Dissertation

submitted to the
Combined Faculties of Natural Sciences and Mathematics
of the Ruperto-Carola University Heidelberg, Germany
for the degree of

Doctor of Natural Sciences

Presented by
M. Sc. Julia Fakhiri

Born in: Tartous, Syria
Oral examination: September 24th, 2019

With small viruses come giant responsibilities

Next-generation parvoviral vectors for human gene therapy with extended packaging capacity and enhanced safety profile

Referees:

Prof. Dr. Christof von Kalle

Prof. Dr. Dirk Grimm

CONTENT

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

AAV	Adeno-associated virus
Adh	Adenoviral helper
AF	Alexa fluor
AA	Amino acid
ALU	Arbitrary light units
bp	Basepair
BocaSR	Bocaviral noncoding small RNA
BoV	Bocavirus
BPV	Bovine parvovirus
BSA	Bovine serum albumin
CnMV	Canine minute virus
cap/Cap	Capsid gene/proteins
CBA	Chicken β-actin
CRISPR	Clustered regularly interspaced short palindromic repeats
Cre	Cre recombinase
Cas	CRISPR-associated genes
Cas9	CRISPR-associated protein 9
crRNA	CRISPR RNA
CFTR	Cystic fibrosis transmembrane conductance regulator
CMV	Cytomegalovirus
CMV + I	Cytomegalovirus promoter plus Intron (SV40)
dNTP	Deoxynucleoside triphosphate
DNA	Deoxyribonucleic acid
dg	Diploid genomes
(pA)d	Distal poly A signal
DFS	DNA family shuffling
DSB	Double-stranded break
ddPCR	Droplet digital PCR
E. coli	*Escherichia coli*
EDTA	Ethylenediaminetetraacetic acid
EGTA	Ethylene glycol tetraacetic acid
FBS	Fetal bovine serum

FLuc	Firefly luciferase
GLuc	*Gaussia* luciferase
g	Gram
GFP	Green fluorescent protein
gRNA	Guide RNA
h	Hour
indels	Insertions/Deletions
IVIg	Intravenous immunoglobulin
ITR	Inverted terminal repeats
J	Joule
kb	Kilobase
kDA	Kilodalton
L	Liter
LP1	Liver promoter 1
mRNA	Messenger ribonucleic acid
µF	Microfarad
µg	Microgram
µL	Microliter
µm	Micrometer
µM	Micromolar
mA	Milliampere
mg	Milligram
mL	Milliliter
mM	Millimolar
mV	Millivolt
miCMV	Minimal CMV
min	Minute
MVM	Minute virus of mice
M	Molar
MMTV	Mouse mammary tumor virus
ng	Nanogram
nm	Nanometer
nM	Nanomolar
NS	Non-structural proteins
NP1	Nuclear phosphoprotein 1
nt	Nucleotide
Oligo	Oligonucleotide

ORF	Open reading frame
OD	Optical density
O/N	Overnight
PFA	Paraformaldehyde
P/S	Penicillin/Streptomycin
PAGE	Polyacrylamide gel electrophoresis
pA	Poly A signal
Pol II	Polymerase II
Pol III	Polymerase III
PCR	Polymerase chain reaction
Pre-mRNA	Precursor messenger RNA
pHAE	Primary human airway epithelia
PAM	Protospacer-adjacent motif
(pA)p	Proximal polyA
qPCR	Quantitative PCR
rAAV	Recombinant AAV
RI	Refractive index
g	Relative centrifugal force
RLuc	*Renilla* luciferase
rep/Rep	Replication gene/proteins
rpm	Revolutions per minute
RNA	Ribonucleic acid
RT	Room temperature
RSV	Rous sarcoma virus
SSC	Saline sodium citrate
scAAV	Self-complementary AAV
shRNA	Short hairpin RNA
SALANTO	Shuffling alignment analysis tool
SV40	Simian virus 40 enhancer and early promoter
ssAAV	Single-stranded AAV
SDS	Sodium dodecyl sulfate
SaCas9	*Staphylococcus aureus* Cas9
SpCas9	*Streptococcus pyogenes* Cas9
TBG	Thyroxine-binding globulin
tracrRNA	*Trans*-activating CRISPR RNA
TTR	Transthyretin
UV	Ultraviolet

VP	Viral capsid proteins
V	Volts
v/v	Volume per volume
w/v	Weight per volume
wt	Wild-type
YFP	Yellow fluorescent protein

SUMMARY

Over the last decade, the field of gene and cell therapy has experienced a major turning point and has finally begun to fully realize its potential as a very attractive, versatile and innovative platform for the development of gene-based drugs. Gene therapy encompasses a spectrum of approaches, ranging from supplying missing genes to the correction of diseases at their molecular level, that all have in common the need of a vehicle (“vector") for specific and efficient delivery of therapeutic DNA or RNA. One branch of small viruses - the parvoviruses - have gained increasing attention as such vectors due to their non-pathogenicity, ease of engineering and low genotoxicity. Particularly, the adeno-associated virus (AAV) emerged as a top candidate, culminating in the authorization of three AAV-based gene therapy products, Glybera, Luxturna and Zolgensma. However, despite all the successes using recombinant (r)AAVs, there is still a demand for more specific vectors with larger DNA cargo capacity and lower immunogenicity. This need defined the scope of this doctoral thesis, which aimed at the construction and evaluation of new parvoviral vectors (derived from bocaviruses [BoVs]) and to increase the safety of vector application in humans.

The first part of this work was fueled by a seminal study by Ziying Yan and colleagues in 2013, who used parvovirus cross-genera pseudotyping to combine an oversized rAAV2 genome of 5.5 kilobases (kb) with the capsid of the human bocavirus 1 (HBoV1). As reported, the rAAV2/HBoV1 vector could be produced efficiently and potently transduced primary human airway epithelial cells (pHAE). Here, we have validated and expanded on these intriguing findings by more comprehensively exploring the upper DNA packaging limit of the HBoV1 capsid. Notably, we found that up to 6.2 kb single-stranded (ss) - or 3.2 kb self-complementary (sc) - AAV genomes can be efficiently packaged into the HBoV1 capsid, as compared to only 5.1 (ssAAV) and 2.8 kb (scAAV) for AAV2, which has important ramifications for the delivery of complex rAAV vector DNA.

Next, we further expanded this system to other primate BoV serotypes - three from humans (HBoV2, 3 and 4) and one from Gorilla (GBoV) - that have not been studied as vectors before. To this end, we successfully assembled the capsid genes of HBoV2-4/GBoV and produced chimeric rAAV/BoV vectors of all studied serotypes. With the help of reporter genes, we subsequently started to study and unravel the so-far unknown tropism of the new viral vectors. Strikingly, our screens on various primary cells and cell lines revealed that BoVs (especially GBoV) have a much wider tropism *in vitro* than previously anticipated. We found a wide range of primary and therapeutically relevant cells to be amenable to BoV infection, including human hepatocytes, T-cells and skeletal muscle cells. In addition, we obtained the first evidence that pseudotyped rAAV/BoV vectors also differ in their reactivity to pooled human

antibodies (intravenous immunoglobulin, IVIg), which implies the possibility of vector re-dosing in rAAV/BoV-treated human gene therapy patients. Finally, we aimed to increase the fitness of BoV vectors and therefore employed a high-throughput diversification method called DNA family shuffling (DFS), to create the first library of chimeric BoV capsids. As hoped for, the library was packaging-competent, increased in titer over selection rounds and acquired a unique footprint when cycled in pHAE.

Despite an excellent safety record of rAAV vectors, undesirable toxicity resulting from permanent gene expression represents a clinical concern. So far, the ensuing need to gain temporal control over vector persistence or expression has been addressed by using Cre recombinase or inducible systems that necessitate complex vector re-engineering. Thus, in the second part of this work, we aimed to overcome these limitations by introducing novel rAAV vectors that harbor a kill-switch (KS) based on the bacterial CRISPR II system (clustered regularly interspaced short palindromic repeats). This approach has two major components: (i) a (g)uide RNA expressed from the rAAV vector and (ii) the CRISPR/Cas9 endonuclease, which is supplied *in trans* and directed by the gRNA to a target site in the vector/transgene itself. We tested our KS system extensively *in vitro* and show a 10- to 100-fold reduction in transgene expression (Firefly luciferase) after supplying Cas9 *in trans* using ss and scAAV vectors for the expression of full-length and split Cas9, respectively. Moreover, we expanded our study to an *in vivo* application in mice, where we could recapitulate our findings in cell culture and trigger an up to 50% reduction in transgene expression. Finally, we devised a universal approach to inactivate any rAAV vector without further modifications. Therefore, we developed and experimentally validated self-inactivating (SIN) CRISPR vectors based on split Cas9 and ssAAVs that harbor the anti-target and anti-Cas9 gRNA and hence allow concurrent targeting of both. Moreover, we utilized different RNA polymerase III promoters (Pol III) to study and eventually optimize the effect of differential gRNA expression on the kinetics of both processes.

Collectively, this work has yielded original BoV helper constructs and chimeras that represent valuable new tools to investigate fundamental and applied aspects of bocaviral biology, from the discovery of antigenic domains to the construction of designer viral vectors. Concomitantly, we have implemented and validated novel concepts to increase the safety of recombinant vectors including rAAV KS or SIN constructs that can be harnessed in future work, either alone or in combination with BoV capsids, to form the next generation of parvoviral vectors.

ZUSAMMENFASSUNG

Das Fachgebiet der Gen- und Zelltherapie hat im letzten Jahrzehnt immens an Bedeutung gewonnen, und sein Potenzial als vielseitige und innovative Plattform für die Entwicklung genbasierter Arzneimittel wird nun endlich voll ausgeschöpft. Die Gentherapie umfasst ein breites Spektrum therapeutischer Ansätze, die vom Ersatz fehlender Gene bis hin zur Therapie von Krankheiten auf molekularer Ebene reichen. Allen gemeinsam ist die Verwendung eines Trägervehikels („Vektors“) für die gezielte und effiziente Einschleusung therapeutischer DNA oder RNA. Eine Familie kleiner Viren, der Parvoviren, hat aufgrund ihrer Apathogenität, ihrer technischen Unkompliziertheit und ihrer geringen Genotoxizität bei der Herstellung von Vektoren zunehmend an Bedeutung gewonnen. Insbesondere das Adeno-assoziierte Virus (AAV) hat sich zu einem Spitzenkandidaten entwickelt, der die Zulassung von drei AAV-basierten Gentherapeutika ermöglichte, Glybera, Luxturna und Zolgensma. Trotz aller Erfolge mit rekombinanten (r)AAVs besteht nach wie vor ein Bedarf an spezifischeren Vektoren mit größerer DNA-Ladekapazität und geringerer Immunogenität. Diese offenenen Punkte waren die Motivation für die vorliegende Dissertation, deren Ziel die Entwicklung neuer Vektoren war, die von anderen Parvoviren (den Bocaviren [BoV]) abgeleitet werden und die die Anwendungssicherheit im Menschen erhöhen.

Der erste Teil dieser Arbeit wurde durch die Pionierarbeit von Ziying Yan im Jahr 2013 inspiriert, der mit Hilfe der gattungsübergreifenden Parvovirus-Pseudotypisierung ein über-dimensionales rAAV2-Genom von 5,5 Kilobasen (kb) mit dem Kapsid des humanen Bocavirus 1 (HBoV1) kombinierte. Der resultierende rAAV2/HBoV1-Vektor konnte effizient hergestellt werden und erwies sich als äußerst wirksam bei der Transduktion primärer menschlicher Epithelzellen der Atemwege („primary human airway epithelial cells“ - pHAE). In dieser Arbeit haben wir diese faszinierenden Ergebnisse validiert und weiter ausgearbeitet, indem wir die obere DNA-Verpackungsgrenze des HBoV1-Kapsids genauer untersuchten. Insbesondere stellten wir fest, dass bis zu 6,2 kb einzelsträngige (ss) - oder 3,2 kb selbst-komplementäre (sc) - AAV-Genome effizient in das HBoV1-Kapsid verpackt werden können, verglichen mit nur 5,1 kb (ssAAV) bzw. 2,8 kb (scAAV) für AAV2, was bedeutende Auswirkungen auf die Übertragung von komplexer rAAV-Vektor-DNA hat.

Als nächstes weiteten wir dieses System auf andere Serotypen des Primaten-Bocavirus aus - drei vom Menschen (HBoV2, 3 und 4) und einer vom Gorilla (GBoV) - die bisher nicht auf ihre Eignung als Vektoren untersucht worden waren. Zu diesem Zweck assemblierten wir die Kapsidgene von HBoV2-4/GBoV erfolgreich und stellten chimäre rAAV/BoV-Vektoren aller untersuchten Serotypen her. Mit Hilfe von Reportergenen begannen wir, den bisher unbekannten Tropismus der neuen viralen Kapside zu erforschen und zu entschlüsseln.

Erstaunlicherweise zeigten unsere Analysen verschiedener primärer Zellen und Zelllinien, dass Bocaviren (insbesondere GBoV) *in vitro* einen viel breiteren Tropismus aufweisen als bisher angenommen. Wir fanden heraus, dass eine Vielzahl von primären und therapeutisch relevanten Zellen für eine BoV-Infektion zugänglich ist, darunter menschliche Hepatozyten, T-Zellen und Skelettmuskelzellen. Des Weiteren konnten wir erstmals belegen, dass sich pseudotypisierte rAAV/BoV-Vektoren auch im Hinblick auf ihre Reaktivität gegenüber humanen Antikörpern (intravenöses Immunglobulin, IVIg) unterscheiden, was die Möglichkeit impliziert, mit rAAV/BoV behandelte Menschen mehrfach zu therapieren. Darüber hinaus war unser Ziel eine Erhöhung der Fitness von BoV-Vektoren, weshalb wir eine Hochdurchsatz-Diversifikationsmethode namens „DNA family shuffling“ (DFS) einsetzten, um die erste Bibliothek für chimäre BoV-Kapside zu erstellen. Wie erhofft, erwies sich die Bibliothek als verpackungskompetent, verzeichnete im Laufe der Selektionsrunden einen Titeranstieg und erwarb bei wiederholter Amplifikation in pHAE ein einzigartiges Profil.

Trotz der ausgezeichneten Sicherheitsbilanz der rAAV-Vektoren stellt eine unerwünschte Toxizität aufgrund permanenter Genexpression ein klinisches Sicherheitsrisiko dar. Bisher wurde die daraus resultierende Notwendigkeit, die Vektorpersistenz oder -expression zeitlich zu steuern, durch den Einsatz von Cre-Rekombinase oder induzierbaren Systemen gelöst, die allerdings eine komplexe Vektor-Modifikation erfordern. Daher versuchten wir im zweiten Teil dieser Arbeit, diese Einschränkungen zu überwinden, indem wir neuartige rAAV-Vektoren mit einem „Kill-Switch" (KS) einführten, der auf dem bakteriellen CRISPR-II-System („clustered regularly interspaced short palindromic repeats") basiert. Dieser Ansatz besteht aus zwei Hauptkomponenten: (I) einer Guide-RNA (gRNA), exprimiert aus dem rAAV-Vektor und (ii) der CRISPR/Cas9-Endonuklease, die *in trans* geliefert und von der gRNA an eine Zielstelle im Vektor/Transgen selbst gelenkt wird. Die umfassende Erprobung unseres KS-Systems *in vitro* zeigte eine 10- bis 100-fache Reduktion der Transgenexpression (Leuchtkäfer-Luciferase) nach Zuführen von Cas9 *in trans* unter Verwendung von ss und scAAV-Vektoren für die Expression von volllständigem oder geteiltem ("split") Cas9 Protein. Darüber hinaus weiteten wir unsere Studie auf eine *in vivo* Anwendung bei Mäusen aus, im Rahmen derer wir eine bis zu 50% Reduktion der Transgenexpression erreichten. Anschließend entwickelten wir einen universellen Ansatz, mit dem jeder rAAV-Vektor ohne weitere Modifikationen deaktiviert werden kann. Zu diesem Zweck entwarfen und testeten wir selbst-inaktivierende (SIN) CRISPR-Vektoren auf der Basis von split Cas9 und ssAAVs, die sowohl die Anti-Target- als auch die Anti-Cas9-gRNA enthalten und somit eine gleichzeitige Inaktivierung beider (Vektorgenom und Cas9) ermöglichen. Darüber hinaus verwendeten wir verschiedene Promotoren der RNA-Polymerase III (Pol III), um die Wirkung der differentiellen gRNA-Expression auf die Kinetik beider Prozesse zu untersuchen und gegebenenfalls zu optimieren.

Zusammengefasst hat diese Arbeit BoV-Helferkonstrukte und -Chimären hervorgebracht, die wertvolle neue Werkzeuge darstellen, um grundlegende und angewandte Aspekte der Biologie des Bocavirus zu untersuchen, von der Entdeckung antigener Domänen bis zur Konstruktion von Designer-Virusvektoren. Gleichzeitig haben wir neuartige Konzepte zur Verbesserung der Sicherheit rekombinanter Vektoren implementiert und validiert, einschließlich rAAV-SIN-Konstrukte, die in zukünftigen Arbeiten entweder allein oder in Kombination mit BoV-Kapsiden zur Bildung der nächsten Generation parvoviraler Vektoren genutzt werden können.

1 INTRODUCTION

Over five decades ago, scientists began to use microinjection to introduce foreign DNA and RNA into eukaryotic cells - a technology that eventually revolutionized our understanding of biology and diseases [1]. Soon after, the term “genetic engineering” was coined, which describes the change in the genetic landscape of a cell triggered by the direct transfer of foreign DNA [2]. Originally, these DNA molecules were composed of any desired nucleotide sequences in any combination and were synthesized using recombinant DNA technology. The DNA was then either directly introduced to the organisms using physical methods (*e.g.* the aforementioned microinjection) or indirectly through a carrier, also known as a “vector”. Among the different vector systems, viral vectors are nowadays considered as the most efficient and thus in the focus of the current doctoral thesis. The success of gene transfer culminated in the creation of the first gene-modified organism (GMO) in 1973 [3] and the first transgenic mouse in 1974 [4]. At the same time, scientists realized the potential of gene transfer for treatment of diseases at their molecular level. This especially applies to monogenic diseases, which are caused by one defective gene copy and would only require its replacement by a healthy equivalent. To this end, French Anderson performed the first therapeutic use of gene transfer in 1990 to treat severe combined immunodeficiency (SCID) in a child [5, 6].

Since then, several viruses have been developed as vectors for human gene therapy including (i) RNA viruses, such as retroviruses and lentiviruses, and (ii) DNA viruses, such as adenoviruses, herpes simplex virus (HSV) and parvoviruses. The latter have gained increasing attention due to their broad applicability and unique characteristics, including the anti-oncogenicity of autonomous parvoviruses [7] or the non-pathogenicity and long-term expression of dependoviruses [8]. Moreover, the establishment of methods for efficient vector production was a milestone and prerequisite for diverse *in vitro* and *in vivo* applications.

One particular Dependovirus - the adeno-associated virus or short AAV - has emerged as a lead candidate for sustained clinical gene expression, due to its amenability to genetic manipulation and the excellent safety profile. However, one limitation of the so-called recombinant (r)AAV vectors is their small cargo capacity of 4.7 kb DNA as compared to other parvoviruses (*e.g* ~5.5 kb for bocaviruses [BoVs] or ~6 kb for densoviruses [9]). Moreover, the typically broad cell specificity of natural AAVs requires the administration of high doses to reach a therapeutic benefit in a given cell or tissue type. Therefore, methods for the development of a new generation of rAAVs were established, and other parvoviruses with larger packaging ability and/or higher specificity were concurrently studied and vectorized. One example is a recent study [10], in which the parvovirus human BoV1 (HBoV1) has been utilized, for the first time, as a vector for gene transfer. This vector showed a unique ability to cross-package 5.5

kb rAAV genomes and to efficiently target primary human airway epithelia (pHAE). This advance in vector technology inspired the first aim of this doctoral thesis, which was to expand on this published data and to characterize other primate BoV serotypes with as-of-yet unknown tropisms that can be harnessed, and may be useful, for therapeutic gene transfer.

Next to size, a second potential limitation of AAV and BoV vector-delivered DNA is that remains extra-chromosomal in the cells and consequently gets lost during cell division. This restricts the application to mainly non-dividing but long-living cells like, for instance, liver and muscle cells. The target range of these vectors could, however, be extended to dividing cells by packaging a cargo that introduces a permanent change in the DNA that is subsequently transmitted to the daughter cells. Here, the adaptation of the bacterial CRISPR machinery (clustered regularly interspaced short palindromic repeats) for gene editing in eukaryotes set a milestone in the history of genetic engineering and opened new possibilities for gene therapy of human diseases (see sections 1.4 to 1.8).

To date, CRISPR has already been used in combination with rAAV vectors (rAAV-CRISPR) to permanently and efficiently modify genes with clinical relevance in cells of the brain [11], liver [12], eye [13] and muscle [14], thereby correcting disease-causing genotypes in these cells. Still, packaging of all the CRISPR components (expression cassettes for Cas9 and gRNA, and, if desired, also a DNA repair template for homologous recombination) into one rAAV vector is challenging and only possible if minimal forms of promoters and regulatory elements are used. Consequently, a second seminal goal in this thesis was to address and resolve the size issue, by harnessing the larger cargo capacity of the HBoV1 capsid to package all-in-one rAAV-CRISPR cassettes. Moreover, we aimed at increasing the safety of our applications by constructing simple but powerful genetic circuits that allow controlling both, Cas9 and transgene expression from rAAV vectors.

The following chapters will provide an overview of parvovirus biology and associated vectors for gene transfer, with a special focus on pseudotyped and chimeric parvoviral vectors. The second part of the introduction will focus on the emerging role of CRISPR technology, especially its clinical translation for treatment of human diseases.

1.1 The biology of parvoviruses

The *Parvovirinae* family includes small, non-enveloped viruses with a single-stranded (ss)DNA genome. The name of the family is derived from the Latin term “parvus”, which means small and describes the hallmark of these viruses, *i.e.* their diminutive size. The viral capsid is around 23 – 25 nm in diameter and encapsidates a genome with a length between 4.7 – 6 kb. Because of the limited size, the genome is tightly packed with overlapping open reading frames (ORFs), which comprise two major genes (**Figure 1**) encoding: (i) the non-structural proteins

(NS, Rep or NP1) and (ii) the capsid proteins (VP or Cap). Consequently, parvoviruses are strongly dependent on the host cell for providing factors for their genome replication and gene expression. Extreme examples are the dependoviruses, which even require the help of other, more complex viruses to be able to replicate and form progeny.

Primate parvoviruses are classified into four genera: *Parvovirus*, *Dependovirus*, *Erythrovirus* and *Bocavirus*. The genome organization across the genera is similar but there are differences in the numbers of promoters (from 1 to 3) and of expressed VP or Cap proteins (from 2 to 3), as shown in **Figure 1**.

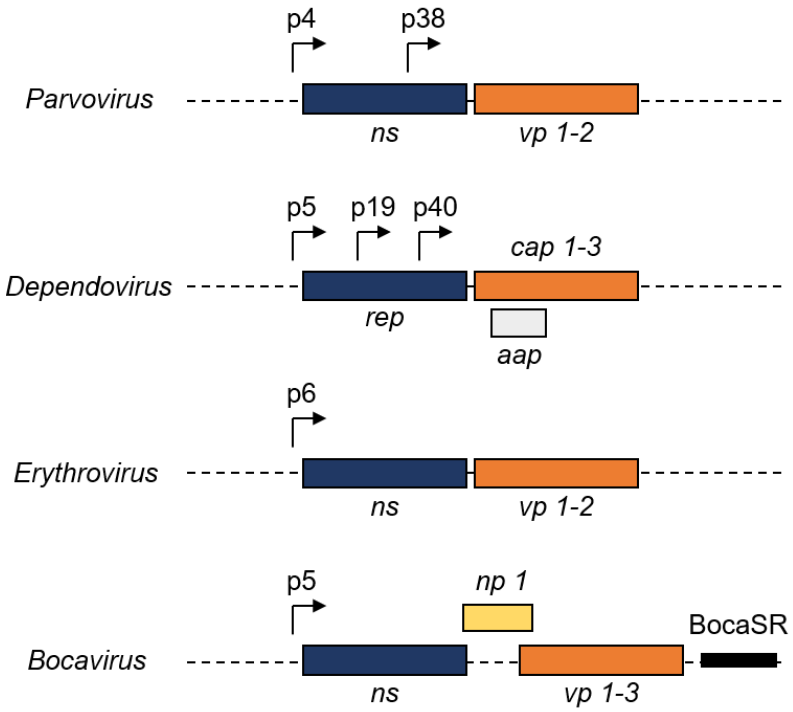


Figure 1. Genome organization of different genera in the *Parvovirinae* family.

The total genome size of the shown parvoviruses ranges from 4.7 to 6 kb, including the (inverted) terminal repeats that flank the coding region and that were omitted here. There are two major open reading frames encoding structural (VP or Cap) and non-structural (NS, NP1 or Rep) proteins. The number of promoters (p) at different genomic map units ranges from 1 to 3. *ns* = non-structural; *vp* = viral (cap) protein; *rep* = replication protein, *aap* = assembly-activating protein; *np 1* = nuclear phosphoprotein 1; BocaSR = Bocaviral non-coding small RNA. Modified from reference 9.

