. Steeve Boulant, mit denen wir gemeinsam das SARS-CoV-2-Projekt bearbeiten durften, und die hierbei stets zuverlässig und engagiert zur Seite standen. Weiterhin bin ich sehr dankbar für die Unterstützung von Prof. Dr. Holger Willenbring und Pervinder Choksi, welche die Untersuchung des GFAP-Promotors und die *in vivo*-Screens der AAV6-Library ermöglichten, und diese Projekte durch ihren Beitrag enorm voranbrachten.

Zwei meiner Projekte wären ohne vorangehende Arbeiten früherer Gruppenmitglieder so nicht möglich gewesen: Ich bin sehr dankbar für die Hilfe und Betreuung durch Dr. Julia Fakhiri, die mir während meiner vorangehenden Praktika und Masterarbeit ermöglichte, die DEPOOL-Strategie mitzuentwickeln. Ebenfalls möchte ich mich bei Dr. Claire Domenger für das Vertrauen bedanken, mir das Promotor-Projekt mit auf den Weg zu geben.

Mein besonderer Dank gilt allen früheren und aktuellen Mitgliedern der Gruppe „Viral Vector Technologies“, die mich während meiner Zeit als Doktorand unterstützten. Dies betrifft insbesondere die TAs Ellen Wiedtke, Lena Naber, Emma Gerstmann, die während ihrer Bachelorarbeit das Cas13d-Projekt vorantrieb, und Chiara Krämer, die unter anderem bei etlichen Gewebsextraktionen behilflich war. Ich möchte mich bei Nico Fischer für unzählige (auch wissenschaftliche) Diskussionen und breit gefächerte Hintergrundmusik bedanken. Weiterhin danke ich Dr. Julia Fakhiri, Dr. Jihad El Andari, Dr. Kleopatra Rapti und Dr. Mischa Schwendy, die mir besonders am Anfang essenzielle methodische Hilfestellung gaben.

Zutiefst dankbar bin ich meiner Familie, allen voran meinen Eltern. Vielen Dank für die unerschöpfliche Unterstützung über all die Jahre in und vor der Doktorarbeit. Dringend gebrauchte Ablenkung und emotionalen Beistand verdanke ich Nataša und Mino. Nataša, neben so vielem anderen hast Du mir immer Perspektive, Kraft und Durchhaltevermögen geschenkt. Vielen Dank für alles.