1.1.1 Dependovirus

The genus *Dependovirus* includes three main members: the AAV, the goose parvovirus (GPV) and the muscovy duck parvovirus (MDPV).

AAV was first discovered in 1965 as a contaminant of an adenovirus preparation, when it appeared as small, icosahedral particles in electron microscopy images [15]. In this original work, Atchison and co-workers showed that these particles were infectious and exhibited a different immunological profile than adenovirus. Moreover, they also reported that AAV is incapable of replicating in the absence of adenovirus and consequently stated that the virus might be defective [15]. Later, several reports revealed that other viruses can also act as helpers

for AAV replication such as herpes simplex virus (HSV) 1 and 2 [16], papilloma virus [17], vaccinia virus [18] and, discovered most recently, HBoV1 [19].

In 1966, Hoggan *et al.* reported another important feature of the AAV life cycle. As soon as AAV starts replicating, it forms a negative-feedback loop that inhibits the further replication and propagation of the helper virus. This has been observed for adenovirus [20] and later also for HSV [21]. By contrast, HBoV1 infection seems not to be influenced by the presence of AAV [19]. This is interesting since HBoV1 has not been clearly associated with diseases in humans, in contrast to adenovirus and HSV, implying that natural AAV infection may protect against co-infection with certain viral pathogens.

In the absence of a helper virus, AAVs establish latency by persisting as episomes or by integrating into specific loci in the host chromosome (*e.g.* the AAVS1 locus for AAV2 [22]). Here, the AAV Rep proteins play an important role by facilitating the site-specific integration [23, 24] and silencing of the integrated genome [8]. Recently, a study by Nault *et al.* [25] associated wtAAV2 integration in human liver with insertional mutagenesis in hepatocellular carcinomas (HCC). Interestingly, a study by Logan and colleagues [26] identified a region in the 3' UTR of wtAAV2 that contains an enhancer–promoter element composed of binding sites for liver-specific transcription factors. Importantly, this element was found to be part of the integrated wtAAV2 stretch in 10 of the 11 reported HCC cases and was thus suggested as a cause of HCC-associated gene dysregulation. This challenges all published studies to date, which consistently showed that infection with wtAAV was not associated with any pathology and, on the contrary, might even protect from cancer [27-29]. Collectively, while these two mentioned studies [25, 26] certainly revealed interesting biological observations, they could neither prove the involvement of wtAAV in promoting HCC, nor exclude a secondary integration after the initiation of tumorgenesis [30]. The integration propensity of AAV-based viral vectors and the development of systems that integrate in a specific and safe manner [31, 32] are discussed in section 1.2.2.

The genome of AAV contains three promoters (p5, p19 and p40) that drive the expression of the three viral genes: *rep*, *cap* and *aap* (as shown in **Figure 1**). Two of the promoters (p5 and p19) regulate the expression of the *rep* gene, which encodes four Rep proteins designated Rep78, Rep68, Rep52 and Rep40. The p40 promoter drives the expression of the *cap* gene, which encodes the three capsid proteins VP1/VP2/VP3 and the recently discovered AAP [33].

AAVs have been found in all animal species studied so far, including various mammals, snakes or birds [34-36]. Among the primate AAVs, 13 serotypes and hundreds of variants have been isolated and shown to have a wide cell and tissue tropism. The latter can partly be explained by the ubiquitous expression of the known AAV receptors. For example, two common cell surface molecules, heparan sulfate proteoglycans (HSPG) [37] and sialic acids,

were identified as receptors for AAV2, 3, 6 and AAV1, 4, 5, 6 [38], respectively. Even more remarkable was the discovery of a universal (and broadly distributed) AAV receptor (AAVR) that governs the infectivity of a majority of AAV serotypes [39, 40]. From an evolutionary perspective, it is unclear why AAV displays such a large tissue distribution as compared to other parvoviruses such as B19. It has been hypothesized that the promiscuity might offer an advantage for the virus because it increases the probability of meeting a helper virus to support its replication [41]. On the other hand, the host also benefits from a broad tropism, as the ability of AAVs to suppress their helper viruses increases the host's chances of surviving infection with a range of pathogenic viruses that target different tissues [27-29].

Unlike AAVs, GPV and MDPV can autonomously replicate in permissive and dividing cells [42]. Curiously, in contrast to other autonomous parvoviruses, GPV is also able to replicate in resting cells but only in the presence of a helper virus. These two viruses were still classified as dependoviruses because of their close phylogenetic relationship to AAV2 [43]. In 2005, Qiu *et al.* performed a detailed analysis of the transcriptional landscape of GPV and revealed, not surprisingly, features of both dependo- and autonomous parvoviruses [44].

1.1.2 Autonomous parvoviruses

Helper-independent or autonomous parvoviruses, contrary to AAVs, can replicate and spread in the absence of any helper viruses. The first autonomous parvoviruses were detected in tumor tissue and thus, for a long time, considered as etiologic agents for cancer. However, a role for these viruses in provoking tumorigenesis could never be shown *in vitro* or *in vivo* [45]. By contrast, several early *in vivo* studies in laboratory animals proved that autonomous parvoviruses can actually interfere with tumor formation and act as oncosuppressive agents in two different ways: (i) a protective manner by reducing the incidence of spontaneous tumor formation [45, 46] or (ii) a strong inhibition of tumor growth [47] up to complete regression [48].

The autonomous parvoviruses belong to four genera: *Parvovirus*, *Erythrovirus*, *Bocavirus* and the more distantly related *Amdovirus.* The *Parvovirus* genus includes a variety of viruses with different cell tropisms that infect vertebrates and non-vertebrates. Among the most studied members of this genus are the minute virus of mice (MVM), the rat parvovirus H1 and LuIII. By contrast, viruses in the *Erythrovirus* genus have a very specific tropism for red blood cell progenitors. The first isolated member is B19, a pathogenic erythrovirus responsible for the fifth disease in humans [49]. Parvoviruses belonging to the genus *Bocavirus* have a unique additional ORF that encodes the protein NP1 (see **Figure 1**), whose function and relevance will be further discussed in section 1.1.2.3. The *Amdovirus* genus includes only one highly pathogenic member, namely, the Aleutian mink disease virus (AMDV).

The permissiveness of cancer cells for many autonomous parvoviruses, especially members belonging to the *Parvovirus* genus, is complex and was shown to be dependent on different factors comprising the tissue of origin [50, 51], the expressed oncogenes [52, 53] or the

combination thereof [54]. A detailed discussion of all these factors is beyond the scope of this introduction but can be found in numerous excellent reviews (see references 55, 56). In general, the oncotropism can be linked to the coupling of the viral life cycle to the S-phase of the cell [7, 56]. Particularly, the replication of the single-stranded parvoviral DNA is strictly dependent on cyclin A, which is induced when a cell enters the S-phase [57]. The mechanisms by which these viruses induce dell death in later stages are not well understood. It has been shown, however, that certain oncoproteins increase the replication and expression of parvovirus proteins, some of which undergo changes in post-translational modifications (*e.g*. phosphorylation state [58]) and accumulate to cytotoxic levels in the cells [52, 53].

The next chapter will focus on the biology of members of the *Bocavirus* genus, which are extensively studied and utilized as gene transfer vectors in the first part of this work.

1.1.2.1 Bocavirus

BoVs are composed of a small icosahedral capsid with a ssDNA genome of ~5.5 kb. The genus name “Boca” stems from the first isolated animal bocaparvoviruses, (Bo)vine parvovirus (BPV) and (ca)nine minute virus (CnMV), which were described in 1960 [59] and 1970 [60], respectively. Both viruses were associated with reproductive failure and neonatal respiratory diseases [61-63]. Since then, many other BoVs have been discovered in primates (such as HBoV [64, 65] and GBoV [66] in human and gorilla, respectively) and non-primates including pigs, dogs, mice, rabbits, bats and sea lions [67]. Primate BoVs share ~40% sequence identity with all animal BoVs but more than 70% among each other. BPV is the most distinct bocaparvovirus with a capsid composed of four VP proteins [68]. By contrast, the capsid in the more closely related (to each other) CnMV and HBoV1 is composed of only three capsid proteins, VP1-3 [69, 70].

1.1.2.2 The primate BoV

The first discovered primate BoV is HBoV1, which was isolated in 2005 from human respiratory samples [64]. Since then, the virus was extensively studied to unravel its role as an etiologic agent of lower respiratory tract infections. Many studies have revealed a global prevalence with infection rates of up to 10% [71]. However, studies of disease association have yielded very conflicting results and, surprisingly, more than one decade later, it is still unclear whether the virus is a true pathogen or not. Notably, the high degree of coinfections with disease-causing agents and the lack of animal models or easy-to-handle cell culture systems further hampered the true classification of the virus. Fortunately, the discovery that HBoV1 replicates in pHAE allowed for the study of its transcriptional profile and pathogenesis. For example, Huang *et al.* reported a loss of cilia and a disruption of the tight junctions upon infection with HBoV1 [72]. Moreover, Luo and Liu *et al.* [73, 74] showed that infection with HBoV1 results in an innate immune response, which the virus escapes in different ways to establish latency (see discussion section 4.1.4).

In the years following the report of HBoV1, other human BoVs called HBoV2-4 were discovered in stool and were readily linked to gastroenteritis [65]. These viruses show a high sequence identity to HBoV1 (>74% for the *ns*, *np1* and *cap* ORFs) but were not or only sporadically detected in respiratory samples [75]. At the same time of their discovery, Kantola *et al.* described the first gorilla BoV (GBoV), which was detected in stool [66]. This variant shares a high sequence identity with the human isolates but was shown to carry a unique *ns* ORF that encodes an extended version of the NS protein [66]. As for HBoV1, it is unclear whether these BoVs are true pathogens or innocent bystanders.

1.1.2.3 The transcriptional landscape of HBoV1

The transcriptional profile of HBoV1 is similar to the one observed in other animal bocaparvoviruses such as BPV [76] and CnMV [77]. A single promoter drives the transcription of a precursor (pre)-mRNA that is alternatively spliced and polyadenylated to generate the different non-structural (NS/NP) and structural (VP) proteins.

In HBoV1, the pre-mRNA is transcribed from the p5 promoter and alternatively spliced using the different donor (D) and acceptor (A) sites shown in **Figure 2**. The non-structural genes *ns* and *np1* encode numerous major transcripts that translate into four (NS1-4) and one (NP1) protein(s), respectively. Transcription can be terminated at the proximal (pA)p or distal (pA)d polyA signal, which results in a short (S) or long (L) mRNA transcript, respectively. There are also minor forms (m) of mRNA that retain the D2/A2 intron, which generates an in-frame stop codon. This results in the formation of smaller NS proteins such as NS1-70 (**Figure 2**) and NS2‘, NS3‘ and NS4‘ [78] (not shown but described in more detail in reference [78]). The abundance of these proteins is low and their function remains enigmatic. Moreover, additional small proteins NS1/2*35-40 and NS3*50 were detected during a productive infection of pHAE with wild-type (wt) HBoV1. Likewise, the nature and function of these proteins is not well understood [78].

The unspliced NS protein (NS1) contains a DNA-binding/endonuclease domain, a helicase-activity domain, a potential oligomerization site and several transactivation (TAD) domains [78]. This protein is critical for efficient viral replication in HEK293T cells, whereas NS2-4 are dispensable. Curiously, NS2 is strictly required for wtHBoV1 replication in pHAE, which may reflect a cell-dependent protein function. The exact roles of NS3 and NS4 are currently unknown.

The unique BoV protein NP1 is expressed from an ORF in the middle of the genome that overlaps with the *ns* and *cap* ORFs. This protein has been shown to be important for replication of the viral DNA and regulation of transcription [77, 79, 80]. In more detail, Zou *et al.*[70] showed that in the HBoV1 genome, NP1 regulates VP expression by (i) mediating splicing at the A3 acceptor site, which is a prerequisite for production of VP-encoding transcripts and (ii)

suppression of the (pA)p signal, which lies within the *cap* ORF. This allows the transcription machinery to reach the (pA)d signal, which consequently results in the production of full-length Cap polycistronic pre-mRNA. The mechanism by which NP1 interacts with the transcriptional machinery remains obscure.

This VP pre-mRNA expresses the three viral capsid proteins, VP1/VP2/VP3, by using distinct initiation codons. This is different to what has been observed for CnMV, where VP3 results from the proteolytic cleavage of VP2 by host proteases [69].

Another role for NP1 in CnMV was reported by Fasina *et al.* [80]. In this study, three novel NS transcripts that fuse their carboxy termini in frame to the carboxy terminus of NP1 were identified. NP1 regulates the expression of these transcripts by mediating splicing of the pre-mRNA. In HBoV1, a similar subset of NS proteins has not been reported so far.

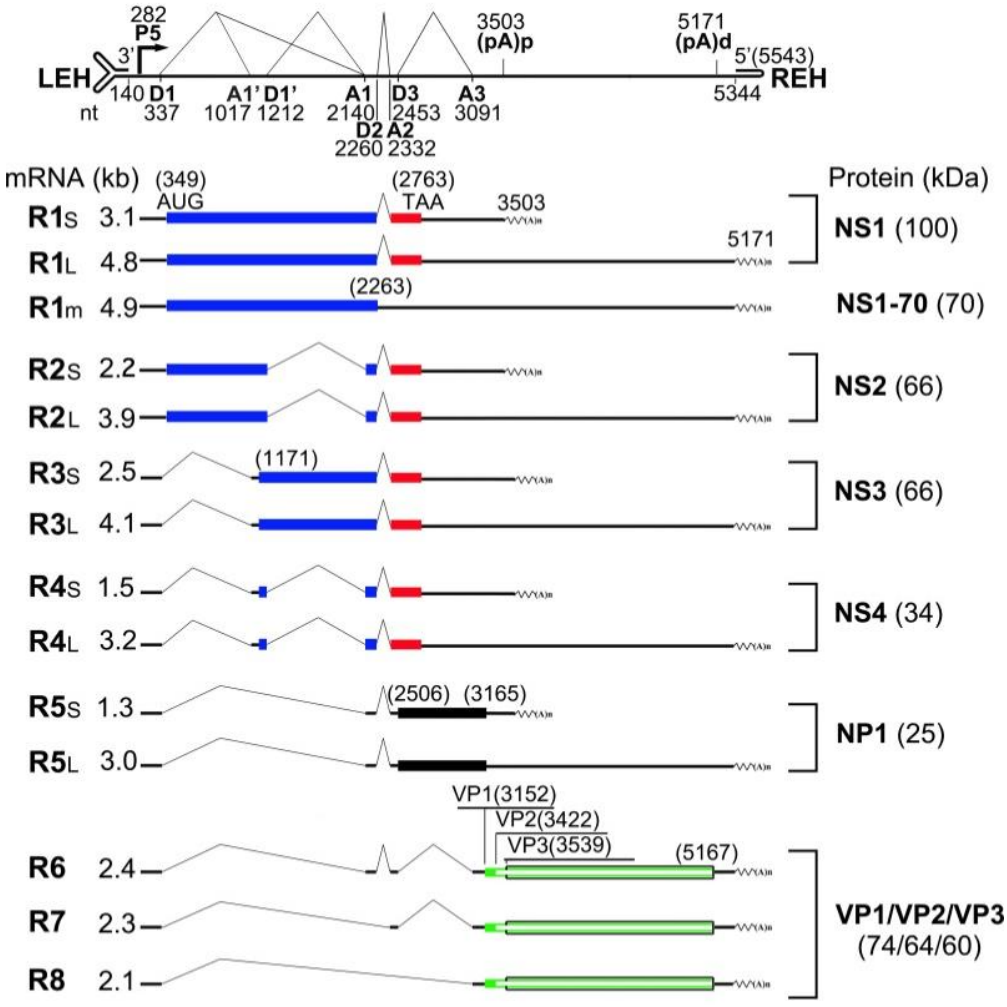


Figure 2. The transcriptional profile of HBoV1.

Shown is the wtHBoV1 genome with the different splice donor (D) and acceptor (A) sites that allow for alternative splicing of mRNA (the numbers indicate the nucleotide positions within the genome). This results in the shown mRNA splice forms R1 to R8 that express the different indicated proteins. LEH, REH = Left-end and right-end hairpin, respectively. Modified from reference 70.

Little is known about how protein expression is regulated in HBoV1 to result in the 1:1:10 ratio observed for VP1:VP2:VP3. In AAVs, which use the same protein stoichiometry, the VP pre-mRNA is produced from the p40 promoter and spliced into a minor (for VP1) and a major (for VP2/VP3) form. Then, a non-canonical start codon (ACG) is used for VP2 to limit the expression of this protein. Likewise, VP2 of HBoV1 also uses a less efficient start codon (GTG). However, a similar splicing mechanism within the HBoV1 *cap* gene was not observed. Recently, it has been shown that upper (u)ORFs residing in the 5‘ untranslated (UTR) region of the HBoV1 Cap mRNA play an important role in the regulation of alternative translation of VP proteins, a process which might involve leaky scanning of ribosomes [81].

An additional unique feature of primate BoVs as compared to the animal bocaparvoviruses is the presence of a small non-coding RNA (BocaSR) in the 3‘ UTR region [82]. Notably, the sequence and structure of this RNA is highly conserved: HBoV2-4 and GBoV share 97% sequence identity, while the sequence is more divergent in HBoV1 with only 87% sequence similarity. Moreover, BocaSR is closely related to the adenovirus viral-associated RNA I and II (VAI and VAII, ~60% sequence identity) and to the Epstein-Barr virus-encoded small RNAs 1 and 2 (~57% identity). BocaSR is also expressed at levels close to the adenoviral VAI in cells but unlike VAI, it accumulates in the nucleus at sites of viral DNA replication. Interestingly, it has been shown that this localization plays an important role in the regulation of viral DNA replication and expression of the non-structural proteins NS1-3 and NP1 but not NS4. The molecular mechanisms by which BocaSR facilitates these events are still unclear.

1.2 Parvovirus vectorization

1.2.1 Helper-independent parvovirus vectors

The cancer-specific cytopathic effect exerted by autonomous parvoviruses in cancer tissue fueled the first attempts to harness these viruses for medical applications. From the wt parvovirus (**Figure 3A**), three types of vector systems have been derived. The first are (i) *cap*-replacement vectors (**Figure 3B**) in which ~800 bp in the *cap* region are replaced by a transgene of interest. The deletion of larger sequences is also possible but results in a dramatic drop in infectious viral titers [83]. One application of this type of vectors is their use for vaccination. For example, Palmer *et al.* [84] replaced a part of the *cap* gene with the *Borrelia burgdorferi* outer surface protein A (OspA) and injected these vectors into naive mice. A single dose already resulted in high and sustained levels of specific anti-OspA antibodies. In another application, therapeutic interleukins or chemokines were incorporated to increase the antineoplastic effect of the cognate wt viruses [85, 86]. The second vector type are (ii) "gutted" vectors (**Figure 3C**) in which the complete genome comprising the whole *ns* and *cap* ORFs is deleted and replaced with a transgene of interest. These types of vectors have so far only been

used in combination with reporter genes such as luciferase and β-galactosidase to allow a transient expression of transgenes [87-89]. Finally, there are (iii) promoter-replacement vectors, in which the endogenous p4 promoter of rodent parvoviruses has been modified or replaced to add a new level of regulation. This feature was then combined with either wt viruses [51] or vectors. For example, Maxwell *et al.* [90] replaced the p4 promoter in a recombinant LuIII with a liver-specific promoter to drive the expression of a luciferase transgene (Pol II promoter in **Figure 3C**). This resulted in preferential luciferase expression in a human hepatoma cell line (HepG2). Moreover, transgene expression could be turned on or off by adding tetracycline-responsive elements (combination of binding sites with the promoters in **Figure 3**).

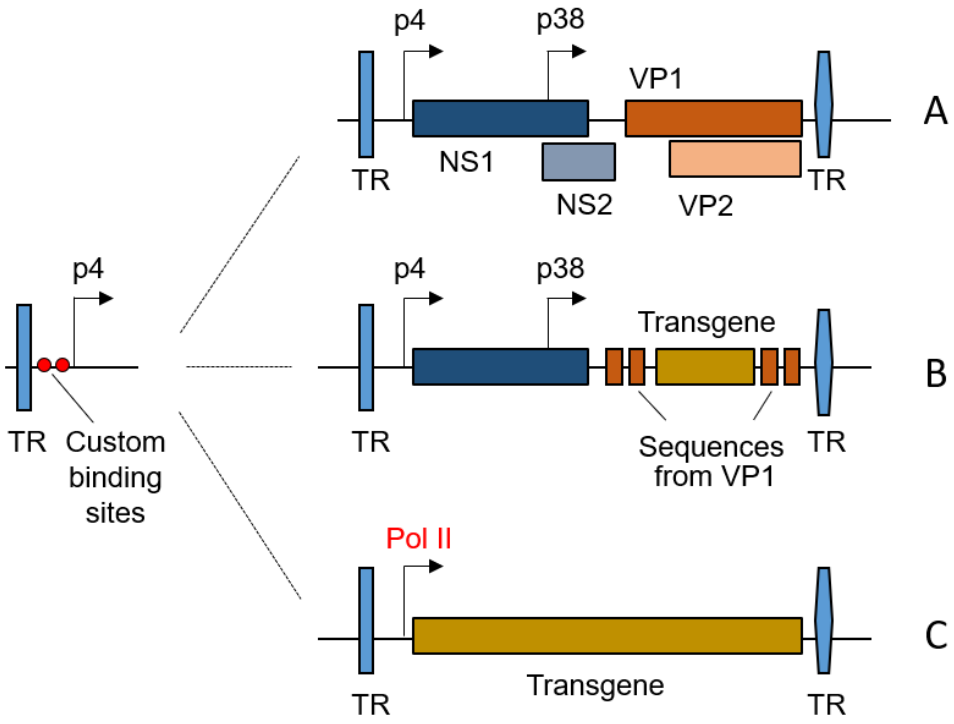


Figure 3. Genome organization of wt and recombinant rodent autonomous parvoviruses.

(A) Schematic representation of the genome of wt autonomous rodent parvoviruses that encompasses two major genes encoding non-structural (NS) and structural (VP) proteins. The terminal repeats (TR) are asymmetric and required for packaging of the genome. **(B)** *cap*-replacement vectors. A part of the *vp* ORF is replaced by a transgene of interest. **(C)** Complete genome deletion ("gutted") vectors. The *ns* and *vp* ORF are completely replaced by a transgene of interest driven by an RNA pol II promoter. Custom binding sites can be combined with A, B or C to allow cell-type-specific or inducible expression.

1.2.2 Helper-dependent parvovirus vectors

Helper-dependent parvovirus vectors have so far been exclusively based on AAV. As previously mentioned in section 1.1.1, the site-specific Rep-mediated integration of wtAAV decreases the risk of insertional mutagenesis (as compared *e.g.* to lentiviruses [91]) and thus inspired the development of AAV-based vectors for integration of transgenes [31, 32]. In these types of vectors, the Rep proteins were usually supplied *in trans* to promote the integration into the target genome [92-94]. In a study by Henckaerts and colleagues [95], such an AAV2-based

vector was used to integrate a GFP reporter into mouse embryonic stem cells. This resulted in strong transgene expression, which persisted throughout the differentiation of the cells *in vitro* and *in vivo*. One additional interesting finding in this study was a partial duplication of the target sequence at the integration site, which suggests the possibility that AAV has evolved a mechanism for targeted integration into a gene-rich area in the human DNA while avoiding site-directed mutagenesis.

Despite the promise of Rep-mediated gene targeting, the cytostatic and anti-proliferative effect of Rep protein expression [96, 97] hampers the generation of AAV vectors that express Rep *in cis*. Consequently, currently used AAV vectors are typically gutted, *i.e.*, devoid of all viral ORFs (**Figure 4**). Accordingly, the *cap* and *rep* genes are replaced by a transgene of interest and the endogenous viral promoters are exchanged with ubiquitous or tissue-specific promoters, such as the cytomegalovirus (CMV) promoter or the liver promoter 1 (LP1), respectively. The only remaining viral elements are the inverted terminal repeats (ITRs), which are strictly required for the replication of the viral genome and its packaging into the capsid. The production requirements for rAAV vectors are further discussed in section 1.2.5.1. Importantly, gutted AAV vectors lose their ability to integrate in a site-specific manner [98] and show a different integration profile than the cognate wt viruses [99, 100]. A preferential integration of rAAV into the *Rian* locus has been observed in mice and shown to increase the incidence of HCC [101, 102]. Integration at this site was especially reported in neonatal mice that received high doses of rAAV with strong promoter/enhancer elements [101, 103], which raises important considerations for vector design and time of application. Importantly, numerous long-term studies in primates showed that integration events rarely result in insertional mutagenesis [104-106] and that the majority of rAAV genomes persist episomally as concatemeric structures [107]. Thus, despite the importance of animal disease models for the optimization of vector design and proof-of-concept studies, it remains questionable how far these studies can predict the outcome of a certain strategy in humans.

The majority of rAAV plasmids used nowadays are based on AAV serotype 2, *i.e.* they carry the AAV2 ITRs. The ITRs are also the only sequence element required for release of the viral genome from the plasmid backbone during production of rAAV vectors in HEK293T cells. This process is believed to resemble the rescue of AAV from its genomic integration site in a latently infected cell [108]. Interestingly, it has been shown that the ITR sequences among the different AAV serotypes are highly conserved and that the Rep proteins are interchangeable (except for AAV5 [109]). This enabled a universal packaging of the same AAV backbone into AAV capsids from different serotypes. This process of cross-packaging is called pseudotyping and will be further discussed in section 1.2.3.

1.2.2.1 Next-generation AAV vectors

The promiscuity of AAVs represents a hurdle in the application of the natural serotypes for gene therapy, where tissue or cell specificity is of critical importance. Moreover, the high seroprevalence of AAV antibodies in the human population (especially to AAV2 [110]) further restricts their use in individuals, even in those with a low antibody titer. Thus, rational or random approaches have been applied to generate new AAV variants with unique assets for gene transfer and gene therapy. Rational approaches have been of limited success because they require a profound understanding of virus biology including cell entry and trafficking. By contrast, random methods do not require such comprehensive knowledge but rather rely on the power of selection. The latter usually starts with a library of highly diverse AAV capsids, generated through *e.g.* DNA family shuffling (DFS), random mutagenesis, peptide display or a combination thereof. The libraries are then subjected to various positive and/or negative selection pressures to enrich promising candidates [111]. In DFS, the *cap* genes of multiple AAV serotypes are first fragmented and then reassembled based on their partial homologies, resulting in libraries of chimeric AAV capsids. One example of a successful DFS approach is AAV-DJ [111], which displays an increased specificity for hepatocytes and an improved immunological profile. Another example is AAV2.5T [112], a potent transducer of pHAE, which resulted from a combination of DFS and random mutagenesis.

In AAV peptide display, small peptide sequences are incorporated into exposed loops in the AAV capsid. This often involves insertions at or close to the native receptor binding site (*e.g.* the heparin binding motif in AAV2 [113]) to ablate the natural tropism and concomitantly redirect the capsid to another cellular target. Peptide sequences either involved common motifs like RGD (for integrin binding) [114] or were completely randomized at the start of a selection campaign [113]. Peptide display resulted in highly specific AAV variants such as AAV2-ESGHGYF [115] and AAV-PHP.B [116], which selectively transduce the mouse lung and brain, respectively.

1.2.3 Pseudotyped parvoviral vectors

Pseudotyping refers to the packaging of the genome of one parvovirus into the capsid of another. The majority of rAAV vectors available today are produced by packaging of an AAV2-based genome into an AAV capsid of interest. The first report of a pseudotyped rAAV vector was by Beck *et al.* [117] in an attempt to bypass neutralizing antibodies and allow vector re-administration. In this work, an AAV2-CFTR cassette (cystic fibrosis transmembrane conductance regulator) was efficiently packaged into an AAV3 capsid using the Rep proteins from AAV2. Thereafter, it became clear that the Rep proteins from AAV2 support the packaging of AAV2-based genomes not only into AAV2 and AAV3 but also into all other AAV serotypes [118, 119]. Consequently, most of the AAV vector plasmids used nowadays for pseudotyping rely on an AAV2 background despite the fact that other AAV genotypes have also been harnessed

for vector construction [118, 120]. Interestingly, the Rep proteins displayed a high plasticity and were largely interchangeable across AAV serotypes. This was probably facilitated by the high sequence identity of the AAV ITRs, to which the Rep proteins bind and which they cleave during AAV genome replication (>95%) [119, 121]. One exception is AAV5, whose Rep proteins and ITRs are distinct from other AAV serotypes. Thus, AAV5 Rep proteins do not support packaging of AAV2-based genomes and *vice versa*, AAV2 Rep proteins fail to package AAV5-based genomes [122]. This is especially due to significant differences in the terminal resolution site (TRS) in the AAV5 ITR sequence, which its nicked by Rep78/68 during AAV replication [109].

Maxwell and co-workers were the first to perform autonomous parvovirus pseudotyping, in order to elucidate tropism determinants of rodent parvoviruses. Therefore, a recombinant LuIII genome (encoding *luciferase*) was packaged into the capsids of numerous autonomous parvoviruses such as MVM. Transduction of mouse lymphocytes and fibroblasts with the two MVM strains, MVMi and MVMp, revealed that the capsid is the sole determinant of virus tropism [123]. A similar observation was made by Spitzer *et al.* who pseudotyped a LuIII genome with the capsids of canine parvovirus (CPV) and the feline panleukopenia virus (FPV) [123]. Also in these experiments, the resulting virus tropism was determined by the virus capsid but not by the packaged genome. Still, genomic determinants might also be important for virus tropism. For example, the murine species-specificity of H1 parvovirus is determined by DNA elements in the viral genome [124].

Most autonomous parvoviruses predominantly package minus rather than plus strand genomes (~90:10), while others encapsidate both strands with equal efficiencies (*e.g.* LuIII). Pseudotyping was used to study the determinants of viral genome packaging. For example, when rLuIII genomes were pseudotyped with MVM capsids, the packaging polarity of LuIII was retained [125]. Likewise, pseudotyping a rAAV genome with a HBoV1 capsid results in equal packaging of minus and plus genomes, which is a feature of AAV but not of HBoV1 [10]. These observations hint at an intrinsic property in the viral genomes that determines packaging polarity.

Finally, pseudotyping was utilized to create new cross-genera viral vectors for therapeutic gene transfer. The underlying idea is to combine the best properties of two genera, *e.g.* the low genotoxicity or immunogenicity of AAV with the more specific tropism or enlarged packaging capacity of another parvovirus (*e.g.* B19 or HBoV1).

The ability to package a rAAV genome into the B19 capsid was first demonstrated in an attempt to simplify the study of B19 biology. In this work, Srivastava *et al.* packaged a chimeric viral genome composed of AAV2 ITRs and the whole B19 genome into an AAV2 capsid [126] (for more details, see section 1.2.4). In another study, Ponnazhagan *et al.* [127] used this unique property in an inverted approach, by packaging rAAV genomes into the B19 capsid. The

pseudotyped rAAV2/B19 viral vector very efficiently and specifically transduced erythroid progenitor cells. One limitation of these vectors was the low viral titers of 10^7 genome copies per mL (gc/mL), which in general restricted their application to small-scale *in vitro* experiments. For example, the pseudotyped vector was used to study B19 tropism and cell entry [128].

The most recent report of a cross-genera parvoviral vector was published in 2013 [10]. In this study, Yan *et al.* packaged a rAAV2 genome into the capsid of HBoV1, yielding rAAV2/HBoV1 vectors that potently transduced pHAE [10]. Moreover, an oversized (with respect to AAV capsids) ssAAV2 vector genome of up to 5.5 kb could be efficiently packaged into the HBoV1 capsid. [10]. Further increasing the attractiveness of this latest pseudotyped parvoviral vector is the high viral titers that were obtained (> 2×10^{11} gc/mL) as compared to the rAAV2/B19 vectors.

1.2.4 Chimeric parvoviral vectors

The genomes of chimeric parvoviral vectors are composed of elements belonging to more than one parvovirus, which distinguishes them from pseudotyped vectors where the genome is derived from a single viral isolate (see previous chapter). Because of their hybrid nature, these vectors usually share minimal or no homology to the packaging-helper plasmids. Thus, this strategy was commonly used to eliminate contaminating replication-competent wt virus in parvoviral vector stocks by reducing the risk of recombination between vector and helper plasmids [124].

Chimeric parvoviral vectors were also used to answer questions surrounding the parvoviral life cycle. As shortly mentioned in section 1.2.3, the first chimeric parvoviral vector was constructed by Srivastava *et al.* [126], who packaged a hybrid rAAV2-B19 genome (AAV2 ITRs flank the whole B19 genome) into an AAV2 capsid (rAAV2-B19/AAV2). This vector was able to autonomously and specifically replicate in erythroid progenitor cells. Strikingly, despite the distant phylogenetic relationship of AAV2 and B19, the hybrid genome was able to replicate in the sole presence of B19 proteins. The was only possible in the B19 target cells because of the restricted activity of the p6 promoter. Importantly, production of the hybrid vector was not restricted to the B19 target cells because the B19 genome was flanked by AAV2 ITRs, which are the only viral requirement for standard rAAV production (see section 1.2.2). In view of the strict growth requirements of wtB19 virus [129], the construction of a rAAV-based, infectious and replication-competent B19 clone that can be easily produced in standard human KB cells (and probably also in HEK293 cells) represented a major advance, which simplified the dissection of B19 biology. In a subsequent report, Wang *et al.* [130] used a different hybrid vector to assess the role of the B19 p6 promoter in B19 infectivity and cell specificity. Therefore, the p5 promoter in wtAAV2 was replaced by the B19 p6 promoter. Interestingly, the B19 promoter was sufficient to allow autonomous replication of the rAAV genome in erythroid progenitor cells [130].

Krüger *et al.* [131] followed a similar approach with the aim to specifically target and kill tumor cells. The idea here was to combine the oncolytic properties of the H1 parvovirus with the broad tissue tropism of AAV2. Therefore, the p4/p38 promoters and the *ns* gene from the H1 virus were first cloned between rAAV2 ITRs. The H1 *cap* gene driven by the p38 promoter was then replaced with the suicide gene thymidine kinase from herpes simplex virus (HSV-*tk*). This enables an inducible killing of HSV-TK-expressing cells by the addition of the prodrug ganciclovir [132].

Flanking the H1 hybrid vector with AAV ITRs allowed the packaging of the genome into an AAV2 capsid (rAAV2-H1/AAV2). Importantly, the p4 promoter, which drives the expression of the *ns* gene, has a higher activity in transformed cells owing to the binding of specific transcription factors [133]. Yet, in contrast to the rAAV2-B19/AAV2 hybrid vector, the rAAV2-H1/AAV2 vector was unable to replicate autonomously. This missing feature is, however, believed to be important for an efficient oncolytic effect. Accordingly, the vector showed a preference for transformed cells but no advantage over a CMV promoter-driven standard AAV vector.

1.2.5 Production of parvoviral vectors

The different parvoviral vectors described in the previous chapters can be produced by flanking the modified genomes by terminal repeats and supplying the missing viral proteins *in trans* from separate plasmids. For example, *cap*-replacement vectors (**Figure 3B**) would require the *trans*-complementation of the *cap* ORF, whereas gutless vectors (**Figure 3C**) devoid of all viral genes require both *cap* and *ns* genes *in trans*. Importantly, the degree of homology between the different plasmids should be as low as possible to prevent recombination events that could lead to the contamination of the vector stocks with wt viruses. As noted above, chimeric viruses and pseudotyping were used as efficient approaches to diminish this type of contamination [88, 123, 124].

In this work, we used pseudotyped AAV and BoV vectors for gene transfer into primary cells and cell lines. Thus, their production and purification will be in the focus of the next two chapters.

1.2.5.1 Production of pseudotyped rAAV vectors

In standard AAV vectors, the complete AAV genome is usually replaced by a transgene of interest, *i.e.*, they are gutted. Only the ITRs are kept as they are the sole sequences required for DNA replication and packaging of the genomes into the capsids (AAV vector in **Figure 4**). The *cap* and the *rep* genes are provided *in trans* together with the genes from adenovirus (AAV and Ad helpers in **Figure 4**, respectively).

In this work, all AAV vector backbones were based on the AAV2 genotype. Pseudotyping with different capsids was performed by simply replacing the *cap* ORF in the helper plasmid.

AAV vectors were then produced by co-transfecting the three plasmids into HEK293T cells (see method section 2.2.5.2)

Interestingly, it has been shown that reduced Rep protein expression favours the production of AAV vectors [134, 135]. This led to the development of AAV helper plasmids that express lower levels of Rep78/68 than conventional helper constructs, due to a shift of the AAV2 p5 promoter (in plasmid pWHC2 [136]) or its replacement with the weaker MMTV promoter (in plasmids pDG and pDGΔVP) [135, 137].

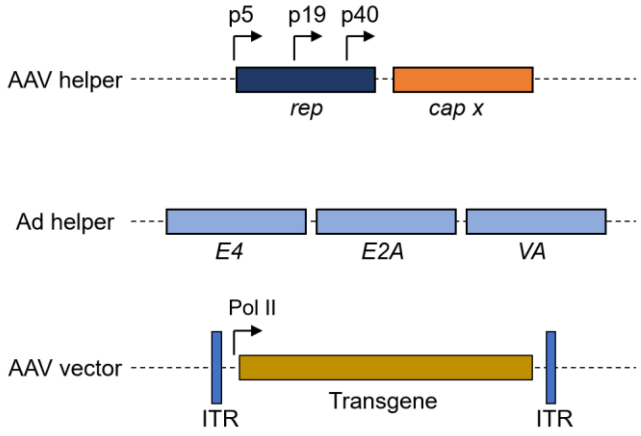


Figure 4. Recombinant AAV production system.

Shown are the three plasmids needed for AAV vector production. The plasmids are co-transfected into 70% confluent HEK293T cells. The x denotes the used AAV serotype, which can be 1-13 for naturally occurring primate AAVs.

1.2.5.2 Production of pseudotyped rAAV2/HBoV1 vectors

Pseudotyped rAAV2/HBoV1 vectors were first described by Yan *et al*. [10] in 2013. In this original report, the vectors were produced by co-transfection of four plasmids (**Figure 5**, plasmids 1 to 4). Notably, at that time, the factors strictly required for pseodotyped vector production were still unknown. Thus, all non-structural proteins of BoV and AAV2 were supplied *in trans* (**Figure 5**, plasmids 2 and 3). Moreover, an adenoviral helper plasmid was co-transfected that expresses the adenovirus genes needed for AAV production (**Figure 5**, plasmid 4). Pseudotyped vectors could be produced at ~1.25×10^{11} gc/mL, which is around 5 – 10% of cognate rAAV2/AAV2 preparations [10].

Recently, Yan and colleagues have reported an optimized rAAV2/HBoV1 production system that results in higher yields than the prototype system [138]. This system was rationally developed based on the increased understanding of parvovirus, especially bocaparvovirus biology. The authors showed that overexpression of the non-structural proteins (from a CMV promoter) negatively influences the production of viral particles. Consequently, deletion of the NS1-4 proteins resulted in a 9-fold increase in vector yields.

The NP1 protein is the only non-structural protein in HBoV1 which is strictly required for VP expression. Curiously, despite its important role, overexpression of NP1 resulted in a reduction in the viral titer [138]. Thus, along their aim to optimize the rAAV2/HBoV1 production system, the authors sought to design a HBoV1 helper that is independent of NP1. This was achieved by introducing a silent mutation into the (pA)p site that lies within the *cap* ORF [70]. The resulting bocaviral helper carrying the (pA)p mutation and lacking all NS/NP1 proteins promoted higher *cap* expression but a 3-fold lower virus production. This led the authors to conclude that the 1:1:10 stoichiometry of the VP1:VP2:VP3 proteins was disturbed and that NP1 may play a role in its regulation.

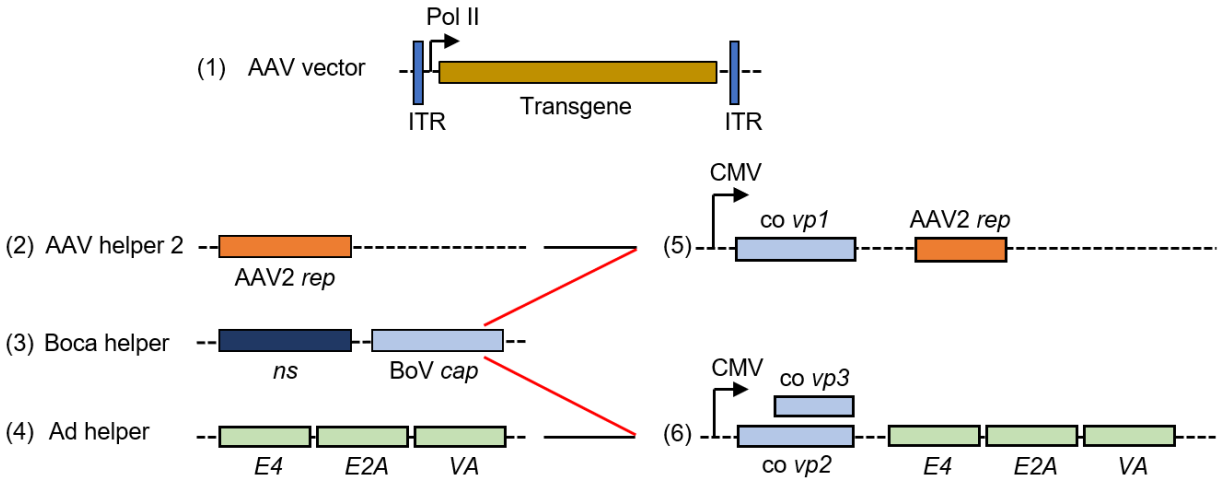


Figure 5. Pseudotyped rAAV/HBoV1 production system.

The prototype production protocol as described by Yan *et al.* in 2013 [10] involves plasmids 1 to 4. The transgene of interest (1) is expressed from a standard rAAV vector with AAV2 ITRs. The non-structural proteins required for AAV replication and packaging are supplied *in trans* from the AAV and Ad helpers (2 and 4, respectively). The HBoV1 structural and non-structural proteins are provided from the BoV helper (3). The improved rAAV/HBoV1 production system includes plasmids 1, 5 and 6 [138]. The non-structural proteins from BoV are not needed. Instead, the BoV *cap* ORF was split onto two new helpers (5 and 6; codon-optimized *vp1* to *vp3*). Moreover, the non-structural proteins required for AAV replication and packaging (*rep* and adenoviral helper genes) are included.

The silencing of the (pA)p signal made it possible, for the first time, to express the HBoV1 *cap* ORF independently and in the absence of any non-structural proteins. The major Cap protein is VP3 with 60 copies in the viral capsid. VP1 is the largest VP protein and plays an important role in the infectivity of the viral particles, because of its N-terminal phospholipase A domain and the nuclear localization sequence (NLS) [68]. The role of the VP2 protein in HBoV1 infectivity is still unknown. However, it is believed that VP2 in parvoviruses participates in receptor recognition and cell entry [68]. Consistent with this, Yan and colleagues experimentally verified that VP2 is not limiting, whereas high levels of VP1 and VP3 are important for infectious HBoV1 virus particle formation [138].

To restore the correct stoichiometry of viral protein expression, the authors mimicked the mechanism used by AAV, which is to uncouple VP1 expression from VP2/VP3. This was

achieved by juxtaposing the VP1 ORF on one plasmid with the Rep proteins from AAV2 (**Figure 5**, plasmid 5), whereas the VP2/VP3 ORFs and the adenoviral helper genes were placed on a second plasmid (**Figure 5**, plasmid 6). Moreover, the GTG start codon of VP2 and the surrounding sequence were replaced by a stretch from AAV2 that includes the Kozak consensus sequence and the start codon ACG. Akin to the original start codon GTG in HBoV1, ACG is a non-canonical start site that results in reduced VP2 protein levels. This hybrid constellation was successful and resulted in a 16-fold increase in viral production (~4.5×10^{12} gc/mL) as compared to the prototype production system.

1.3 Clinical application of parvoviral vectors

AAV vectors have many advantageous traits such as their low immunogenicity, distinct tissue tropism and high vector yields. This makes them the most attractive parvoviral vectors available to date for many pre-clinical and clinical applications. The success of AAV vectors was reflected in the licensing of the first European AAV gene therapy product, Glybera, in 2012 for the treatment of lipoprotein lipase deficiency. Since then, impressive therapeutic efficacy has been frequently reported in numerous clinical trials for treatment of ocular [139], neurological [140] and other monogenic diseases (*e.g.* alpha-1 antitrypsin deficiency [141] and hemophilia B [142]). Most recently, the FDA approved the first two AAV gene therapy products in the US, namely, Luxturna for the treatment of various forms of retinal dystrophy [143], or Zolgensma for the treatment of spinal muscular atrophy [140, 144].

Replication-competent autonomous parvoviruses showed promising results in targeting and resolving different types of cancer in mice [48, 145]. The first clinical trial involving oncolytic parvoviruses (ParvOryx 01) started in 2011 and used the murine H1 parvovirus. The study involved 18 patients and revealed a good safety profile of the virus, broad distribution and an immunostimulatory property against tumor cells [146]. Despite all these promising aspects, complete remission was not obtained. The efficacy of the treatment was now enhanced with the inclusion of immune-modulating agents [147]. This showed a promising synergistic oncolytic effect in 78% of patients with two complete responses. In 2016, a second clinical trial was initiated for the treatment of inoperable pancreatic cancer (NCT02653313).

1.4 A brief history of CRISPR discovery

In 1987, Ishino and co-workers reported a very unusual cluster of repetitive sequences near the 3' end of the *iap* gene in *E. coli*. Despite the absence of any sequence similarity to the previously identified repetitive extragenetic palindromic (REP) sequences, the authors speculated that they might share a similar function because of their repetitive nature, which is believed to contribute to mRNA stability [148, 149].

A few years later, the mysterious repetitive regions were also found in other bacteria such as *Shigella dysenteriae*, *Salmonella typhimurium* [150] and *Mycobacterium bovis* [151]. In 1993, Mojica *et al.* [152] made the first observation of a similar cluster in halophilic archaea (*Haloferax mediterranei).* Thereafter, many researchers detected similar clusters in a variety of microbes, archaea and even mitochondria, implying a high biological relevance.

Different terminologies were proposed to describe these sequences: regularly spaced repeats [152] and spacers interspersed direct repeats (SPIDR) [153]. Eventually, the naming conflict was resolved and the acronym CRISPR for clustered regularly interspaced short palindromic repeats was adopted and is still in use, although it is nowadays known that the repeats are not always palindromic. The unknown ORFs next to CRISPR were found to encode proteins that directly interact with the repeats and were hence named CRISPR-associated proteins (Cas) [154].

With the development of bioinformatic tools and the availability of broadly-accessible databases of full-length genomes, the spacer sequences that separated the highly conserved repeats could be matched to sequences from phages and mobile genetic elements. This revealed that the CRISPR locus maintains a memory of previous invaders [155] and thus might represent an adaptive immune system. However, initially, it was thought to resemble the eukaryotic RNA interference (RNAi) mechanism.

Bolotin *et al.* showed in 2005 [156] that there is a correlation between the number of spacers and the resistance to infection. Moreover, they identified a short and conserved motif next to the target sequence (the protospacer) in the invader, which was later named the protospacer-adjacent motif (PAM). This PAM sequence is important for self- and non-self-recognition that protects the organism form attacking its own genome [157]. Moreover, the PAM sequence was shown to be seminal for both, spacer acquisition and interference [158]. It is noteworthy that some CRISPR systems have evolved to function without a specific PAM sequence, by using different levels of complementarity between the CRISPR (cr)RNAs and the target sequences [159].

The first experimental evidence that CRISPR is indeed an adaptive immune system in prokaryotes was provided by Barrangou and co-workers in 2007 [160], who showed that the addition or deletion of spacer sequences into the CRISPR locus changed the response of *S. thermophilus* to infection with phages carrying the corresponding protospacer sequences. They moreover demonstrated a pivotal role of the Cas proteins in this process.

1.5 CRISPR interference in nature

Currently, there are six known CRISPR-Cas systems, which differ in their protein components and mode of action [161]. Due to the complexity of the topic, the focus here will be

on the best-characterized CRISPR types I to III. A comprehensive review of all these systems and their subtypes can be found in reference 161.

In all the described CRISPR systems, there are three levels of CRISPR interference: adaptation, expression and interference. In the adaptation phase, spacer sequences are acquired from invaders and inserted at the leader end of the CRISPR locus. The mechanism underlying this process is poorly understood and varies between the different CRISPR systems [162]. However, two universal key players are known to be involved: Cas1 and Cas2 [158]. These two proteins are believed to originate from transposon-like elements and to function similar to viral integrases and transposases in mediating the insertion of the DNA into the genome [163]. There are two types of spacer acquisition: (i) naive, when an invader has not been previously encountered [164], and (ii) primed, following a pre-exposure to the invader [165]. The latter is usually more efficient and results in insertion of additional spacers from one invader, increasing the resistance of the organism to this invader.

The CRISPR locus encodes a long primary transcript called the pre-crRNA (the precursor CRISPR RNA), which is expressed by the host RNA polymerase and is subsequently processed into short mature crRNAs [166]. This process varies across the different CRISPR systems and subtypes but in general involves endoribonucleases originating from the *cas* ORF (such as Cas6 in type I and III [167]) or from a different region (RNase III for type II [168]). Next, the crRNAs are loaded into the respective Cas proteins to form a ribonucleoprotein complex. This involves the CRISPR-associated complex for anti-viral defense (Cascade), Cas9 or the Cascade-like effectors (Csm/Cmr) for type I, II and III CRISPR systems, respectively.

In the last stage of interference, the Cas/RNA ribonucleoprotein complexes target and cleave a complementary sequence of foreign DNA. Also here, differences were observed among the CRISPR systems. For example, type I and II require the presence of the PAM sequence next to the target site, whereas type III systems use a different mechanism for self- and non-self-discrimination. Here, a part of the repeat sequences, which are usually only present in the host CRISPR locus, are included in the crRNA and result in inhibition of the endonuclease activity if the complementarity to the target site is high enough. Moreover, the protein responsible for cleaving the target DNA is different. In the type II system, only the Cas9 protein is required for interference and DNA cleavage together with an additional small RNA called the *trans*-activating CRISPR RNA (tracr) RNA [168]. By contrast, in type I CRISPR systems, different proteins in the Cascade complex are involved in crRNA binding, DNA targeting and PAM recognition. The endonuclease Cas3 is believed to be recruited later for a processive cleavage of target DNA [169, 170]. Type III CRISPR systems are even more complex and less well understood, involving numerous proteins (such as Cas10) and an ability to target both, DNA and RNA [171].

Apart from its function in bacterial immunity, the CRISPR system has proved to be involved in many other regulatory functions in bacteria such as virulence [172], DNA repair and gene expression regulation [173].

1.6 Anti-CRISPR mechanisms

Bacteriophages are viruses that infect bacterial cells, which can result in lysis and death of the host organism. Phages are the most abundant forms of life on the planet. It has been estimated that in seawater, there are at least 10 bacteriophages for each bacterial cell [174, 175]. Therefore, bacteria have evolved different ways to defend themselves against bacteriophage infection including: (i) the change or masking of surface receptors [176], (ii) Sie (superinfection exclusion) and RM (restriction modification) systems that block phage DNA injection or destroy the incoming viral DNA through self/non-self discrimination, respectively [177], and (iii) the use of the previously described CRISPR-Cas system.

The wide distribution of CRISPR in ~50% of bacteria puts a high selection pressure on phages to develop anti-CRISPR mechanisms to breach these barriers [178]. The same applies to viruses infecting eukaryotic cells, which evolved RNA silencing suppressors (RSS) to counteract the host's RNAi machinery and hence promote infection [179]. Thus, it came as no surprise when the first anti-CRISPR type I proteins (Acr F1-5) were discovered in phages infecting *Pseudomonas aeruginosa* [180]*.* Later studies demonstrated the existence of Acr proteins in a wide range of gram-positive- or -negative bacteria [181]. The expression of these proteins from lysogenic phages (*i.e.* integrated in the bacterial genome) that harbor targets for the bacterial CRISPR system became an essential requirement for the cell's and the phage's survival [182].

The mechanisms by which Acr proteins inhibit CRISPR activity are diverse but usually involve a physical interaction between the Acr protein and subunits of the Cas complex or the endonuclease component [183]. Consequently, Acr proteins perturb different steps of CRISPR interference from target binding to target degradation. Moreover, some types of interactions were also shown to impede new spacer acquisition [184].

Most of the Acr proteins discovered so far are directed against the CRISPR I and II systems. Two of these proteins (AcrIIA2 and AcrIIA4) inhibit the Cas9 protein derived from *Streptococcus pyogenes* (SpCas9) [185]. The ability to heterologously express these proteins in eukaryotic cells along with the SpCas9 (see section 1.7) has been harnessed to allow for spatial and temporal control of Cas9 activity [186].

1.7 Mammalian genome editing using CRISPR/Cas9

In 2011, Sapranauskas *et al.* showed that CRISPR regions can be transferred between bacteria as completely functional units, which demonstrated their autonomous nature [187]. Two years later, the machinery was expressed in eukaryotic cells and used for genome editing [188, 189].

Of all the CRISPR systems described so far, the type II systems are the simplest, as they rely on only one protein (the Cas9) for DNA targeting and cleavage [191]. Moreover, and in contrast to the other CRISPR types, the target DNA has been shown to be cleaved in a specific manner, 3 bp upstream of the PAM sequence [191]. Both characteristics facilitated the development of this CRISPR type as a tool for genome editing. The ability to fuse the crRNA and tracrRNA into one chimeric gRNA further increased the simplicity and applicability of the system [190]. Consequently, Jinek *et al.* showed that only two components were required to efficiently target plasmid DNA *in vitro* (**Figure 6**): (i) the endonuclease Cas9 and (ii) the chimeric gRNA composed of the crRNA and the tracrRNA. The crRNA confers the specificity by directing the Cas9 to a complementary target sequence, whereas the tracrRNA plays a role in recruiting and binding of the Cas9. Once the gRNA/Cas9 complex finds its target sequence, the Cas9 protein uses two endonuclease homologous domains HNH and RuvC to cleave the target and non-target strand, respectively [191, 192]. Thereby, any DNA sequence can be chosen as target for Cas9 as long as it harbors a compatible PAM sequence at its 3' end.

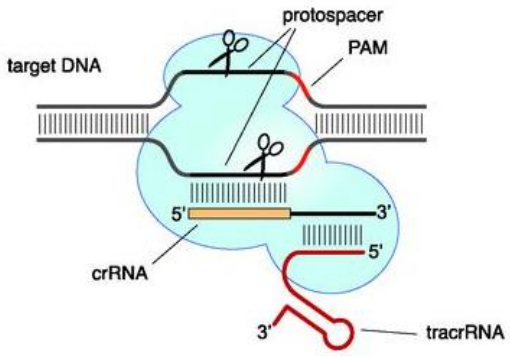


Figure 6. Mammalian genome editing using the CRISPR type II system.

The Cas9 protein is directed to its target DNA sequence (the protospacer) using a chimeric gRNA, which is composed of the crRNA and a tracrRNA. The 5'-PAM sequence is crucial for target recognition. Two endonuclease domains, HNH and RuvC, cleave the target and non-target strand, respectively. (Adopted from reference 190)

When Cas9 cuts its defined target sequence, the result is a double-stranded DNA break. The cell attempts to repair its DNA by using either homologous recombination (HR) or non-homologous end-joining (NHEJ). HR is template-dependent and results in high-fidelity repair, whereas NHEJ is error-prone and can eventually result in gene knockout [193]. Usually, NHEJ is the dominant repair pathway in eukaryotic cells, especially when mutations are biallelic and no repair template is supplied *in trans*.

The desired outcome of gene editing always depends on the application. For example, Cas9-mediated knockout allows loss-of-function studies while the ability of gRNA multiplexing

facilitates parallel genome-wide knock-out screens, as demonstrated by reports studying cellular signalling [194] and tumor growth [195]. On the other hand, HR is beneficial when the aim is to introduce defined changes in target DNA sequences (knock-in). However, the intrinsic frequency of HR in eukaryotic cells is very low, which poses an obstacle for harnessing its full potential. Therefore, several strategies have been implemented to increase the incidence of HR events by *e.g*, using inhibitors of the NHEJ pathway [196] or by physically positioning the donor DNA close to the cut site [197].

Noteworthy, a catalytically inactive form of Cas9, the (d)ead Cas9, has been constructed and used alone or in combination with transcription activators/repressors to introduce transcriptional or epigenetic changes at various sites in the genome [198, 199].

1.7.1 Cas9 delivery for *in vivo* genome editing

The simplicity and flexibility of the CRISPR/Cas9 system led to its widespread use for *in vitro* genome editing. Soon, the promise of the technology for *in vivo* gene editing and medical applications was recognized. The challenge here was to find a suitable carrier for the delivery of all CRISPR components *in vivo*. In first attempts, hydrodynamic injection of plasmid DNA [200, 201] was used to deliver the Cas9, a gRNA and a repair template. In a study by Yin *et al.* [200], low rates of *Fah* mutation correction were obtained in mouse hepatocytes. Importantly, an expansion of the corrected cells resulted in alleviation of the pathogenic phenotype of hereditary tyrosinemia.

The process of hydrodynamic injection is now widely applied for delivery of proteins, oligonucleotides and small molecules [202]. However, due to the invasiveness of the method, it is only applicable in rodents and not considered for gene delivery in humans [203]. Instead, other methods are currently under investigation, which show great promise for human gene therapy: (i) non-viral carriers including lipid nanoparticles (LNP) [204], gold nanoparticles [205] and polymeric vectors such as polyethylamine (PEI) [206], as well as (ii) the previously mentioned viral vectors including adenoviruses [207], lentiviruses [208] and AAVs [12, 209, 210] (the latter is discussed in detail in section 1.2.2).

Non-viral methods are attractive because of their low immunogenicity and safety advantage. However, the specificity and the resulting editing efficiencies were lower than those obtained so far with viral vectors. Therefore, most applications nowadays focus on viral vectors, especially rAAVs that show a high transduction efficiency, but low immunogenicity as compared to adeno- and lentiviruses. As mentioned in section 1, one hurdle is the small packaging capacity of rAAVs of only 4.7 kb, which makes the packaging of both CRISPR requirements - the Cas9 and the gRNA - challenging. This especially applies to the first identified SpCas9 variant with a size of 4.1 kb. Still, all-in-one vectors that carry both parts on one vector were possible when minimized forms of promoters were used [209]. High expression levels are, however, required to achieve phenotypic changes or clinical benefit. Thus, to

overcome the AAV packaging limit, several approaches have been applied: (i) packaging of Cas9 and the gRNA expression cassette into two separate rAAV capsids [11], or (ii) the identification and use of smaller Cas9 orthologs including those from *Streptococcus thermophilus* [188], *Neisseria meningitidis* [211] and *Staphylococcus aureus* (SaCas9) [12]. The latter is now most commonly used because of its comparable cutting efficiencies to SpCas9 [14, 212]. Moreover, (iii) split Cas9 variants were developed in which the N-terminal part of the Cas9 is expressed from one AAV capsid and the C-terminal part from another [213]. After co-transduction, the Cas9 is reconstituted in the cell and results in editing efficiencies that can be comparable to a full-length Cas9, depending on the target cells and specific splicing strategy. The split Cas9 approach is discussed in more detail in the next section due to its high relevance for this work.

1.7.2 Split Cas9 approaches for overcoming AAV size restrictions

Several previous studies have attempted to overcome the packaging limit of rAAVs by distributing larger transgenes over two or even three separate rAAV vectors (called split AAVs) [214]. When a cell is co-transduced by all the different rAAVs, the transgene is reconstituted on the DNA level via recombination or *trans*-splicing, which consequently restores transgene expression [214]. Despite many promising results, transgene expression remained less than that observed with full-length versions of the gene. Nevertheless, this certainly represented a promising approach to deliver transgenes that surpass or exceed the AAV packaging capacity.

With a length of 3.1 to 4.2 kb, the Cas9 ORF belongs to the larger transgenes, which, when packaged into rAAVs, only leave little space for regulatory elements or additional cassettes (*e.g.*, fluorescent reporters or gRNA cassettes). Therefore, many research groups harnessed the split AAV approach to distribute Cas9 on two or three different rAAVs. In more detail, one can distinguish three different strategies, the first being (i) physical methods. Here, the SpCas9 was split in two parts that assembled only in the presence of a suitable and modified gRNA [215]. The second strategy is (ii) rapamycin-inducible assembly, whereby the N- and C-termini of Cas9 are fused to FK506-binding protein 12 (FKBP) or FKBP rapamycin-binding (FRB) domains, respectively. Addition of rapamycin triggers the dimerization of the two domains and hence reconstitutes full Cas9 expression [216]. Finally, (iii) the split Cas9 parts were fused to intein moieties. When the two split parts are expressed in one cell, the inteins recognize each other, splice themselves out and join the two protein halves to recover the full-length Cas9.

As seen before with other split transgenes, the split Cas9 variants were usually less efficient than a full-length Cas9 [216, 217]. However, the lower level of activity might still be sufficient for distinct applications that are only possible with a split rather than a full-length Cas9. One example is the combination of Cas9 with a variety of biologically active modules for transcriptional perturbation or inducible expression. These CRISPR on/off systems usually

exceed the packaging limit of standard AAV vectors but were readily combinable with a split Cas9 [216]. Moreover, tight spatial control of Cas9 expression should be possible using different cell-type specific promoters for the split parts, which would restrict Cas9 activity to a defined subset of cells.

1.8 Clinical trials for genome editing

Over the last few years, CRISPR has been used successfully for *ex vivo* gene therapy applications. One of the most intriguing examples is the so-called CAR-T therapy for children with B-cell acute lymphoblastic leukemia and for adults with advanced lymphomas [218, 219]. CAR-T stands for chimeric antigen receptor T cells, which are cells taken from patients and engineered *ex vivo* to express a modified receptor (CAR) that targets the T cells to the cancer cells. First-generation CAR-T cells were created using lentiviruses or the *Sleeping Beauty* transposon system. Both integrate exogenous DNA sequences randomly into the genomic DNA and hence allow for a stable expression of the chimeric receptor [220]. Recently, the FDA authorized Kymriah, a CAR-T therapy directed against CD19+ tumor cells [221].

To enhance the safety and to broaden the applicability of CAR-T technology, several laboratories around the world are now using the CRISPR/Cas9 system to knock out the endogenous T-cell receptor and knock in an exogenous CAR cDNA at the same time. The resulting T-cells are allogeneic, which reduces the risk of immune rejection (graft-versus-host disease) [222, 223]. Another improvement to CAR-T cell therapy was the combination with blocking antibodies against surface molecules such as PD-1, whose expression causes an immune exhaustion and loss of functional activity of the T-cells [224]. The clinical benefit of these strategies is currently being studied in multiple clinical trials (NCT03399448, NCT03166878, NCT03398967 and NCT02650999).

Another *ex vivo* clinical trial was announced in 2018 by CRISPR Therapeutics and Vertex Pharmaceuticals in Canada and Europe (NCT03655678). The medication called CTX001 is a CRISPR-based modification of human hematopoietic stem and progenitor cells (hHSPCs). This therapy is proposed for individuals with β-thalassemia or sickle-cell disease, who suffer from abnormalities in hemoglobin production. The molecular target of the approach is a gene called *BCL11A*, which represses the production of fetal hemoglobin. The quintessential hope is that when this repression is lifted, enough hemoglobin will be produced to alleviate the disease phenotype.

Editas Medicine and Allergan plan the first *in vivo* CRISPR gene-editing trial (EDIT-101) in 2019 [225]. Here, the gene editing machinery will be packaged into a rAAV vector and delivered by subretinal injection to directly reach the photoreceptor cells. CRISPR should then target and

correct the point mutation IVS26 in the gene *CEP290,* which is believed to play a role in Leber's congenital amaurosis type 10.

Seven years after its adaptation to mammalian cells, the CRISPR technology broke a record for entering the first clinical trials faster than any other technology before. Whilst the true clinical value has yet to be proven, there is every reason to be excited about this rapidly emerging and broadly applicable technology.

1.9 Aim

The gene therapy field has a growing demand for new vector systems that efficiently and specifically deliver any cargo of interest. Recently, a novel lung-specific pseudotyped rAAV2/HBoV1 system has been described that relies on the use of the HBoV1 capsid and a rAAV2 genome. The advantages of this system are manifold, including the larger packaging capacity of the BoV capsid as compared to AAV, the compatibility with all available rAAV vectors and the high vector yields in contrast to other previously reported chimeric systems. Intrigued by this advance in vector technology, we focused in the first part of this work on BoVs as potential tools for delivery of transgenes that exceed the packaging capacity of AAVs. To this end, we aimed to: (i) establish and optimize the pseudotyped rAAV/BoV production system in our lab, (ii) extend the genome size of both ss and scAAVs to examine the upper HBoV1 packaging capacity, and (iii) study the impact of targeted tyrosine-to-phenylalanine mutations in the viral capsid on the transduction with rAAV2/HBoV1 vectors.

Subsequently, we aimed to expand the repertoire of bocaviral vectors by exploiting four newly discovered primate BoVs, namely, HBoV2-4 and GBoV. Importantly, despite the high homology to HBoV1, these viruses were only sporadically detected in the airways and hence believed to have a different tropism. Thus, the first goal was to *de novo* assemble the *cap* ORFs of the four primate BoVs based on published sequences and to create hybrid helper plasmids for efficient pseudotyping with rAAV genomes. The success of the first goal laid the foundation for the second step, which was to test the cell tropism and thereby prove the functionality of the viral vectors. The final goal was to extend the DNA family shuffling approach, commonly applied to AAV vectors, to all five BoV capsids in order to create chimeric libraries that should serve as valuable tools for both, vector development and dissection of the biology of BoV infection.

In the second major part of this work, we aimed to increase the safety of rAAV vectors in gene therapy through two different approaches. The first concept was to integrate a “safety switch” into the rAAV vector, which, when needed, can inactivate the expression of the encoded transgene from the vector. To this end, we exploited the CRISPR/Cas9 system to target the vector DNA, leading to a permanent knock-out of gene expression and - in the best

scenario - a complete degradation of the target vector. Here, two components are needed, the first being a gRNA complementary to a site in the transgene. To create an activatable kill-switch, the gRNA expression cassette (Pol III promoter and gRNA) is cloned into the vector harboring the transgene. The second component is the Cas9 endonuclease that is supplied *in trans* when the activation of the kill-switch is required and then loads the gRNA and triggers vector degradation. Another goal in this part of the work was to optimize Cas9 delivery in terms of expression amplitude and duration. Therefore, we aimed to construct and validate a self-inactivating (SIN) rAAV-CRISPR system based on a split Cas9 approach in which the two parts of the split Cas9 (N- and C-parts) are positioned on two distinct rAAV vectors. To create a SIN circuit, an anti-Cas9 gRNA targeting the N-terminus of the Cas9 is added to the vector harboring the C-terminal part. This overcomes a previous hurdle in the construction of SIN CRISPR vectors, which is the mutation and inactivation of Cas9 during vector production. The smaller size of the split parts offers additional advantages, such as (i) a flexibility in choice of promoters and regulatory elements, which is crucial towards optimization of *in vivo* Cas9 expression. Moreover, (ii) it allows packaging of Cas9 and the gRNA into scAAVs, which result in higher transgene expression than ssAAVs [226].

In summary, this doctoral work addresses and unites two central aspects of viral vector development, which are capsid and genome engineering. Importantly, these two parts are complementary and readily combinable. For example, the larger packaging capacity of the BoV capsid can be harnessed in future work to optimize the SIN vectors by (i) making use of promoters, regulatory elements and functional domains which allow higher or tissue-restricted expression and which usually do not fit together with the full-length Cas9 into standard rAAVs, and by (ii) including all required elements for DNA editing and repair, *i.e.*, the Cas9, the gRNAs and a homology repair template. Beyond CRISPR applications, we envision that different pseudotyped bocaviral vectors will expand the range of primary cells and tissues, which are refractory to efficient rAAV transduction, but which could be targeted using our new generation natural or synthetic BoV vectors and thus also become amenable to therapeutic intervention.

2 MATERIALS AND METHODS

2.1 Materials

2.1.1 Bacterial strains

Strain	Description	Source
E. coli ccdB Survival^{TM} T1^{R}	Chemocompetent	Thermo Fisher Scientific (Massachusetts, USA)
E. coli MAX Efficiency DH5α^{TM}	Chemocompetent	Thermo Fisher Scientific (Massachusetts, USA)
SURE2 supercompetent cells	Chemocompetent	Agilent Technologies (California, USA)
E. coli MegaX DH10B^{TM} T1^{R}	Electrocompetent	Thermo Fisher Scientific (Massachusetts, USA)
E. cloni 10G elite/supreme	Electrocompetent	Lucigen (Madison, USA)

2.1.2 Cell lines

Cell line	Origin	Description
HEK293T	*H. sapiens*	Human embryonic kidney cells expressing the SV40 large T-antigen [227]
Panc-I	*H. sapiens*	Cell line established from a human carcinoma of the exocrine pancreas [228]
HeLa	*H. sapiens*	Cell line derived from a cervical carcinoma [229]
Huh7	*H. sapiens*	Human hepatoma cell line [230]
RAW 264.7	*M. musculus*	Macrophage mouse cell line derived from a tumor induced by Abelson murine leukemia virus [231]
LX-2	*H. sapiens*	Human hepatic stellate cell line established by spontaneous immortalization under low serum conditions (1% FBS) [232]
MCF-7	*H. sapiens*	Human cell line derived from a breast carcinoma [233]
T84	*H. sapiens*	Human cell line derived from immortalized HAE of a CF genotype [234]
Cufi-8	*H. sapiens*	Cell line derived from a colorectal carcinoma [235]

2.1.3 Primary cells

Cell type	Origin	Description
Saphenous vein endothelial cells	*H. sapiens*	Purchased from PromoCell, Heidelberg, Germany #C-12231
Skeletal muscle cells	*H. sapiens*	Purchased from PromoCell, Heidelberg, Germany #C-12530
Pulmonary fibroblasts	*H. sapiens*	Purchased from PromoCell, Heidelberg, Germany #C-12360
Cardiac myocytes	*H. sapiens*	Purchased from PromoCell, Heidelberg, Germany #C-12810
Primary human hepatocytes (pHep)	*H. sapiens*	Purchased from Cytes Biotechnologies, Barcelona, Spain
pHAE	*H. sapiens*	Provided by Marc Schneider, Thoraxklinik, Heidelberg University Hospital, Heidelberg, Germany
Primary T cells	*H. sapiens*	Provided by Manuela Nickl, Dept. of Infectious Diseases/Virology, Heidelberg University Hospital, Heidelberg, Germany
Primary peripheral blood mononuclear cells (PBMCs)	*H. sapiens*	Provided by David Bejarano, Dept. of Infectious Diseases/Virology, Heidelberg University Hospital, Heidelberg, Germany
Primary macrophages	*H. sapiens*	Provided by David Bejarano and Kathleen Börner, Dept. of Infectious Diseases/Virology, Heidelberg University Hospital, Heidelberg, Germany
Primary gut organoids	*H. sapiens*	Provided by Megan Stanifer, Dept. of Infectious Diseases/Virology, Heidelberg University Hospital, Heidelberg, Germany
Primary lung organoids	*H. sapiens*	Provided by Jens Puschhof, Hubrecht Institute, Utrecht, the Netherlands

2.1.4 Viruses

Virus type	Description	Source
rAAV2	Recombinant adeno-associated virus type 2 encoding Firefly and *Renilla* luciferase and/or different CRISPR components	Produced in this study

rAAV8	Recombinant adeno-associated virus type 8 encoding different transgenes: Firefly and *Renilla* luciferase, the KS reporter or Cre recombinase	Produced in this study
rAAV2/HBoV1	Pseudotyped virus composed of a HBoV1 capsid and a rAAV2 genome encoding *Gaussia* luciferase or YFP	Produced in this study
rAAV2/HBoV2	Pseudotyped virus composed of a HBoV2 capsid and a rAAV2 genome encoding *Gaussia* luciferase or YFP	Produced in this study
rAAV2/HBoV3	Pseudotyped virus composed of a HBoV3 capsid and a rAAV2 genome encoding *Gaussia* luciferase or YFP	Produced in this study
rAAV2/HBoV4	Pseudotyped virus composed of a HBoV4 capsid and a rAAV2 genome encoding *Gaussia* luciferase or YFP	Produced in this study
rAAV2/GBoV	Pseudotyped virus composed of a GBoV capsid and a rAAV2 genome encoding *Gaussia* luciferase or YFP	Produced in this study

2.1.5 Chemicals and reagents

	Product	**Company**
Chemicals	Iodixanol (Optiprep™)	Progen (Heidelberg, Germany)
	Agarose	Biozym Scientific GmbH (Hessisch Oldendorf, Germany)
	10× TGS (tris/glycine/SDS)	Bio-Rad (Hercules, USA)
	Polyethylenimine (PEI)	Polysciences Europe GmbH (Eppelheim, Germany)
	Albumin fraction V (BSA)	Roth (Karlsruhe, Germany)
	Calcium chloride ($CaCl_2$)	Roth (Karlsruhe, Germany)
	Magnesium chloride ($MgCl_2$)	Applichem (Darmstadt, Germany)
	Magnesium sulphate ($MgSO_4$)	MERCK (Darmstadt, Germany)
	Cesium chloride (CsCl)	Roth (Karlsruhe, Germany)
	Sodium chloride (NaCl)	GRÜSSING GmbH (Filsum, Germany)

Potassium acetate (KAc)	GRÜSSING GmbH (Filsum, Germany)
Potassium chloride (KCl)	GRÜSSING GmbH (Filsum, Germany)
di-Potassium hydrogen phosphate (K_2HPO_4)	Applichem (Darmstadt, Germany)
Potassium dihydrogen phosphate (KH_2PO_4)	MERCK (Darmstadt, Germany)
Monopotassium phosphate (KPO_4)	MERCK (Darmstadt, Germany)
Sodium hydroxide (NaOH)	MERCK (Darmstadt, Germany)
Hydrochloric acid (HCl)	MERCK (Darmstadt, Germany)
Ethidium bromide	Roth (Karlsruhe, Deutschland)
Milk powder	Roth (Karlsruhe, Germany)
EDTA	GRÜSSING GmbH (Filsum, Germany)
EGTA	Roth (Karlsruhe, Germany)
1,4-dithiothreitol (DTT)	Roth (Karlsruhe, Germany)
Bromophenol blue	Chroma Technology Corp. (Bellows Falls, USA)
Glycerol	VWR chemicals (Fenenay-sous-Bais, France)
Phenol red	MERCK (Darmstadt, Germany)
Tween 20	Roth (Karlsruhe, Germany)
Tris-EDTA (TE) solution	Thermo Fisher Scientific (Waltham, USA)
Tris (hydroxymethyl)-aminomethan (TRIS)	Roth (Karlsruhe, Germany)
β-mercaptoethanol	Roth (Karlsruhe, Germany)
Hoechst 3000	Dianova (Hamburg, Germany)
Dimethyl sulfoxide (DMSO)	MERCK (Darmstadt, Germany)
Dodecylsulfate-Na-salt pellets (SDS)	SERVA Electrophoresis GmbH (Heidelberg, Germany)
Triton X-100	MERCK (Darmstadt, Germany)
Sodium azide	Roth (Karlsruhe, Germany)
IGEPAL® CA-630 (NP-40)	MERCK (Darmstadt, Germany)
Paraformaldehyde (PFA)	MERCK (Darmstadt, Germany)

Glucose	MERCK (Darmstadt, Germany)
MOPS	SERVA Electrophoresis GmbH (Heidelberg, Germany)
Lipofectamine 2000	Life Technologies GmbH (Paisley, UK)
Isopropanol	MERCK (Darmstadt, Germany)
Ethanol absolute	MERCK (Darmstadt, Germany)
Methanol	MERCK (Darmstadt, Germany)
Nuclease-free water	Ambion, Life Technologies GmbH (Paisley, UK)
PonceauS	MERCK (Darmstadt, Germany)
Acetic acid	VWR chemicals (Fenenay-sous-Bais, France)
Trisodium citrate	Applichem (Darmstadt, Germany)
Ficoll (Type 400)	MERCK (Darmstadt, Germany)
Maleic acid	Roth (Karlsruhe, Germany)
Sodium deoxycholate	MERCK (Darmstadt, Germany)
cOmplete™, EDTA-free Protease Inhibitor Cocktail	Roche (Penzberg, Germany)
Bacto™ yeast	BD (Franklin Lakes, USA)
Bacto™ tryptone	BD (Franklin Lakes, USA)
Bacto™ agar	BD (Franklin Lakes, USA)
Rotiophorese® Gel 40	Roth (Karlsruhe, Germany)
UltraPure™ TEMED	Thermo Fisher Scientific (Waltham, USA)
Ammonium persulfate (APS)	GRÜSSING GmbH (Filsum, Germany)
Coelenterazine	p.j.k GmbH (Kleinblittersdorf, Germany)
D-luciferin	p.j.k GmbH (Kleinblittersdorf, Germany)
Doxorubicin	Santa Cruz Biotechnology (Dallas, USA)
Calpain inhibitor 1 ALLN	G-Biosciences (St. Louis, USA)
Cultrex® growth factor reduced BME type 2	Trevigen (Gaithersburg, USA)

Kits	QIAquick Gel Extraction Kit	QIAGEN (Hilden, Germany)
	QIAquick PCR Purification Kit	QIAGEN (Hilden, Germany)
	QIAprep Spin Miniprep Kit	QIAGEN (Hilden, Germany)
	DNeasy Blood & Tissue Kit	QIAGEN (Hilden, Germany)
	DNA Clean & Concentrator	Zymo research (Irvine, USA)
	Dual-Luciferase® Reporter Assay System	Promega (Madison, USA)
	PureYield™ Plasmid Midiprep System	Promega (Madison, USA)
	NucleoBond® Xtra Midi / Maxi	Macherey-Nagel (Hoerdt, France)
	SensiMix™ II Probe Kit	Bioline (London, UK)
	Western Lightning® PLUS-ECL	PerkinElmer (Waltham, USA)
	DIG High Prime DNA Labeling and Detection Starter Kit II	Roche (Basel, Switzerland)
	Qubit dsDNA BR Assay Kit	Thermo Fisher Scientific (Waltham, USA)
	Pierce BCA Protein Assay Kit	Thermo Fisher Scientific (Waltham, USA)
	Fixable Aqua Dead Cell Stain Kit	Thermo Fisher Scientific (Waltham, USA)
Enzymes	Restriction enzymes	NEB (Ipswich, USA) / Fermentas (St. Leon-Rot, Germany)
	Benzonase	MERCK (Darmstadt, Germany)
	DNaseI	Roche (Basel, Switzerland)
	OneTaq® 2× Master Mix with Standard Buffer	NEB (Ipswich, USA)
	Phusion Hot Start II DNA Polymerase	Thermo Fisher Scientific (Waltham, USA)
	Antarctic Phosphatase	NEB (Ipswich, USA)
	RNase A	QIAGEN (Hilden, Germany)
	Proteinase K	Roche (Basel, Switzerland)
	T4 DNA ligase	NEB (Ipswich, USA)
Nucleotides	dNTPs (dATP, dCTP, dGTP, dTTP)	NEB (Ipswich, USA)

Standard markers	100 bp DNA ladder	Thermo Fisher Scientific (Waltham, USA)
	1 kb Plus DNA ladder	Thermo Fisher Scientific (Waltham, USA)
	PageRuler™ Plus Prestained Protein Ladder	Fermentas (St. Leon-Rot, Germany)
	DNA Molecular Weight Marker VII, DIG-labeled	MERCK (Darmstadt, Germany)

2.1.6 Equipment

Device	Company
Centrifuges and rotors	
Allegra X-12R centrifuge	Beckman Coulter (Brea, USA)
Avanti J-26 XP centrifuge	Beckman Coulter (Brea, USA)
Benchtop centrifuge 5415R	Eppendorf (Hamburg, Germany)
Optima L-90K ultracentrifuge	Beckman Coulter (Brea, USA)
Neuation iFuge BL08VT	Neuation Technologies (Gandhinagar, India)
Fixed angle type 70 Ti rotor	Beckman Coulter (Brea, USA)
Fixed angle type 70.1 Ti rotor	Beckman Coulter (Brea, USA)
JA-10 rotor	Beckman Coulter (Brea, USA)
Beckman tube sealer	Beckman Coulter (Brea, USA)
Beckman rotor NVT 65	Beckman Coulter (Brea, USA)
Electrophoresis systems	
PowerPac basic/HV/HC	Bio-Rad (Hercules, USA)
Mini-PROTEAN Tetra cell chamber	Bio-Rad (Hercules, USA)
Mini-Sub cell GT	Bio-Rad (Hercules, USA)
Sub-Cell GT	Bio-Rad (Hercules, USA)
Western and Southern blotting	
Trans-Blot SD semi-dry transfer cell	Bio-Rad (Hercules, USA)
PeqLab PerfectBlot hybridization oven	VWR (Radnor, USA)
Film developing cassettes	Dr. Goos-Suprema (Heidelberg, Germany)
Tube roller TRM-V	NeoLab (Heidelberg, Germany)
Tube rotator	VWR (Radnor, USA)
DNA/Protein detection	
UV-transilluminator	Biostep GmbH (Jahnsdorf, Germany)
Gel Doc XR	Bio-Rad (Hercules, USA)
Mitsubishi P93D	Mitsubishi Electric (Cypress, USA)

ChemoCam (ECL Imager)	INTAS Science Imaging Instruments (Göttingen, Germany)
X-OMAT 2000 processor (film developer)	KODAK (Rochester, USA)
PCR	
FlexCycler	Analytik Jena (Jena, Germany)
C1000 Touch thermal cycler	Bio-Rad (Hercules, USA)
Corbett RG 6000	QIAGEN (Hilden, Germany)
Vapo protect	Eppendorf (Hamburg, Germany)
Luciferase assays	
GloMax 96 microplate luminometer	Promega (Madison, USA)
Sterile cell culture	
HERA cell 150 incubator	Thermo Fisher Scientific (Waltham, USA)
HERA safe sterile work bench	Thermo Fisher Scientific (Waltham, USA)
Countess automated cell counter	Thermo Fisher Scientific (Waltham, USA)
Bacterial incubators	
Heraeus function line incubator	Thermo Fisher Scientific (Waltham, USA)
Shaking incubator (Multitron)	INFORS HT (Basel, Switzerland)
Microscopes	
CKX 419F	OLYMPUS (Hamburg, Germany)
U-RPL-T	OLYMPUS (Hamburg, Germany)
Spectrophotometers	
NanoVue spectrophotometer	Thermo Fisher Scientific (Waltham, USA)
NanoDrop 2000	Thermo Fisher Scientific (Waltham, USA)
BioPhotometer plus	Eppendorf (Hamburg, Germany)
Qubit fluorometer	Thermo Fisher Scientific (Waltham, USA)
TECAN Infinite M200	Tecan Group (Männedorf, Switzerland)
Mixing blocks and magnetic stirrers	
Rotilabo magnetic stirrer with heating	Roth (Karlsruhe, Germany)
Mixing block MB-102	Bioer Technology (Hangzhou, China)
MSH Basic - magnetic stirrer with stainless steel heating plate	IKA Laboratory Equipment (Staufen, Germany)
Other applications	
Shaker DOS-10 L	neoLab (Heidelberg, Germany)
Sonorex ultrasonic bath	Bandelin (Berlin, Germany)
Vacuum pump	Promega (Madison, USA)

Gene Pulser Xcell	Bio-Rad (Hercules, USA)
FACSVerse	BD (Franklin Lakes, USA)
Vortex Genie 2	Scientific Industries (Bohemia, USA)
Water bath TW12	Julabo Labortechnik (Seelbach, Germany)
Microwave	Sharp Electronics (Hamburg, Germany)
Weighing scale	KERN & SOHN GmbH (Balingen, Germany)
Bunsen burner	Carl Friedrich Usbeck KG (Radevormwald, Germany)
Accujet pro	BrandTech Scientific (Essex, UK)
pH meter PB-11	Sartorius (Göttingen, Germany)
Pipettes	Gilson (Middleton, Germany) / Eppendorf (Hamburg, Germany)
Refractometer	Exacta Optech (San Prospero, Italy)
Bio-Dot® SF microfiltration apparatus	Bio-Rad (Hercules, USA)

2.1.7 Materials

Material	Description	Company
Pipet tips	2.5 / 10 / 100 / 200 / 1,000 (µL)	Sarstedt (Nümbrecht, Germany) / Greiner Bio-One (Kremsmünster, Austria) / Kisker (Steinfurt, Germany)
PCR tubes	0.2 mL 8-Strip	STARLAB (Hamburg, Germany)
Microcentrifuge tubes	0.5 / 1.5 / 2.0 / 5 (mL)	SARSTEDT (Nümbrecht, Germany), Eppendorf (Hamburg, Germany)
Tubes	15 / 50 (mL)	Greiner Bio-One (Kremsmünster), Austria), BD (Franklin Lakes, USA)
Amicon Ultra-15	Centrifugal filter units (100,000 NMWL)	MERCK (Darmstadt, Germany)
Centrifuge tubes	500 mL capacity	Corning (New York, USA)
Petri dishes	15 cm	Greiner Bio-One (Kremsmünster), Austria)
Cell culture dishes	15 cm	Nunc, Thermo Fisher Scientific (Waltham, USA)
Cell culture flasks	75 / 175 (cm^2)	Greiner Bio-One (Kremsmünster, Austria)
Cell culture plates	6 / 24 / 48 / 96 (wells)	Greiner Bio-One (Kremsmünster, Austria)

Lumitrac microplate	96 well, white	Greiner Bio-One (Kremsmünster, Austria)
ThinCert cell culture inserts	24 well plates, 0.4 µm pore size	Greiner Bio-One (Kremsmünster, Austria)
Sterile filters	0.22 µm pore size	Greiner Bio-One (Kremsmünster, Austria)
Steritop filter	0.22 µm pore size	MERCK (Darmstadt, Germany)
Serological pipettes	5 / 10 / 25 (mL)	Greiner Bio-One (Kremsmünster, Austria)
Countess cell counting chamber slides	-	Thermo Fisher Scientific (Waltham, USA)
Coster reagent reservoir	50 mL	Corning (New York, USA)
Cell lifter	-	Corning (New York, USA)
Glass culture tubes	-	DWK life sciences (Wertheim, Germany)
Beckman OptiSeal tubes	26×77 mm 16×67 mm	Beckman Coulter (Brea, USA)
Beckman Quick-Seal centrifugation tubes	14×89 mm 25×89 mm	Beckman Coulter (Brea, USA)
Erlenmeyer flasks	250 / 500 / 1,000 / 2,000 (mL)	Simax (Křížová, Czech Republic)
Cuvettes	Polystyrene	Sarstedt (Nümbrecht, Germany)
Electroporation cuvettes	25×1 mm gap	Peqlab (Erlangen, Germany)
VacConnectors	-	QIAGEN (Hilden, Germany)
Pasteur capillary pipettes	230 mm	NeoLab (Heidelberg, Germany)
BD plastipak syringes	3 / 5 / 10 (mL)	BD (Franklin Lakes, US)
BD microfine syringes	300 µL	BD (Franklin Lakes, US)
Microlance canules	21G 0.8×40 mm, 19G 1.1×40 mm	BD (Franklin Lakes, US)
Glass hybridization bottles	-	Wheaton DWK life sciences (Wertheim, Germany)
Slide-A-Lyzer dialysis cassette	MWCO 20,000	Thermo Fisher Scientific (Waltham, USA)

Whatman paper	3 mm	Whatman (Maidstone, UK)
Mini-PROTEAN precast gels	15 / 50 (µL) capacity	Bio-Rad (Hercules, USA)
Nitrocellulose membrane	-	Whatman (Maidstone, UK)
Nylon membrane	Positively charged	GE Healthcare (Chicago, USA)
X-ray films	Amersham Hyperfilm ECL	GE Healthcare (Chicago, USA)

2.1.7.1 Buffers and solutions

Name	Composition
Buffers for virus production and purification	
PEI	0.323 g Polyethylenimine (25 kDa) in 1 L ddH_2O autoclaved and subjected to 3 freeze-thaw cycles
Benzonase buffer (pH 8.5)	50 mM Tris / HCl (pH 8.0), 150 mM NaCl, 2 mM $MgCl_2$
10 mM Tris-HCl	Prepared from 1 M Tris-HCl stock (Thermo Fisher Scientific, Waltham, USA)
PBS-MK	1× PBS, 1 mM $MgCl_2$, 2.5 mM KCl
PBS-MK-NaCl	1× PBS, 1 mM $MgCl_2$, 2.5 mM KCl, 1 M NaCl
Phenol red solution	0.5% (w/v) in ddH_2O
Iodixanol (Optiprep) stock	60% iodixanol solution (Progen, Heidelberg, Germany)
Iodixanol (15%)	12 mL 60% iodixanol and 36 mL 1 M PBS-MK-NaCl buffer
Iodixanol (25%)	20 mL 60% iodixanol, 28 mL 1× PBS-MK buffer, phenol red (for red color)
Iodixanol (40%)	30 mL 60% iodixanol and 15 mL 1× PBS-MK buffer
Iodixanol (60%)	60% iodixanol plus 2.5 µl/mL phenol red
10% DOC	100 g sodium deoxycholate in 1 L H_2O
Buffers for mini-preps of plasmid DNA	
P1 buffer (pH 8.0)	50 mM Tris/HCl (pH 8.0), 10 mM EDTA, 100 µg/mL RNase A

P2 buffer	200 mM NaOH, 1% SDS
P3 buffer (pH 5.1)	2.8 M KAc
Buffers for alkaline gel electrophoresis	
Gel casting buffer (pH 7.5)	30 mM NaCl, 2 mM EDTA
Alkaline electrophoresis buffer	30 mM NaOH, 2 mM EDTA
Alkaline gel-loading buffer	300 mM NaOH, 6 mM EDTA, 18% (w/v) Ficoll (Type 400), 0.25% (w/v) Xylene cyanol
Buffers for Southern blotting	
20× SSC (pH 7.0)	3 M NaCl, 300 mM trisodium citrate (in ddH_2O)
Depurination buffer	0.2 M HCl
Gel-denaturation buffer	1.5 M NaCl, 0.5 M NaOH
Gel-neutralization buffer (pH 7.5)	3 M NaCl, 0.5 M Tris HCl
Maleic acid buffer (pH 7.5)	0.1 M Maleic acid, 0.15 M NaCl
Washing buffer	Maleic acid buffer, 0.3% (v/v) Tween 20
Detection buffer (pH 9.5)	0.1 M Tris-HCl, 0.1 M NaCl
Stringency wash-buffer I	2× SSC, 0.1% (v/v) SDS
Stringency wash-buffer II	0.5× SSC, 0.1% (v/v) SDS
Buffers for protein Isolation, SDS-PAGE and Western blotting	
RIPA buffer	50 mM NaCl, 25 mM Tris-HCl (pH 8.0), 0.5% NP40 0.1% (v/v) SDS plus protease inhibitor prior to use
Stacking gel buffer (pH 8.8)	1.5 M Tris-HCl, 0.4% (v/v) SDS
Running gel buffer (pH 6.8)	0.5 M Tris-HCl, 0.4% (v/v) SDS
Protein sample buffer (2×)	2 mM EDTA, 100 mM Tris-HCl (pH 8.0) 4% SDS, 20% glycerol, 10% β-mercaptoethanol, 0.02% Bromophenol blue
Running buffer	1× TGS (Bio-Rad, Hercules, USA)
Transfer buffer	1× TGS, 20% methanol
10× TBS	0.25 M Tris-HCl, 1.25 M NaCl

Washing buffer (TBST)	1× TBS, 0.05% (v/v) Tween 20
Buffers for preparation of chemocompetent bacteria	
TFBI (pH 5.8)	30 mM KAc, 80 mM $MgCl_2$, 25 mM KCl, 16 mM $CaCl_2$, 13% glycerol
TFBII (pH 7.0)	10 mM MOPS, 75 mM $CaCl_2$, 2.5 mM KCl, 13% glycerol
Buffers for cultivation of bacteria	
LB medium (pH 7.0)	1% (w/v) Bacto™ tryptone, 0.5% (w/v) Bacto™ yeast extract, 1% (w/v) NaCl
SOB medium (pH 7.0)	2% (w/v) Bacto™ tryptone, 0.5% (w/v) Bacto™ yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$
SOC medium (pH 7.0)	2% (w/v) Bacto™ tryptone, 0.5% (w/v) Bacto™ yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 20 mM Glucose
Solution for bacterial glycerol stocks	50% (v/v) in ddH_2O, autoclaved
Buffers for agarose gel electrophoresis	
50× TAE (pH 8.3)	2 M Tris Base, 50 mM EDTA, 1 M acetic acid
Gel casting and running buffer	1× TAE
10× DNA loading dye	50 mM Tris pH 7.6, 60% glycerol, 0.25% (w/v) bromophenol blue
6× Purple loading dye	NEB (Ipswich, USA)
Buffers for Hirt extract	
Proteinase K solution	20 mg/mL stock solution in ddH_2O
Hirt extraction buffer	10 mM Tris-HCl (pH 8.0), 10 mM EDTA, 1% SDS
Buffers for luciferase assays	
Firefly luciferase buffer	25 mM Glycylglycine, 15 mM KPO_4 buffer (pH 7.8), 15 mM $MgSO_4$, 4 mM EGTA Freshly added before use: 1 mM DTT, 2 mM ATP, 70 µM Luciferin
Renilla/*Gaussia* luciferase buffer	1.1 M NaCl, 2.2 mM, Na_2EDTA, 0.22 M K_xPO (mixture of KH_2PO_4 und K_2HPO_4), 0.44 mg/mL BSA, 1.3 mM NaN_3

2.1.7.2 Cell culture media and additives

Product	Company
1× DPBS 0.25% Trypsin / EDTA Dulbecco's modified eagle medium (DMEM) GlutaMAX™ high Glucose (4.5 g/l) without sodium pyruvate Fetal bovine serum gold (FBS) Penicillin / Streptomycin (P/S) RPMI 1640 medium with GlutaMAX™ supplement	Thermo Fisher Scientific (Waltham, USA)
PneumaCult ALI 10× Supplement PneumaCult ALI Basal Medium Heparin Solution 0.2% PneumaCult ALI Maintenance Supplement 100× Hydrocortisone 200× ROCK Inhibitor	Stemcell technologies (Vancouver, Kanada)
Phytohemagglutinin (PHA)	Merck (Darmstadt, Germany)
Interleukin-2	Biomol (Hamburg, Germany)
Fibroblast Growth Medium 2 Supplement Mix / Fibroblast Growth Medium 2 Myocyte Growth Medium Supplement Mix / Myocyte Growth Medium Endothelial Cell Growth Medium Supplement Mix / Endothelial Cell Growth Medium Skeletal Muscle Cell Growth Medium Supplement Mix / Skeletal Muscle Cell Growth Medium	PromoCell (Heidelberg, Germany)
Maintenance media for primary human hepatocytes (Williams medium, P/S, ITS+, GlutaMAX™, Dexametason 10 nM in DMSO)	Cytes Biotechnologies (Barcelona, Spain)

2.1.7.3 Oligonucleotides and Gene blocks

All oligonucleotides used in this work were synthesized by Sigma Aldrich (now Merck KGaA, Darmstadt, Germany) or Integrated DNA Technologies (Coralville, USA). Probes used for qPCR analysis and BoV gene blocks were synthesized by Integrated DNA Technologies.

2.1.7.4 Software

Software	Reference
MUSCLE	(Chojnacki *et al*., 2017) [236]
Serial cloner 2.6	Freeware by Franck Perez (2013) http://serialbasics.free.fr/Serial_Cloner.html
ApE v2.0.55	Freeware by M. Wayne Davis (2018) http://jorgensen.biology.utah.edu/wayned/ape/
ImageJ	(Schneider *et al.* 2012) [237]
MEGAX	(Kumar *et al*. 2018) [238]
Quantity One	Bio-Rad (Hercules, USA)
Flowing software v2.5	Freeware by Perttu Terho (2013) http://flowingsoftware.btk.fi/
Salanto v3	(Schürmann *et al*. 2013) [239]

2.2 Methods

2.2.1 Cloning procedures

2.2.1.1 Construction of HBoV1 tyrosine-phenylalanine mutants

All the tyrosine *cap* mutants generated in this work are based on a previously reported HBoV1 helper plasmid (pCMVNS1*HBoV1) [70] encoding the complete HBoV1 genome (except for the terminal repeats). Point mutations in the *cap* ORF were generated by overlap extension PCR (2.2.3.4). For each mutant, two fragments (5' and 3') were amplified by regular PCR using the external (#1 and #2) and overlapping primers (#3 to #14) listed in **Table 1** and the primer combinations shown in **Table 2**. Next, the two fragments served as templates in a second PCR to generate full-length products using only the external primers #1 and #2 that contain BstBI and NotI restrictions sites, respectively.

Table 1. Primers used in OE-PCR for cloning of HBoV1 tyrosine mutants.

No.	Oligo Name	Sequence (5'- 3')
#1	HBoV1_BstbI_fwd	CACTGC**TTCGAA**GACCTCA
#2	HBoV1_NotI/EagI_rev	ATGA**GCGGCCGC**TCTAGAT
#3	HBoV1_Y147F_fwd	GACACAACATTCAACAATGACCTCA
#4	HBoV1_Y147F_rev	TGAGGTCATTGTTGAATGTTGTGTC
#5	HBoV1_Y274F_fwd	GAAGAGTTCAGTTCATAAGACAAAACG
#6	HBoV1_Y247F_rev	CGTTTTGTCTTATGAACTGAACTCTTC
#7	HBoV1_Y355F_fwd	CTAACCTAGAATTCAAACTTCAGTGGTAC

#8	HBoV1_Y355F_rev	GTACCACTGAAGTTTGAATTCTAGGTTAG
#9	HBoV1_Y394F_fwd	AAACCAAACCACATTCAATCTAGTG
#10	HBoV1_Y394F_rev	CACTAGATTGAATGTGGTTTGGTTT
#11	HBoV1_Y466F_fwd	TGCAGACTCATTCCTAAACATATACTGTAC
#12	HBoV1_Y466F_rev	GTACAGTATATGTTTAGGAATGAGTCTGCA
#13	HBoV1_Y528F_fwd	AGCCCACGTCATTCGATCAGT
#14	HBoV1_Y528F_rev	CACTGATCGAATGACGTGG

***Bold**: restriction sites.*

The PCR reactions were loaded onto a 1% agarose gel and the correctly sized fragments were extracted from the gel using the QIAquick Gel Extraction Kit (Qiagen) following the manufacturer's instructions. Next, the purified fragments and the vector (pCMVNS1*HBoV1) were digested with BstBI / NotI O/N (as described in section 2.2.3.5). The digested vector was loaded onto an agarose gel, the required fragment was cut out and extracted from the agarose as described above. By contrast, the digested PCR products were purified over columns using the Zymo DNA clean & concentrator kit according to the supplied protocol.

Digested fragments were ligated to the vector as described in section 2.2.3.6 to yield the HBoV1 *cap* mutants listed in **Table 2**. All sequences were verified by Sanger sequencing (section 2.2.3.1)

Table 2. List of HBoV1 tyrosine to phenylalanine mutants.

No.	Mutant	PCR fragment	Template	Forward primer	Reverse primer
#2261	pCMVNS*HBoV1 Y147F	5‘	#2224	#1	#4
		3‘	#2224	#3	#2
#2262	pCMVNS*HBoV1 Y274F	5‘	#2224	#1	#6
		3‘	#2224	#5	#2
#2263	pCMVNS*HBoV1 Y355F	5’	#2224	#1	#8
		3’	#2224	#7	#2
#2264	pCMVNS*HBoV1 Y394F	5’	#2224	#1	#10
		3’	#2224	#9	#2
#2265	pCMVNS*HBoV1 Y466F	5’	#2224	#1	#12
		3’	#2224	#11	#2
#2266	pCMVNS*HBoV1 Y528F	5’	#2224	#1	#14
		3’	#2224	#13	#2
#2267	pCMVNS*HBoV1 Y147F + Y394F	5’	#2261	#1	#10

		3’	#2261	#9	#2
#2268	pCMVNS*HBoV1 Y147F + Y528F	5’	#2261	#1	#14
		3’	#2261	#13	#2
#2269	pCMVNS*HBoV1 Y274F + Y394F	5’	#2262	#1	#10
		3’	#2262	#9	#2
#2270	pCMVNS*HBoV1 Y274F + Y528F	5’	#2262	#1	#14
		3’	#2262	#13	#2
#2271	pCMVNS*HBoV1 Y394F + Y528F	5’	#2264	#1	#14
		3’	#2264	#13	#2

*Numbers refer to the internal AG Grimm library. Primer sequences are listed in **Table 1**.*

2.2.1.2 Construction of BoV helper plasmids

To allow for seamless cloning of the different HBoV2-4 and GBoV *cap* ORFs, an intermediate acceptor plasmid for cloning was first generated by deleting the entire 2 kb HBoV1 *cap* ORF in the HBoV1 helper plasmid (pCMVNS1*HBoV1) and replacing it with two BsmBI sites, which were positioned in an inverted orientation to each other (starting at nucleotide position 646, *np1* numbering). To this end, an OE-PCR was performed using the two flanking primers #15 / #18 and the two overlapping primers #16 / #17 (all primers are listed in **Table 3**). The resulting 1918 bp fragment was cloned into pCMVNS1*HBoV1 using HindIII and NotI restriction sites (also present in primer #15 or #18, respectively). This resulted in the *cap* acceptor plasmid pCMVNS*ΔVP-2×BsmBI.

Each BoV *cap* ORF (GenBank IDs: HBoV2 FJ170278, HBoV3 EU918736, HBoV4 FJ973561, GBoV HM145750) was ordered as two gene blocks (each ~1,000 bp) from IDT. For HBoV3 and HBoV4, one PCR reaction was performed to amplify each gene block, using primers #19 to #22 (HBoV3) and #23 to #26 (HBoV4). For the first and second gene block of HBoV2 and GBoV, respectively, no PCR product could be obtained possibly due to secondary structures in the sequence. This was solved by performing three separate PCR reactions for each bocaviral isolate. Accordingly, gene block I of GBoV was amplified using primers #27 and #28, while gene block II was amplified in two separate reactions using primers #29 to #32. Likewise, gene block I of HBoV2 was amplified in two separate reactions using primers #33 to #36, and gene block II was amplified using primers #37 and #38. All PCR reactions were gel-purified using the QIAquick Gel Extraction Kit (Qiagen) following the manufacturer’s instructions. For each bocaviral *cap* ORF, the fragments were assembled and cloned via a Golden Gate reaction into an empty pBSII-KS (+) plasmid (#48) using NotI and ClaI restriction sites. Correct assembly and integrity of the resulting 2004 to 2016 bp long *cap* ORFs were validated by Sanger sequencing (section 2.2.3.1). Next, the complete *cap* ORFs were amplified using primers #39 to #46 with BsmbI restriction sites. This type IIS restriction enzyme

cleaves outside of its recognition sequence and thereby allows a seamless subcloning into pCMVNS*ΔVP-2×BsmBI using Golden Gate cloning (see **Figure 7** for a schematic representation). This last step resulted in the final helper plasmids pCMVNS*HBoV2-4 and pCMVNS*GBoV (**Table 4**, plasmids #2272 to #2275).

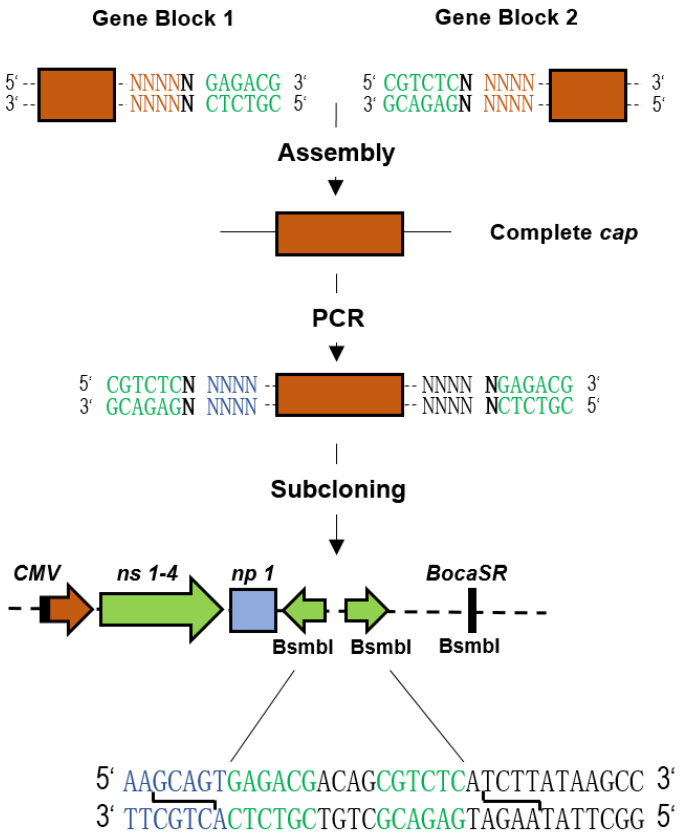


Figure 7. Golden Gate strategy used for cloning of BoV *cap* sequences.

Gene blocks (GB) 1 and 2 were PCR-amplified with primers including the BsmBI recognition sites (green): 5' CGTCTCN 3'. The enzyme cuts outside of its recognition sequence and results in a sticky end with a 4 bp overhang. The overhangs generated for GB 1 and 2 are compatible and thus allow a directional assembly of the two fragments into a full-length *cap* ORF using standard ligation or Golden Gate cloning. In a next step, the *cap* ORFs were PCR-amplified again and cloned likewise into the final acceptor plasmid pCMVNS*ΔVP-2×BsmBI. A closer view of the acceptor site with the two inverted BsmBI sites is shown.

Table 3. Primers used for Golden Gate cloning of BoV *cap* sequences.

No.	Oligo Name	Primer sequence (5'- 3')
#15	F1_HIND III	GACAATA**AAGCTT**TACAGCTTTTG
#16	R1_2xBsmbI	CAAAAAAGAGGCTTATAAGA**TGAG** **ACG**CTGT**CGTCTCA**CTGCTTCCAT GCTTTCAGC
#17	F2_2xBsmbI	TGCTGAAAGCATGGAAGCAG**TGAG** **ACG**ACAG**CGTCTCA**TCTTATAAGC CTCTTTTTTGCTTCTGC

#18	R2_NotI	ATGA**GCGGCCGC**TCTAGATGTA
#19	Boca3_partI_notI_fwd	ATTAG**GCGGCCGC**ATGCCTCCAAT TAAAAGGCAAC
#20	Boca3_partI_bsmbI_rev	GATC**CGTCTCG**CTTCCATTTCTGC AAGTTCATG
#21	Boca3_part2_BsmBI fwd	GATC**CGTCTCG**GAAGACTCCAATG CAGTAGAAAAAGCA
#22	Boca3_part2_claI_rev	GATC**ATCGAT**TTACAACACTTTATT GATGTTTGTTTTAACTGG
#23	Boca4_partI_notI_fwd	GATC**GCGGCCGC**ATGCCTCCAATT AAACGC
#24	Boca4_partI_bsmbI_rev	GATC**CGTCTCG**CTTCCATTTCAGC AAGTTCATG
#25	Boca4_part2_BsmbI_fwd	GATC**CGTCTCG**GAAGACTCAAATG CTGTAGAAAAAGCA
#26	Boca4_part2_claI_rev	GATC**ATCGAT**TTACAACACTTTATT GATGTTTGTTTTAACTGGAAAG
#27	BocaGo_partI_notI_fwd	GATC**GCGGCCGC**ATGCCTCCAATT AAAAGGCA
#28	BocaGo_partI_BsmbI_rev	GATC**CGTCTCG**TTCCATCAAGATC GGCAAGC
#29	BocaGo_part2_BsmbI_fwd	GATC**CGTCTCG**GGAACTACTGCTG GAGGAACTGC
#30	JBocaGo_part2.1_BsmbI_rev	GATC**CGTCTCG**CTCTGTAGAGGAG TTGGTCTCTAAGC
#31	Goboca_part2.2_BsmbI_fwd	GATC**CGTCTCG**AGAGGAAACCAAA CAACATAC
#32	BocaGo_part2_claI_rev	GATC**ATCGAT**TTACAACACTTTATT GATGTTTGTTTTTACAGGCATA
#33	Boca2_partI_notI_Fwd	ATTAG**GCGGCCGC**ATGCCTCCAAT TAAACGC
#34	Boca2_sub1_BsmBI rev	GATC**CGTCTCG**CCAG ATGGTTGTTGGTCTTG
#35	Boca2_sub2_BsmbI_fwd	GATC**CGTCTCG**CTGG CTCCATGGAGGAGCGAGG
#36	Boca2_partI_BsmbI_rev	GATC**CGTCTCG**CGTCTTCCATTTC AGCCAGT

#37	Boca2_part2_BsmbI_Fwd	GATC**CGTCTCG**GACGCAAATGCTG TAGAAAAAGCTATAGC
#38	Boca2_part2_ClaI_rev	GATC**ATCGAT**TTACAACACTTTATT GATGTTTGTTTTGA
#39	JF_boca4_BsmbI_fwd	GTTA**CGTCTCT**GCAGATGCCTCCA ATTAAACGC
#40	JF_boca4_BsmbI_rev	GTTA**CGTCTCT**AAGATTACAACACT TTATTGATGTTTGTTTTAAC
#41	JF_boca3_BsmbI_fwd	GTTA**CGTCTCT**GCAGATGCCTCCA ATTAAAAGGCAAC
#42	JF_boca3_BsmbI_rev	GTTA**CGTCTCT**AAGATTACAACACT TTATTGATGTTTGTTTTAAC
#43	JF_Goboca_BsmbI_fwd	GTTA**CGTCTCT**GCAGATGCCTCCA ATTAAAAGGCA
#44	JF_Goboca_bsmbI_rev	GTTA**CGTCTCT**AAGATTACAACACT TTATTGATGTTTGTTTTTAC
#45	JF_boca2_BsmbI_fwd	GTTA**CGTCTCT**GCAGATGCCTCCA ATTAAACGC
#46	JF_boca2_BsmbI_rev	GTTA**CGTCTCT**AAGATTACAACACT TTATTGATGTTTGTTTTG

Bold: *restriction sites.*

For cloning of the infectious variants of the BoV helper plasmids (*i.e.*, #2277 to #2280, **Table 4**) the respective *cap* ORFs were subcloned from plasmids (#2272 to #2275) into pIFHBoV1 [72] using a BstBI and BsmBI double digest. The integrity of the BoV terminal repeats was ensured by performing a digest with SalI and XhoI.

Table 4. BoV helper and replication-competent plasmids.

No.	Plasmid Name	Description	Origin
#48	[a] pBSII-KS (+)	Standard cloning vector with a multiple cloning site	Agilent, Waldbronn, Germany
#2224	pCMVNS*HBoV1	Helper plasmid with *ns* 3-4, *np* 1 (both HBoV1 origin) and the HBoV1 *cap* ORFs. Used for production of chimeric rAAV2/HBoV1 particles	Ref. 70
#2225	[a] pCMVNS*ΔVP-2×BsmBI	Acceptor plasmid derived from #2224 for cloning of BoV gene blocks. The HBoV1	This study

		cap ORF was deleted and replaced by two inverted BsmBI sites	
#2272	pCMVNS*HBoV2	Same as #2224 but with a HBoV2 *cap* ORF. Used for production of chimeric rAAV2/HBoV2 particles	This study
#2273	pCMVNS*HBoV3	Same as #2224 but with a HBoV3 *cap* ORF. Used for production of chimeric rAAV2/HBoV3 particles	This study
#2274	pCMVNS*HBoV4	Same as #2224 but with a HBoV4 *cap* ORF. Used for production of chimeric rAAV2/HBoV4 particles	This study
#2275	pCMVNS*GBoV	Same as #2224 but with a GBoV *cap* ORF. Used for production of chimeric rAAV2/GBoV particles	This study
#2276	pIFHBoV1	Plasmid harbouring the *ns 1-4*, *np 1* (both HBoV1 origin) and the HBoV1 *cap* ORF. HBoV1 LEH and REH flank the viral genome. Used for production of infectious HBoV1 particles	Ref. 72
#2277	pIFHBoV2	Same as #2276 but with a HBoV2 *cap* ORF. Used for production of infectious HBoV2 particles	This study
#2278	pIFHBoV3	Same as #2276 but with a HBoV3 *cap* ORF. Used for production of infectious HBoV3 particles	This study
#2279	pIFHBoV4	Same as #2276 but with a HBoV4 *cap* ORF. Used for production of infectious HBoV4 particles	This study
#2280	pIFGBoV	Same as #2276 but with a GBoV *cap* ORF. Used for production of infectious GBoV particles	This study

Numbers refer to the AG Grimm database. [a] intermediate plasmids used for cloning.

2.2.1.3 Cloning of BoV shuffling acceptors

The wt pTR-BoV shuffling acceptor (#2281) was derived from pIFHBoV1 by first deleting the *cap* ORF and inserting a BsmBI site at nucleotide position 592 (*np* 1 numbering). To this end, an OE-PCR with overlapping primers #47 / #49 and flanking primers #48 / #15 was

performed as described in section 2.2.3.4 (see **Table 5** for a list of primers). The resulting fragment was then gel-purified and cloned into pIFHBoV1 using HindIII / BsmBI restriction sites. The wt shuffled-capsid acceptor harbours the terminal repeats (TR) and thus should theoretically allow the autonomous packaging of a BoV library.

Next, the complete NSΔVP-1×BsmBI sequence was PCR-amplified using primers #50 / #51. The PCR product was digested with EagI / PacI and cloned into plasmid #1914 - a derivative of pSSV9 [240], *i.e.*, a conventional AAV vector construct harboring two AAV2 ITRs. This resulted in the second shuffled-capsid acceptor pAAVNSΔVP-1×BsmBI (#2282). This plasmid can replicate as a conventional AAV vector, *i.e.*, in the presence of the adenovirus helper genes and the Rep proteins from AAV.

For the cloning of shuffled BoV *cap* sequences into #2281 or #2282, primers #52 and #48 were used to PCR-amplify the ~2 kb fragments. Next, O/N digest of the inserts and vectors with BsmBI allowed for a seamless cloning of the fragments.

Table 5. Primers used for cloning of BoV shuffling acceptors.

No.	Oligo Name	Sequence (5‘- 3‘)
#47	Acceptor shuffling fwd1	GACAGAAG**AGAGACG**CGGAAAGTGAAGGGTGACTG
#48	HBoV1 BsmBI Rev1	GGCTAGGTT**CGAGACG**GTAAC
#49	Acceptor shuffling Rev2	CCCTTCACTTTCCG**CGTCTCT**CTTCTGTCTGTGAGG AAACA
#50	HBoV1 wt EagI fwd	GATTC**CGGCCG**CCACAAGGAGGAGTGGTTATA
#51	HBoV1 wt PacI rev	AGTTC**TTAATTAA**ATAAGCAAACAAAACAGCTCC
#52	Fwd_chimBoV_BsmbI	AATGA**CGTCTCC**GAAGCAGACGAGATAACTGACGA GG
#53	HBoV1_cap_PacI_Fwd	GATGC **TTAATTAA**ATGCCTCCAATTAAGAGACA
#54	HBoV1_cap_AscI_Rev	AGTTC**GGCGCGCC**TTACAACACTTTATTGATGTT

Bold: *restriction sites.*

To test whether a hybrid system composed only of AAV *rep* and BoV *cap* can work, plasmid #2283 was created by PCR amplification of HBoV1 *cap* from plasmid #2224 using primers #53 / #54 with restriction sites PacI / AscI, respectively. The PCR amplicon was digested with PacI / AscI and cloned into plasmid #1608 (PacI / AscI digested). Plasmids used for cloning and final products are listed in **Table 6**.

Table 6. List of BoV shuffling-acceptors and plasmids used for cloning.

No.	Plasmid Name	Description	Origin
#1608	[a]pSSV9_Pac_Asc_ccdB	AAV shuffling acceptor plasmid with the *rep* ORF from AAV2 followed by a ccdB resistance cassette	Anne-Kathrin Herrmann
#1914	[a]pSSV9_VECB_YFP	ssAAV backbone plasmid with AAV2 ITRs and PacI/EagI restriction sites for cloning	Anne-Kathrin Herrmann
#2281	pTR-BoV shuffling acceptor	BoV shuffling acceptor plasmid with *ns 1-4*, *np 1* (both HBoV1 origin). The HBoV1 *cap* was deleted and replaced by a BsmBI site. HBoV1 LEH and REH flank the viral genome. Used for production of an autonomously replication- competent BoV library	This study
#2282	pAAVNSΔVP-1×BsmBI	Same as #2281, except that AAV2 ITRs (5' and 3') were used to flank the viral genome instead of BoV TRs. Used for production of a packageable but not autonomously replication-competent BoV library	This study
#2283	pAAV-rep Cap1	Plasmid encoding the *rep* ORF from AAV2 and the HBoV1 *cap* instead of AAV2 *cap*	This study

Numbers refer to the internal AG Grimm library. [a] plasmid used for cloning.

2.2.1.4 Cloning of oversized AAV vectors

To obtain oversized scAAV constructs, portions of the *lacZ* cDNA ranging from 100 to 500 bp were PCR-amplified from plasmid pIRES-LacZ #960 using forward primer #55 in combination with reverse primers #56 to #59 (**Table 7**). All PCR products were digested with EcoRI, which enabled cloning into an EcoRI site downstream of the *yfp* (yellow fluorescent protein) ORF in pBSUF3rev-YFP-sds (#552), resulting in plasmids #2284 to #2287 and #2334 (see **Table 8** for the list of all plasmids). In this backbone, two AAV ITRs, one of which carries a mutation of the terminal resolution site, flank the expression cassettes and allow for their packaging as self-complementary (sc) AAV genomes [241]. The other oversized scAAV-YFP constructs used in this work (plasmids #554 to #556) were cloned by Eike Kienle.

Table 7. Primers used for cloning of oversized AAV constructs.

No.	Oligo Name	Sequence (5‘- 3‘)
#55	LacZ fwd	CG**GAATTC**GTGCCGGAAAGCTGGCTGGAG
#56	LacZ 100 rev	ACTT**GAATTC**GTGTATCTGCCGTGCACTGCA
#57	LacZ 200 rev	ACTT**GAATTC**GACCACTACCATCAATCCGGTAG
#58	LacZ 300 rev	ACTT**GAATTC**GTTTACCCGCTCTGCTACCTGC
#59	LacZ 500 rev	ACTT**GAATTC**GTAGCGGCTGATGTTGAACTGG
#60	CMV_pacI_fwd	GTTAC**TTAATTAA**TTCGGTACCCGTTACATAACTTACGG
#61	SV40_Intron_NheI	GTTA**CGCTAG**CGGTGGCGACCGGTGCGG

***Bold**: restriction sites.*

The cloning of over-sized ssAAV vector constructs expressing SpCas9 was performed together with Yannik Voss (BSc. thesis). Therefore, plasmid #1543 containing the SpCas9 cDNA from the Zhang lab (Addgene plasmid #52961) was used as an acceptor plasmid for cloning. The full-length and minimal CMV promoters (from plasmids #1187 and #1521, respectively) were directly cloned into #1543 by PacI / NheI digest. CMV+I, a CMV promoter with a SV40 intron, was PCR-amplified from plasmid #552 using primers #60 and #61 (**Table 7**) and cloned into #1543 using a PacI / NheI digest. The polymerase III promoters (Pol III; U6 and H1) together with their gRNA expression scaffold [242] carrying two inverted BbsI restriction sites were directly cloned as AscI / NotI fragments. The resulting plasmids are listed in **Table 8** (#2288 to #2291).

Table 8. Oversized sc- and ssAAV vectors generated or used in this work.

No.	Plasmid Name	Description	Origin
#552	pBSUF3rev-YFP-sds	scAAV vector with *yfp* expression cassette	Eike Kienle
#2334	#552-100 bp	Same as #552 with 100 bp stuffer	This study
#2284	#552-200 bp	Same as #552 with 200 bp stuffer	This study
#2285	#552-300 bp	Same as #552 with 300 bp stuffer	This study
#2286	#552-400 bp	Same as #552 with 400 bp stuffer	This study
#2287	#552-500 bp	Same as #552 with 500 bp stuffer	This study
#554	#552-800 bp	Same as #552 with 800 bp stuffer	Eike Kienle
#555	#552-1200 bp	Same as #552 with 1200 bp stuffer	Eike Kienle
#556	#552-1600 bp	Same as #552 with 1600 bp stuffer	Eike Kienle
#1543	EFS_ZhangCas9_H1_mFIX1.1_F+E scaffold	Single-stranded AAV plasmid encoding the SpCas9 and H1-gRNA cassette against the murine Factor IX	Chronis Fatouros

#2288	pSSV9-miCMV-spCas9-H1-gRNA CFTR	ssAAV plasmid encoding the SpCas9 (Shalem *et al*. 2014 [243]) driven by a miCMV promoter. A second cassette encodes a H1 promoter followed by a gRNA against CFTR	Yannik Voss (BSc. thesis)
#2289	pSSV9-miCMV-spCas9-U6-gRNA CFTR	Same as #2288. The gRNA was expressed from the U6 promoter	Yannik Voss (BSc. thesis)
#2290	pSSV9-CMV-spCas9-U6-gRNA CFTR	Same as #2288. The SpCas9 cDNA was expressed from the CMV promoter. The gRNA was expressed from the U6 promoter	Yannik Voss (BSc. thesis)
#2291	pSSV9-CMV+I-spCas9-U6-gRNA CFTR	Same as #2288. The SpCas9 cDNA was expressed from the CMV promoter followed by the SV40 intron. The gRNA was expressed from the U6 promoter	Yannik Voss (BSc. thesis)

Numbers refer to the internal AG Grimm library.

2.2.1.5 Cloning of gRNAs

gRNAs were expressed from the U6 or H1 Pol III promoters and cloned using Golden Gate assembly and type II restriction enzymes (BbsI or BsmBI) as previously described [244]. Oligonucleotides (sense and anti-sense) were annealed and cloned as described in section 2.2.3.6. The length of the used oligonucleotides (**Table 9**) varied from 24 nt (for SpCas9, 20 nt target sequence and 4 nt overhangs for cloning) to 25 nt (for SaCas9, 21 nt target sequence and 4 nt overhangs). The acceptor plasmids for cloning of the gRNAs used in this work are listed in **Table 10** (#1529, #1514 for SpCas9 and #1578, #1581 for SaCas9). The acceptor plasmids were designed to always generate the same 4 nt overhangs upon digestion with BsmBI or BbsI (CACC and AAAC). Some plasmids included a ccdB cassette (#1578, #1581) to increase the efficiency of cloning.

Likewise, the shRNAs used in this work were ordered as forward and reverse oligonucleotides (#98 to #101). Oligos were annealed and cloned as described for gRNAs into plasmid #67 behind the U6 promoter to generate plasmids pBS RSV-GFP-U6-shRNA luc 1 and pBS RSV-GFP-U6-shRNA luc 2 (#2356).

Table 9. Oligonucleotide sequences used for cloning of gRNAs.

No.	Oligo Name	Sequence (5‘- 3‘)
#62	Firefly_Luci_saG1_21mer_fwd	**CACC**GCACTGGCATGAAGAACTGCA
#63	Firefly_Luci_saG1_21mer_rev	**AAAC**TGCAGTTCTTCATGCCAGTGC
#64	Firefly_Luci_saG2_21mer_fwd	**CACC**GTCAACGAGTACGACTTCGTGC
#65	Firefly_Luci_saG2_21mer_rev	**AAAC**GCACGAAGTCGTACTCGTTGAC
#66	Firefly_Luci_saG3_21mer_fwd	**CACC**GAGACAGGTCGTACTTGTCGAT
#67	Firefly_Luci_saG3_21mer_rev	**AAAC**ATCGACAAGTACGACCTGTCTC
#68	Firefly_Luci_saG4_21mer_fwd	**CACC**GCTCTTAGCGAAGAAGCTGAA
#69	Firefly_Luci_saG4_21mer_rev	**AAAC**TTCAGCTTCTTCGCTAAGAGC
#70	JF_SV40polA_gRNA1_fwd	**CACC**GTTGTTTATTGCAGCTTATAA
#71	JF_SV40polA_gRNA1_rev	**AAAC**TTATAAGCTGCAATAAACAAC
#72	JF_SV40polA_gRNA2_fwd	**CACC**GTATGTTTCAGGTTCAGGGGG
#73	JF_SV40polA_gRNA2_rev	**AAAC**CCCCCTGAACCTGAAACATAC
#74	JF_SV40polA_gRNA3_fwd	**CACC**GATAAGATACATTGATGAGTT
#75	JF_SV40polA_gRNA3_rev	**AAAC**AACTCATCAATGTATCTTATC
#76	JF_gRNA1_hLuc_fwd	**CACC**GCATCTCGAAGTACTCGGCAT
#77	JF_gRNA1_hLuc_rev	**AAAC**ATGCCGAGTACTTCGAGATGC
#78	JF_gRNA2_hLuc_fwd	**CACC**GGACTCTAAGACCGACTACC
#79	JF_gRNA2_hLuc_rev	**AAAC**GGTAGTCGGTCTTAGAGTCC
#80	JF_gRNA3_hLuc_fwd	**CACC**GTTGCCGAAAATAGGGTCGC
#81	JF_gRNA3_hLuc_rev	**AAAC**GCGACCCTATTTTCGGCAAC
#82	JF_gRNA4_hLuc_fwd	**CACC**GCGACACCGCTATTCTGAGCG
#83	JF_gRNA4_hLuc_rev	**AAAC**CGCTCAGAATAGCGGTGTCGC
#84	JF_G3sp_Luci_scr_fwd	**CACC**ACCACTACGACTAACGGCGT
#85	JF_G3sp_Luci_scr_rev	**AAAC**ACGCCGTTAGTCGTAGTGGT
#86	JF_G1sa_Luci_scr_fwd	**CACC**AGTAATAAGCCGTACCGGACG
#87	JF_G1sa_Luci_scr_rev	**AAAC**CGTCCGGTACGGCTTATTACT
#88	saRNA1_against N saCas9_fwd	**CACC**GGTGATGCCGATGTCCAGGCC
#89	saRNA1_against N saCas9_rev	**AAAC**GGCCTGGACATCGGCATCACC
#90	saRNA2_against N saCas9_fwd	**CACC**GCTGAAGCGGCGGAGGCGGCA
#91	saRNA2_against N saCas9_rev	**AAAC**TGCCGCCTCCGCCGCTTCAGC
#92	saRNA1_against C saCas9_fwd	**CACC**GAAACCTTCAAGAAGCACATC
#93	saRNA1_against C saCas9_rev	**AAAC**GATGTGCTTCTTGAAGGTTTC
#94	saRNA2_against C saCas9_fwd	**CACC**GCCCTTGCCCTTGGCCAGATT

#95	saRNA2_against C saCas9_rev	**AAAC**AATCTGGCCAAGGGCAAGGGC
#96	sc(sa) gRNA1_N saCas fwd	**CACC**GCAGTCAATTGCCGGGCGCTG
#97	sc(sa) gRNA1_N saCas rev	**AAAC**CAGCGCCCGGCAATTGACTGC
#98	JF_hLuc21-siRNA-1-fwd	**CACC**GAGAAGGAGATCGTGGACTATTCAA GAGATAGTCCACGATCTCCTTCTC
#99	JF_hLuc21-siRNA-1-rev	**AAAA**GAGAAGGAGATCGTGGACTATCTCT TGAATAGTCCACGATCTCCTTCTC
#100	JF_hLuc21-siRNA-2-fwd	**CACC**GAGCACCCTGATCGACAAGTATCAA GAGTACTTGTCGATCAGGGTGCTC
#101	JF_hLuc21-siRNA-2-rev	**AAAA**GAGCACCCTGATCGACAAGTACTCTT GATACTTGTCGATCAGGGTGCTC

Bold: *Overhangs.*

Table 10. Plasmids used for cloning or expression of gRNAs.

No.	Plasmid Name	Description	Origin
#1529	scAAV-RSV-GFP-U6-2×BbsI (sp) scaffold	Acceptor plasmid for cloning of (sp) gRNAs after the U6 promoter	Dirk Grimm
#2292	scAAV-RSV-GFP-U6-anti Luc gRNA1 (sp) scaffold	Plasmid based on #1529 with gRNA1 (sp) against Firefly luciferase (hluc+)	This study
#2293	scAAV-RSV-GFP-U6-anti Luc gRNA2 (sp) scaffold	Plasmid based on #1529 with gRNA2 (sp) against Firefly luciferase (hluc+)	This study
#2294	scAAV-RSV-GFP-U6-anti Luc gRNA3 (sp) scaffold	Plasmid based on #1529 with gRNA3 (sp) against Firefly luciferase (hluc+)	This study
#2295	scAAV-RSV-GFP-U6-anti Luc gRNA4 (sp) scaffold	Plasmid based on #1529 with gRNA4 (sp) against Firefly luciferase (hluc+)	This study
#1514	pBS-H1_F+E RSV:GFP	Acceptor plasmid for cloning of (sp) gRNAs after the H1 promoter	Florian Schmidt
#2296	pBS-H1-luc gRNA3 (sp) scaffold	Plasmid based on #1514 with the gRNA3 (sp) against Firefly luciferase (hluc+)	This study

#1578	scAAV-RSV-GFP-U6-2×BsmBI ccdB (sa) scaffold	Acceptor plasmid for cloning of (sa) gRNAs after the U6 promoter	Florian Schmidt
#2297	scAAV-RSV-GFP-U6-anti Luc gRNA1 (sa) scaffold	Plasmid based on #1578 with gRNA1 (sa) against Firefly luciferase (hluc+)	This study
#2298	scAAV-RSV-GFP-U6-anti Luc gRNA2 (sa) scaffold	Plasmid based on #1578 with gRNA2 (sa) against Firefly luciferase (hluc+)	This study
#2299	scAAV-RSV-GFP-U6-anti Luc gRNA3 (sa) scaffold	Plasmid based on #1578 with gRNA3 (sa) against Firefly luciferase (hluc+)	This study
#2300	scAAV-RSV-GFP-U6-anti Luc gRNA4 (sa) scaffold	Plasmid based on #1578 with gRNA4 (sa) against Firefly luciferase (hluc+)	This study
#1581	scAAV-RSV-GFP-H1-2×BsmBI ccdB (sa) scaffold	Acceptor plasmid for cloning of (sa) gRNAs after the H1 promoter	Florian Schmidt
#2356	pBS RSV-GFP-U6-shRNA luc 2	Plasmid based on #67 with shRNA2 against Firefly luciferase (hluc+)	This study

*Numbers refer to the internal AG Grimm library. Sequences of gRNA and shRNA oligos can be found in **Table 9**.*

2.2.1.6 Cloning of luciferase reporters and kill-switch vectors

All reporter constructs were derived from plasmid #714, which was originally cloned by Dominik Niopek by transferring the two luciferase expression cassettes from the psiCheck-2 reporter (Promega) into a derivative of pSSV9 [240].

The first generation of kill-switch reporters (#2301 to #2304) were constructed by PCR-amplifying the respective H1-gRNA cassettes using common primers #102 and #103 (**Table 11**) with ClaI restriction sites. The fragments were gel-purified, digested and inserted into ClaI digested #714 vector (see **Table 12** for a list of plasmids).

Plasmid #2342 was constructed by first PCR-amplifying the TTR promoter sequence from plasmid #1693 using forward primer #104 and reverse primer #105 with BstZ17I and HindIII overhangs, respectively. The digested and gel-purified fragment was then cloned into BstZ17I / HindIII-digested plasmid #2305. The kill-switch luciferase reporters #2306 and #2343 were derived from #2342 by cloning the respective gRNA oligos using Golden Gate cloning as described in 2.2.3.6 (see also **Table 9** for oligo sequences).

Table 11. Primers used for cloning of luciferase reporters.

No.	Oligo Name	Sequence (5‘- 3‘)
#102	H1_promoter-claI-fwd	GTAC**ATCGAT**GCCCATATTTGCATGTCG
#103	ClaI_E+F_scaffold_rev	GTTACC**ATCGAT**AAAAAAAGCACCGACTCG
#104	Fwd_BstZ17I TTR	GATAT**GTATAC**GGATCTGTAATTCACGCGAG
#105	Rev_HIND III TTR	GATAT**AAGCTT**CAGCTGGGCTTCTCCTGGT
#106	hLuc+_notI_fwd	TAGAT**GCGGCCGC**ATGGCCGATGCTAAGAACAT
#107	SV40 polA_sphI_rev	TGCAT**GCATGC**GTTTATTGCAGCTTATAATG

Bold*: restriction sites.*

For the generation of scAAV-luciferase reporters #2307 to #2309, the Firefly luciferase ORF was PCR-amplified using primers #106 and #107 with NotI / SphI restriction sites, respectively. Next, the *bfp* ORF was replaced in plasmids #1690, #1693 and #1697 by cloning the digested and purified hLuc fagment into the respective SphI / NotI-digested vectors.

Table 12. List of conventional and SIN luciferase reporters.

No.	Plasmid Name	Description	Origin
#714	pSSV9_pSi	ssAAV plasmid encoding Firefly *and Renilla* luciferase	Dominik Niopek
#2301	pSSV9_pSi_H1_anti Luc (sp)_3	ssAAV plasmid based on #714 with the H1-(sp) gRNA3 cassette against Firefly luciferase (hluc+)	This study
#2302	pSSV9_pSi-H1-scr control (sp) 3	ssAAV plasmid based on #714 with a H1-(sp) scr gRNA3 control cassette against Firefly luciferase (hluc+)	This study
#2303	pSSV9_pSi-H1-anti Luc (sa) 1	ssAAV plasmid based on #714 with the H1-(sa) gRNA1 cassette against Firefly luciferase (hluc+)	This study
#2304	pSSV9_pSi-H1-scr control (sa) 1	ssAAV plasmid based on #714 with the H1-(sa) scr gRNA1 control cassette against Firefly luciferase (hluc+)	This study
#2305	ssAAV-TK-hLuc-U6-2×BbsI (sa) gRNA scaffold	ssAAV plasmid encoding Firefly luciferase under TK promoter and U6-2×BbsI (sa) gRNA scaffold	Carolin Schmelas
#2342	ssAAV-TTR-hLuc-U6-2×BbsI (sa) gRNA scaffold	ssAAV plasmid based on #2305 encoding Firefly luciferase under TTR promoter and U6-2×BbsI (sa) gRNA scaffold	This study

#2343	ssAAV_TTR-hLuc-U6-scr control (sa) 1	Plasmid based on #2342 encoding Firefly luciferase under TTR promoter and U6-(sa) scr gRNA1 control	This study
#2306	ssAAV_TTR-hLuc-U6-anti Luc (sa) 1	Plasmid based on #2342 encoding Firefly luciferase under TTR promoter and U6-(sa) gRNA1 cassette against Firefly luciferase (hluc+)	This study
#1693	scAAV_TTR_BFP	scAAV Plasmid encoding blue fluorescent protein under TTR promoter	Florian Schmidt
#1690	scAAV_LP1_BFP	scAAV Plasmid encoding blue fluorescent protein under LP1 promoter	Florian Schmidt
#1697	scAAV_TBG_BFP	scAAV Plasmid encoding blue fluorescent protein under TBG promoter	Florian Schmidt
#2307	scAAV_TTR_hLuc	scAAV Plasmid encoding Firefly luciferase (hluc+) under TTR promoter	This study
#2308	scAAV_LP1_hLuc	scAAV Plasmid encoding Firefly luciferase (hluc+) under LP1 promoter	This study
#2309	scAAV_TBG_hLuc	scAAV Plasmid encoding Firefly luciferase (hluc+) under TBG promoter	This study

Numbers refer to the internal AG Grimm library.

2.2.1.7 Cloning of SIN split Cas9 constructs

Original split Cas9 plasmids that served as templates for cloning were received from Carolin Schmelas. First, to optimize the size of these vectors for packaging as ssAAV genomes, 1.2 or 1.7 kb stuffer sequences were PCR-amplified from plasmid pBSII-KS (+) #48, using primer combinations #108 / #109 or #108 / #110, respectively (see **Table 13** for a list of primers). Gel-purified and PacI-digested PCR amplicons were cloned into plasmid #2314 (the 1.7 kb stuffer) or #2322 (the 1.2 kb stuffer) to yield plasmids #2315 and #2323, respectively (**Table 14**). The final genome sizes of the vectors (including the ITRs) were 3.9 kb for #2315 and 4.0 kb for #2323.

Next, the gRNA acceptor cassettes under the H1 or U6 promoters were PCR-amplified using common primers #111 to #119 with different restriction sites:

A H1-2×BsmBI ccdB-sa scaffold cassette was PCR-amplified from plasmid #1581 using primers #111 and #112. The fragment was gel-purified, digested with PacI and cloned into plasmid #2310 (using the PacI restriction site) to yield plasmid #2311 (together with Daniel Heid).

The ccdB cassette in plasmid #1781 was replaced by annealed oligonucleotides harboring two inverted BbsI sites and compatible overhangs using Golden Gate cloning (with restriction enzyme BsmBI). The resulting plasmid served as a template for PCR amplification of the U6 gRNA acceptor cassette using primers #118 / #119 (ClaI / EcoRV sites). The ClaI / EcoRV-digested fragment was cloned into plasmids #2315 and #2323 (both digested with ClaI / EcoRV) to yield plasmids #2316 and #2324, respectively.

A H1-2×BbsI-sa scaffold cassette was PCR-amplified from plasmid #1533 using primer combinations #113 / #114 (NotI sites) or #117 / #119 (ClaI / EcoRV sites). PCR amplicons were gel-purified and digested with the respective restriction enzyme(s). Next, the NotI-digested products were cloned into the #2331 vector (digested with NotI), close to the AAV2 ITR to yield the plasmid scAAV-mCMV-NLS-flag-SaCas9(C)-split3-gp41_miniPolA-H1-2×BbsI. This plasmid served as a template for cloning of the anti-Cas gRNA against the Cas9 N terminal (#1) and the scr control RNA (listed in **Table 9**). These plasmids were used for the initial validation experiments described in section 3.4.4 but were replaced afterwards by optimized constructs following the recommendations of Xie *et al.* [245] . In this study, the authors showed that the incorporation of cassettes with small hairpins in scAAV vectors close to the non-mutated ITR can be disadvantageous, leading to truncations in the packaged genomes. Therefore, the H1-2×BbsI-sa cassette was re-cloned into the AscI site next to the mutated AAV4 ITR using the same strategy as before (primers #115 / #116), to yield plasmids #2332 to #2334. These optimized plasmids were then used in all transduction experiments (section 3.4.5).

Table 13. Primers used for cloning of SIN split Cas9 vectors.

No.	Oligo Name	Sequence (5‘- 3‘)
#108	Fwd pB KS (+) PacI	GTACC**TTAATTAA**GTGTGGTGGTTACGCGCA
#109	Rev 1200 pB KS (+) PacI	GTACC**TTAATTAA**CCTCTGACTTGAGCGTCGAT
#110	Rev 1700 pB KS (+) PacI	GTACC**TTAATTAA**GGTTTGTTTGCCGGATCAAG
#111	H1_PacI_fwd	TGTAC**TTAATTAA**GCAAAAAAATCTCGCCA
#112	JF_PacI_sa_scaffold_rev	AGAT**TTAATTAA**GCAAAAAAATCTCGCCA
#113	H1_NotI_fwd	TGTAC**GCGGCCGC**GCCCATATTTGCATGTCG
#114	sa scaffold_NotI_rev	TGTAC**GCGGCCGC**GCAAAAAAATCTCGCCA
#115	H1_AscI_fwd	ATATAA**GGCGCGCC**GCCCATATTTGCATGTCGC
#116	sa scaffold_AscI_rev	TATAT**GGCGCGCC**AAAAAAATCTCGCCAACAAGT TGAC
#117	H1_ClaI_fwd	TGTAC**ATCGAT**GCCCATATTTGCATGTCG
#118	U6_ ClaI_fwd	TGTAC**ATCGAT**AGATCTGATATCGGCGCGCC
#119	sa scaffold_EcoRV_rev	TGTAC**GATATC**GCAAAAAAATCTCGCCA

***Bold**: restriction sites.*

The H1-2×BbsI-sa ClaI / EcoRV-digested fragment was cloned into plasmids #2315 and #2323 (both digested with ClaI / EcoRV) to yield plasmids #2319 and #2327.

The gRNA sequences against Firefly luciferase or SaCas9 (sequences shown in **Table 9**) were then cloned as described in section 2.2.3.6 to yield plasmids #2312-13, #2317-18, #2320-21, #2325-26, #2328-29, or #2333-34 (**Table 14**)**.**

Table 14. Split Cas9 plasmids used and SIN variants generated in this work.

No.	Plasmid Name	Description	Source
		ssAAV split Cas9 vectors	
#2310	ssAAV-mCMV-NLS-flag-SaCas9-NLS-miniPolyA	Single-stranded AAV vector with full-length SaCas9 expression cassette driven by a minimal CMV promoter	Carolin Schmelas
#2311	ssAAV-H1-ccdB (sa) scaffold-mCMV-NLS-flag-SaCas9-NLS-miniPolyA	Same as #2310. A H1 promoter and a ccdB cassette were directly cloned before the mCMV promoter into the PacI site	Daniel Heid
#2312	ssAAV-H1-anti Cas gRNA-mCMV-NLS-flag-SaCas9-NLS-miniPolyA	Same as #2311 with a gRNA sequence against SaCas9 (N1; sa gRNA1 against the N terminal)	This study
#2313	ssAAV-H1-scr Cas gRNA-mCMV-NLS-flag-SaCas9-NLS-miniPolyA	Same as #2311 with a scrambled anti-Cas9 gRNA N1 control sequence	This study
#2314	ssAAV-mCMV-NLS-flag-SaCas9(N)-split3-gp41-flag-NLS	Single-stranded AAV vector that encodes the N-terminal part of a split Cas9 driven by a minimal CMV promoter.	Carolin Schmelas
#2315	ssAAV-Stuffer-mCMV-NLS-flag-SaCas9(N)-split3-gp41-flag-NLS	Same as #2314. A stuffer with a multiple cloning site was inserted before the mCMV promoter to optimize the size and allow for subsequent cloning steps	This study
#2316	ssAAV-U6-2×BbsI (sa) scaffold-mCMV-NLS-flag-SaCas9(N)-split3-gp41-flag-NLS	Acceptor plasmid based on #2315 with U6 Pol III promoter followed by two inverted BbsI sites for cloning of gRNA cassettes	This study

#2317	ssAAV-U6- anti Luc gRNA (sa) scaffold-mCMV-NLS-flag-SaCas9(N)-split3-gp41-flag-NLS	Plasmid based on #2316 with sa gRNA1 against Firefly luciferase (hluc+)	This study
#2318	ssAAV-U6- scr Luc gRNA (sa) scaffold-mCMV-NLS- flag-SaCas9(N)-split3-gp41-flag-NLS	Plasmid based on #2316 with a scrambled anti-luc (sa) gRNA1 control sequence	This study
#2319	ssAAV-H1-2×BbsI (sa) scaffold-mCMV-NLS-flag-SaCas9(N)-split3-gp41-flag-NLS	Acceptor plasmid based on #2315 with H1 Pol III promoter followed by two inverted BbsI sites for cloning of gRNA cassettes	This study
#2320	ssAAV-H1- anti Luc gRNA (sa) scaffold-mCMV-NLS-flag-SaCas9(N)-split3-gp41-flag-NLS	Plasmid based on #2319 with (sa) gRNA1 against Firefly luciferase (hluc+)	This study
#2321	ssAAV-H1- scr Luc gRNA (sa) scaffold-mCMV-NLS- flag-SaCas9(N)-split3-gp41-flag-NLS	Plasmid based on #2319 with a scrambled anti-luc (sa) gRNA1 control sequence	This study
#2322	ssAAV-mCMV-NLS-flag-SaCas9(C)-split3-gp41-flag-NLS	Single-stranded AAV vector that encodes the C-terminal part of a split Cas9 driven by a minimal CMV promoter	Carolin Schmelas
#2323	ssAAV-Stuffer-mCMV-NLS-flag-SaCas9(C)-split3-gp41-flag-NLS	Same as #2322. A 1700 bp stuffer with a multiple cloning site was inserted before the mCMV promoter to optimize the size and allow for subsequent cloning steps	This study
#2324	ssAAV-U6-2×BbsI (sa) scaffold-mCMV-NLS-flag-SaCas9(C)-split3-gp41-flag-NLS	Acceptor plasmid based on #2323 with U6 Pol III promoter followed by two inverted BbsI sites for cloning of gRNA cassettes	This study

#2325	ssAAV-U6- anti Cas gRNA (sa) scaffold-mCMV-NLS-flag-SaCas9(C)-split3-gp41-flag-NLS	Plasmid based on #2324 with gRNA N1 against SaCas9	This study
#2326	ssAAV-U6- scr Cas gRNA (sa) scaffold-mCMV-NLS- flag-SaCas9(C)-split3-gp41-flag-NLS	Plasmid based on #2324 with a scrambled anti-Cas9 gRNA N1 control sequence	This study
#2327	ssAAV-H1-2×BbsI (sa) scaffold-mCMV-NLS-flag-SaCas9(C)-split3-gp41-flag-NLS	Acceptor plasmid based on #2323 with H1 Pol III promoter followed by two inverted BbsI sites for cloning of gRNA cassettes	This study
#2328	ssAAV-H1- anti Cas gRNA (sa) scaffold-mCMV-NLS-flag-SaCas9(C)-split3-gp41-flag-NLS	Plasmid based on #2327 with gRNA N1 against SaCas9	This study
#2329	ssAAV-H1- scr Cas gRNA (sa) scaffold-mCMV-NLS- flag-SaCas9(C)-split3-gp41-flag-NLS	Plasmid based on #2327 with a scrambled anti-Cas9 gRNA N1 control sequence	This study
	scAAV split Cas9 vectors		
#2330	scAAV-mCMV-NLS-flag-SaCas9(N)-split3-gp41_miniPolA	Self-complementary AAV vector that encodes the N-terminal part of a split Cas9 driven by a minimal CMV promoter	Carolin Schmelas
#2331	scAAV-mCMV-NLS-flag-SaCas9(C)-split3-gp41_miniPolA	Self-complementary AAV vector that encodes the C-terminal part of a split Cas9 driven by a minimal CMV promoter	Carolin Schmelas
#2332	scAAV-H1-2×BbsI-mCMV-NLS-flag-SaCas9(C)-split3-gp41_miniPolA	Acceptor plasmid based on #2331 with H1 Pol III promoter followed by two inverted BbsI sites for cloning of gRNA cassettes	Together with Carolin Schmelas

#2333	scAAV-H1- anti Cas gRNA (sa) scaffold-mCMV-NLS-flag-SaCas9(C)-split3-gp41_miniPolA	Plasmid based on #2332 with gRNA N1 against SaCas9	This study
#2334	scAAV-H1- scr Cas gRNA (sa) scaffold -mCMV-NLS-flag-SaCas9(C)-split3-gp41_miniPolA	Plasmid based on #2332 with a scrambled anti-Cas9 gRNA N1 control sequence	This study
#1533	ssAAV_CMV_SaCas9_BgH_H1_(BbsIx2) Sa-scaffold	All-in-one SaCas9 vector 'empty', BbsI sites for gRNA cloning (H1 promoter)	Chronis Fatouros
#1781	scAAV-RSV-GFP-U6-2×BsmBI ccdB (sa) scaffold	Acceptor plasmid for cloning of (sa) gRNAs after the U6 promoter	Florian Schmidt

Numbers refer to the internal AG Grimm library.

2.2.2 Microbiological methods

2.2.2.1 Production of $CaCl_2$ chemocompetent bacteria

For the preparation of standard (MAX Efficiency DH5αTM, section 2.1.1) or recombination-deficient chemocompetent cells (SURE), an aliquot from the frozen glycerol stock was resuspended in 50 µL LB medium under sterile conditions. From this suspension, a 1:10 dilution was streaked on an agar plate without antibiotics and was incubated at 37°C O/N. The next day, one colony was picked and inoculated in 6 mL LB medium O/N. Around 16 h later, 3 flasks (250 mL SOC medium each) were inoculated with 1 mL of the O/N culture and were then incubated at 37°C, 180 rpm. The cultures that reached an appropriate OD_{600} of 0.5 were then centrifuged at 4°C, 1,800×g for 15 min. The cell pellet was resuspended in 40 mL TFB I buffer and incubated for 10 min on ice. Next, cells were pelleted again (4°C, 1,800×g for 15 min) and resuspended in 10 mL of TFB II. After an incubation period of 10 min on ice, cells were distributed with a stepper pipette into 50 µL aliquots and snap-frozen in liquid nitrogen. The aliquots were then stored at -80°C or directly used for transformations. To test the transformation efficiency, ~50 pg of the supplied pUC plasmid (section 2.1.1) were transformed and the efficiency was calculated as described in section 2.2.2.3.

2.2.2.2 Heat shock

For the transformation of plasmid DNA or ligation reactions, 50 µL aliquots of chemocompetent bacteria were thawed on ice for 5 min. The plasmid DNA (~10 ng) or ligation reaction (75-100 ng) was added to the cells and the tube was gently flicked. After a 20 min incubation on ice, heat shock was performed by shifting the bacteria to 42°C for 45 sec. Next, a "cold-shock" was achieved by shifting the cells to 4°C for 5 min. To allow recovery of bacteria and production of antibiotic resistance proteins, 1 mL of antibiotic-free LB media was added to the tubes and cells were grown for 40 min at 37°C, 660 rpm. The tubes were then centrifuged at 400×g, 5 min, and the pellets were resuspended in 100 µL LB medium and streaked on antibiotic-containing agar plates.

For ampicillin-encoding plasmids, the 40 min incubation step was not critical and therefore omitted. All plates were incubated at 37°C O/N.

2.2.2.3 Production of electrocompetent bacteria

To produce electrocompetent bacteria, MegaX DH10B^{TM} T1^{R} cells (section 2.1.1) were streaked on agar plates as described in 2.2.2.1. One colony was inoculated in 30 mL LB medium and incubated O/N at 37°C, 180 rpm. Four flasks with 400 mL antibiotic-free LB were inoculated with 5 mL of the overnight culture and were further incubated at 37°C, 180 rpm until an OD_{600} of 0.5 was reached. The bacteria were then transferred to pre-cooled 500 mL flasks and incubated on ice for 15 min. Next, the flasks were centrifuged at 4°C, 4,000×g for 10 min and the pellet was subsequently resuspended in 60 mL ice cold sterile H_2O. The bacteria were then transferred into pre-cooked dialysis tubings (Type 20/32 inch) and dialyzed against 4-8 L ddH_2O at 4°C O/N. The next day, cells were transferred to 50 mL tubes and centrifuged at 4°C, 4,000×g for 15 min. Afterwards, the pellet was suspended in 900 µL 10% glycerol and the OD_{600} was measured. By the gradual addition of 10% glycerol, the OD_{600} was adjusted to 1 (which corresponds to ~2.5×10^{10} cells). Electrocompetent bacteria were aliquoted in 100-300 µL aliquots, snap-frozen in liquid nitrogen and stored at -80°C. To determine the transformation efficiency, the pUC19 control plasmid and the transformation protocol supplied with the MegaX DH10B^{TM} T1^{R} cells (see section 2.1.1) were used. Therefore, 50 pg of pUC19 were added to a 50 µL aliquot of bacteria and electroporation was performed as described in section 2.2.2.4. Transformation efficiency (colony forming units [CFU] / µg) was calculated according to the following formula:

$$\frac{\text{CFU}}{\mu\text{g}} = \frac{\text{number of colonies}}{50\ \text{pg}} \times \frac{1\times10^{6}\ \text{pg}}{\mu\text{g}} \times \frac{\text{total bacterial volume}}{\text{plated volume}} \times \text{dilution factor(s)}$$

2.2.2.4 Electroporation

Chemocompetence was sufficient for most regular cloning applications in this work. However, if the amount of DNA was limiting and/or high transformation efficiencies were required, electroporation was performed. Therefore, bacteria prepared as described in 2.2.2.3 were used. For chimeric BoV library production, even higher transformation efficiencies were needed. Thus, transformations were carried out with commercial electrocompetent bacteria (E. cloni SUPREME electrocompetent cells, section 2.1.1).

In both cases, electroporation cuvettes were cooled on ice and SOC medium pre-warmed at 37°C. Aliquots of competent cells of 300 µL were thawed on ice for 10 min. In the meantime, the ligation reactions were dialyzed against ddH_2O using MF-Millipore™ Membrane Filters. Next, the ligation reaction was mixed with the bacteria (~150 ng DNA per 30 µL bacteria) and placed on ice. Into each 1 mm electroporation cuvette, 35 µL DNA/bacteria mix was transferred and the electroporation was performed in a Gene Pulser Xcell device at 1,800 V, 25 µF and 200 Ω. To allow recovery, bacteria were directly resuspended in 1 mL of SOC medium, transferred into a 1.5 mL tube and shaken at 37°C, 600 rpm for 40 min. For small-scale (1-2) electroporations, cells were spun down after recovery, resuspended in 100 µL medium and directly plated on agar plates with appropriate antibiotics for selection. For library production (6-10 electroporations), bacteria were pooled after recovery and 100 µL of this suspension (undiluted as well as 1:10 and 1:100 dilutions) were plated on LB-ampicillin agar plates, to estimate the number of clones in the library. The rest of the transformation suspension was grown overnight in 400 mL LB-ampicillin to amplify the plasmid library.

2.2.3 Molecular biology methods

2.2.3.1 DNA sequencing

Sequences of plasmids or cloned PCR products were verified using Sanger sequencing analysis, which was performed by GATC (now Eurofins), Konstanz, Germany. One tube contained ~400 ng plasmid DNA and 5 µM of an appropriate sequencing primer in 20 µL total volume.

2.2.3.2 Standard PCR

For PCR amplification of target sequences in plasmid or genomic DNA, Phusion HS II polymerase (Thermo Fisher Scientific) was used according to the manufacturer's instructions. Briefly, one reaction mix (50 µL) contained template DNA, 2.5 µL of each forward and reverse primer (10 µM), 10 µL Phusion buffer (5×, HF or GC), 1 µL dNTP (10 mM), 0.5 µL Phusion DNA polymerase and H_2O up to 50 µL. The amount of template DNA varied from ~20 ng (plasmid DNA) to 200 ng (genomic DNA). The PCR cycling conditions were as follows: an initial denaturation step at 98°C for 10 min, followed by 35-40 cycles of (i) denaturation at 98°C

for 30 sec, (ii) annealing at 56-60°C for 30 sec, (iii) extension at 72°C (for 30 sec/kb) and finally a last elongation step for 8 min at 72°C. The appropriate annealing temperature was determined for every primer set using the online Phusion Tm calculator [246, 247]. In case of low amplification efficiency or unspecific bands, a gradient PCR was performed to determine the best annealing temperature.

2.2.3.3 Colony PCR

For a quick test of cloning efficiency after transformation, a PCR was directly performed from the bacterial colony (usually 10-30 colonies were tested in parallel). Therefore, the Quick-Load Taq 2× Master Mix (NEB) was used according to the manufacturer's instructions. A smaller reaction volume of 10 µL was found to be sufficient for efficient amplification. Therefore, 5 µL of 2× Master Mix were mixed with 0.2 µL forward primer (10 µM), 0.2 µL reverse primer (10 µM) and 4.6 µL H_2O. A colony was picked with a pipet tip, gently streaked on a replica agar plate with appropriate antibiotics (as a backup) and then dispersed into the PCR tube containing the reagent mix. The replica plate was then incubated at 37°C for several hours. PCR reactions were performed with the following conditions: an initial denaturation step at 95°C for 30 sec, followed by 35-40 cycles of (i) denaturation at 95°C for 30 sec, (ii) annealing at 45-68°C for 15 sec, (iii) extension at 68°C (for 1 min/kb) and a final elongation step for 8 min at 68°C. The appropriate annealing temperature was determined for every primer set using the online NEB Tm calculator (https://tmcalculator.neb.com).

PCR reactions were directly loaded and separated on 1% agarose gels containing ethidium bromide. Next, PCR products were visualized under UV light and positive colonies identified. Correct clones were picked from the replica plate and cultured in selective growth media O/N.

2.2.3.4 Overlap extension PCR (OE-PCR)

To introduce mutations, deletions or insertions at specific sites in a sequence, overlap extension PCR was performed. Therefore, two complementary primers (sense and anti-sense) harbouring the desired change were designed. Two additional external primers (forward and reverse) flanking the region of interest and including restriction sites for cloning were needed. In a first step, two separate PCRs were performed using the Phusion DNA polymerase as described in 2.2.3.2 by combining sense and anti-sense with the external reverse and forward primers, respectively. The PCRs were loaded on a 1% agarose gel and the correct fragments were gel extracted using the QIAquick Gel Extraction Kit. Next, equal amounts of each fragment (PCR 1 and 2) were used in a primer-less PCR for 10 cycles, in which the overlapping

sequences act as primers. For the next 30 cycles, the forward and reverse primers were added to amplify the fusion product. PCR reactions were carried out as described in **Table 15**.

2.2.3.5 Restriction enzyme digestion

To prepare DNA for cloning, a digest with restriction enzymes was performed that generates compatible ends. This was accomplished by incubating 6-10 µg of plasmid DNA or PCR amplicons with restriction enzymes (10 U) for 3 h or O/N at the recommended temperature (usually 37°C). Next, the digest was loaded onto a 1% agarose gel and the required fragments were cut out and purified using the QIAquick Gel Extraction Kit. For standard test digests to assess the success of cloning, 300-500 ng of plasmid DNA was incubated with restriction enzymes (1-2 U) in the respective buffers for 1 h at the recommended temperature.

Table 15. PCR reaction mix and protocol for OE-PCR.

PCR protocol			**PCR reaction**	
Temperature	**Time**		**Amount (µL)**	**Component**
98˚C	30 sec		5	5× HF buffer
98˚C	10 sec	**10×**	2.5	dNTPs
72°C	30 sec / kb		1	PCR 1
Addition of external primers			1	PCR 2
			0.25	Phusion HS II
98°C	10 sec		15,25	H_2O
55-60°C	30 sec	**30×**	**Addition of external primers**	
72°C	30 sec / kb			
72°C	4 min		5	5× HF buffer
4°C	hold		2.5	dNTPs
			2.5	Primer forward
			2.5	Primer reverse
			0.25	Phusion HS II
			12.25	H_2O

The inverted terminal repeats (ITRs) harbor the packaging signals of AAV. These sequences are unstable in standard *E. coli* [248]. Therefore, plasmids were always assessed for the presence of both ITRs after retransformation and after every cloning step. For ssAAVs with two identical ITRs (both from AAV2), a digest with XmaI was used that cuts within the ITR sequences. For scAAVs with two distinct ITRs (from AAV2 and AAV4), XmaI and BsaI were used for the AAV2 and AAV4 ITR, respectively.

Likewise, replication-competent BoV plasmids include the 5'- and 3'-end hairpins and could only be propagated in recombination-deficient SURE cells. The integrity of these sequences was confirmed with SalI and XhoI for the 5'- and 3'-end hairpins, respectively.

2.2.3.6 Ligation

For standard cloning procedures, digested and purified DNA fragments (according to section 2.2.3.5) were mixed in a molar ratio of 1:3 or 1:5 vector to insert (total of 80-200 ng of DNA). Next, 10× Ligation buffer, 0.5 µL of T4 DNA ligase and H_2O were added to a final volume of 10 or 20 µL (depending on the volumes of DNA fragments).

For the introduction of short sequences (*e.g.* gRNAs or shRNAs), two oligonucleotides (sense and anti-sense) were ordered with compatible overhangs for cloning. For annealing of the two strands, a mix was prepared including 2.5 µL forward oligo (100 µM), 2.5 µL reverse oligo (100 µM), 5 µL NEB buffer 2 and 40 µL H_2O. The mix was placed at 95°C for 5 min and then cooled down slowly to room temperature (by switching off the heating block). Next, a 1:200 dilution was made and 2 µL were used in the ligation reaction.

All ligations were performed for 2 h at RT or at 16°C O/N. Finally, 5-10 µL of the ligation mix were transformed into 50 µL chemocompetent *E. coli* (section 2.2.2.2).

2.2.3.7 Amplification and purification of plasmid DNA

For amplification of plasmid DNA, chemocompetent bacteria were first transformed with ~10 ng of plasmid DNA as described in section 2.2.2.2. One clone was picked from the agar plate and inoculated in 3 mL, 80 mL or 350 mL LB media supplied with the appropriate antibiotic for mini-, midi- and maxi-preps, respectively. The cultures were incubated at 37°C in a shaking multitron at 180 rpm for 16 h.

For small-scale (mini) plasmid isolation, 2 mL bacterial culture was centrifuged at 8,000 rpm for 5 min at RT. The plasmid DNA was then extracted using the alkaline lysis method and house-made buffers. Therefore, the bacterial pellet was resuspended in 300 µL of neutral buffer (P1). Next, the cells were lysed by adding 300 µL of alkaline lysis buffer (P2). The tubes were carefully inverted several times and incubated at RT for not more than 5 min. Proteins and high-molecular weight DNA were precipitated by adding neutralizing buffer (P3). The lysate was cleared by centrifugation at 13,200 rpm, 15 min and the DNA was precipitated by mixing 800 µL of the supernatant with 600 µL isopropanol. The DNA was then pelleted at 13,200 rpm, 20 min and washed with 70% ethanol. After a final centrifugation step (13,200 rpm, 5 min), the pellet was air-dried and dissolved in 50 µL H_2O. The plasmid DNA prepared using this method was used for subsequent cloning steps and sequencing analysis.

For the preparation of high-purity small-scale plasmid DNA needed for transfection experiments, the QIAprep Spin Miniprep Kit (Qiagen) was used according to the manufacturer's instructions. If larger amounts of plasmid DNA were required, midi-preps (for

~150-240 µg) using the PureYield™ Plasmid Midiprep kit (Promega) or maxi-preps (for ~600-1,500 µg) using the NucleoBond® Xtra Maxi kit (Macherey-Nagel) were performed according to the manufacturer's protocols.

2.2.3.8 Genomic DNA and RNA isolation

2.2.3.8.1 In-house protocol applied for DNA isolation from the mouse liver

Around 200 mg liver pieces were first homogenized in 3 mL genomic DNA lysis buffer supplied with 20 µL (100 mg/mL) RNase A and then incubated for 30 min at 37°C. Next, 15 µL proteinase K (20 mg/mL) were added, and the tubes were inverted several times and incubated at 55°C O/N. To precipitate contaminating proteins, the tissue lysate was mixed with 1.8 mL 6 M LiCl and centrifuged at 13,000 rpm for 30 min. The supernatant was transferred to a new 15 mL tube, and 2.8 mL isopropanol were added and mixed by inverting the tube several times. Precipitating genomic DNA was visible by eye and could be fished out using a P1000 pipette tip. The DNA was transferred to a new 1.5 mL tube and centrifuged at 13,000 rpm for 5 min. The pellet was washed with 70% ethanol (13,000 rpm, 5 min) and carefully dissolved in 100-200 µL H_2O. When the pellet was too dry, the dissolving of genomic DNA in water was performed O/N at 4°C.

2.2.3.8.2 Isolation of genomic DNA from pHAE

Cells growing on transwells were incubated with PBS from both, apical and basolateral sides for 20 min at 37°C. This step reduced the amount of mucus that interferes with the trypsin digest. Next, 150 µL 0.25% Trypsin-EDTA solution was added to the apical compartment. After 10-15 min at 37°C, the trypsin was inactivated with 500 µL 3% BSA or medium and the cells were transferred into 1.5 mL tubes. Next, the tubes were centrifuged at 500×g for 10 min and washed once with PBS (500×g, 10 min). DNA was extracted from the cell pellet using the DNeasy Blood & Tissue Kit (Qiagen) according to the manufacturer's instructions.

2.2.3.9 Low-molecular weight (Hirt) DNA extracts

To detect the rescue and proper replication of viral genomes during recombinant virus production, low-molecular weight DNA was extracted following a modified Hirt protocol [249]. Therefore, HEK293T cells were seeded in 6-well plates at a density of 5×10^5 cells/well. After 24 h, cells were PEI-transfected for production of rAAV2 or rAAV2/HBoV1 as summarized in **Table 16**. Cells were harvested 72 h post-transfection in 1 mL Hirt lysis buffer supplied with 12.5 µL proteinase K (20 mg/mL) and then incubated at 55°C for 3 h. Next, samples were put on ice for 10 min and 5 M NaCl was added dropwise to precipitate high-molecular weight DNA. The mixture was immediately inverted several times and incubated overnight at 4°C. The next day, the tubes were centrifuged for 1 h at 13,200 rpm and the supernatant was carefully

transferred to a new 1.5 mL tube. The solution was centrifuged again (13,200 rpm, 15 min) and stored at -20°C or further processed for low-molecular weight DNA isolation. Therefore, 700 µL of the supernatant were mixed with 700 µL (*i.e.* one volume) of phenol/chloroform/isoamyl alcohol and vortexed well. Then, the tubes were centrifuged for 5 min at 13,200 rpm and 600 µL of the upper aqueous phase were transferred to a new 1.5 mL tube containing 420 µL isopropanol. Tubes were then centrifuged again (30 min, 13,200 rpm) and the resulting DNA pellet was washed with 500 µL 70% ethanol. After a last centrifugation step (5 min, 13,200 rpm), the pellet was dissolved in 18 µL H_2O. To digest residual plasmid DNA, a DpnI digest was performed for 1 h at 37°C. The DNA was then either stored at -20°C or loaded on a neutral 1% agarose gel.

Table 16. PEI transfection protocol for small-scale AAV and BoV production.

Plasmid	**rAAV/BoV**	**rAAV/AAV2**	**Transfection mix (µL)**
rAAV vector	175 ng	850 ng	**DNA mix**
adeno-helper	-	850 ng	45 NaCl
pDGΔVP	1312 ng	-	45 H_2O
AAV-helper	-	850 ng	**PEI mix**
BoV-helper	1400 ng	-	45 NaCl
			27 H_2O
			22 PEI

2.2.3.10 Neutral agarose gel electrophoresis

Standard (neutral) gel electrophoresis was performed to separate DNA molecules by size for subsequent purification of defined fragments or visualization of restriction digest patterns. Therefore, the appropriate amount of agarose was melted in 1× TAE buffer and cooled down to 55°C to obtain 1% or 2% agarose gels. Ethidium bromide was added to the solution (1 µg/mL) before pouring into the gel chambers and inserting the comb. Samples were mixed with 6× purple loading dye (NEB) and loaded together with an appropriate size marker on the gel. Electrophoresis was performed at 90-120 V for 20-30 min, depending on the size and concentration of the gel. DNA was visualized on a Gel Doc XR system or a UV transilluminator. If DNA fragments had to be excised from the gel, UV exposure was minimized and the DNA was purified from the gel using the QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer's protocol. To enhance the elution of DNA from the columns, H_2O was preheated to 55°C and added directly to the filter. After an incubation period of 4 min, the columns were centrifuged at 13,000 rpm for 2 min.

2.2.3.11 Alkaline agarose gel electrophoresis

Alkaline agarose gel electrophoresis was used to denature and separate ssDNA molecules (rAAV genomes in this work).

For virus production, HEK293T cells were seeded in 150 cm^2 dishes and then PEI-transfected as described in section 2.2.5.2. Three days later, cells were harvested in 1 mL 50 mM Tris-HCl and subjected to five freeze-thaw cycles. To digest contaminating plasmid DNA, 540 µg DNase I, 5 µL 1 M $MgCl_2$ and 5 µL 1 M $CaCl_2$ were added and the cell lysate was incubated for 3 h at 37°C. To inactivate the DNase I, 10 µL 0.5 M EDTA and 20 µL 0.25 M EGTA were added and the cell lysate was incubated at 70°C for 20 min. Viral DNA was then purified using the DNeasy Blood & Tissue Kit (Qiagen) following the manufacturer's instructions. The isolated DNA was additionally digested with DpnI for 1 h prior to mixing with 6× alkaline loading dye and loading on 0.7% alkaline agarose gels. Gel electrophoresis was performed in alkaline buffer for ~6 h at 30 V.

2.2.3.12 Southern blot analysis of rAAV vector genomes.

Southern blot analysis was performed to detect DNA molecules not directly visible by ethidium bromide staining and UV light exposure.

Following alkaline or neutral gel electrophoresis, the gel was washed once with ddH_2O and immersed in 0.2 M HCl solution for 7 min. This improves the transfer of larger DNA fragments (ssAAV genomes and replication intermediates >5 kb). Next, the gel was denatured with gel-denaturation buffer on a rocking platform for 30 min, followed by immersion in gel-neutralization buffer (pH 7.5) for additional 30 min. Then, the Southern blot was assembled by first placing the gel (inverted) on a Whatman paper bridge soaked in 20× SSC and connected to a tank with 20× SSC solution. A nylon membrane was soaked in 2× SSC and placed on top followed by three Whatman papers soaked in 20× SSC. Finally, a pile of thick paper cut to size was placed on top and the transfer was performed O/N. The next day, the blot was disassembled, and the membrane soaked in 20× SSC for 20 min. Then, the membrane was air-dried and the DNA UV cross-linked for 2 min at 100 J/cm^2.

Random digoxigenin-labelled probes against SpCas9 or *yfp* were generated using the DIG Starter Kit II (Roche) according to the manufacturer's instructions. The membranes were incubated with the probes (~25 ng/mL) O/N at 40°C. Stringency washes, blocking and antibody incubation were performed according to the manufacturer's protocol. Blots were exposed to X-ray films and developed in an X-OMAT 2000 x-ray film processor.

2.2.3.13 Dot blot analysis for detection of viral genomes

Dot blot analysis was performed to identify the virus-containing fractions in cesium chloride (CsCl) density gradients. Per fraction, 15 µL were mixed with 189 µL Tris-HCl (50 mM,

pH 8), 1 µL 1 M $MgCl_2$, 1 µL 1 M $CaCl_2$, 2 µL 1 mg/mL DNase I and 2 µL 1 mg/mL RNase A, and the entire mixture was incubated for 30 min at 37°C. DNase I was inactivated by adding 2 µL 0.5 M EDTA, 4 µL 0.25 M EGTA and 10 µL 10% sarcosine, and by additionally heating the samples to 70°C for 20 min. Next, 10 µL proteinase K (20 mg/mL) were added and samples incubated O/N at 55°C. The next day, each sample was mixed with 40 µL 5 M NaOH, 20 µL 0.5 M EDTA and 225 µL H_2O, and incubated at 55°C for 10 min before loading on the Bio-Dot® Microfiltration Apparatus. A nylon membrane was cut to size, soaked in 6× SSC for 10 min and assembled into the dot blot device (Bio-Rad) according to the manufacturer´s protocol. The samples were then applied in two aliquots, followed by a final wash with 200 µL 2× SSC per well. Membranes were placed on a Whatman paper soaked in 10× SSC, and the DNA was cross-linked with a UV-cross linker for 2 min at 100 J/cm^2. A probe binding in the *yfp* transgene was generated using the DIG Starter Kit II (Roche) according to the manufacturer's instructions. The blots were pre-hybridized and incubated with the probe (~25 ng/mL) O/N. Blots were exposed to X-ray films and developed in an X-OMAT 2000 x-ray film processor. After background subtraction, dot blots were analyzed using Fiji software.

2.2.4 Cell culture methods

2.2.4.1 Culturing conditions

All standard cell lines (HEK293T, HeLa, Huh7, LX-2, Panc-I, MCF-7 and RAW 264.7) were cultured in Dulbecco's Modified Eagle's Medium (DMEM) with GlutaMAX™ supplemented with 10% fetal bovine serum (FBS) and 100 U/mL penicillin-streptomycin P/S (37°C with 5% CO_2 incubation). The T84 human colon carcinoma cell line was cultured in a 50:50 mixture of DMEM and F12 medium supplemented with 10% FBS and 1% penicillin.

The following primary cells were purchased from PromoCell and grown as monolayers: Saphenous vein endothelial cells, skeletal muscle cells, pulmonary fibroblasts and cardiac myocytes. Primary human hepatocytes were purchased from Cytes Biotechnologies. All primary cells were maintained following the supplier's instructions.

pHAE cells were kindly provided by Marc Schneider (Section Translationale Forschung STF, Thoraxklinik, University Hospital Heidelberg, Heidelberg, Germany) and maintained in DMEM/Ham's F-12 medium supplemented with Airway Epithelial Cell Growth Medium Supplement Pack and ROCK inhibitor.

PBMCs, primary T-cells and macrophages were isolated and purified at the Department of Infectious Diseases, Virology of the University Hospital Heidelberg, Heidelberg, Germany. Macrophages were maintained in RPMI 1640 with GlutaMAX™ medium supplemented with 10% FBS and 100 U/mL penicillin-streptomycin. T-cells and PBMCs were maintained under the same conditions but additionally supplemented with phytohemagglutinin and interleukin-2.

Human lung organoids were expanded by Jens Puschhof (Hubrecht Institute, Utrecht, the Netherlands) as previously described [250]. Differentiation was induced five days prior to viral vector microinjection through medium change as described earlier [250].

For experiments involving primary intestinal organoids, culturing of cells was performed by Megan Stanifer (Boulant lab, Department of Infectious Diseases, Virology of the University Hospital Heidelberg, Heidelberg, Germany). Briefly, human tissue was received from colon and small intestine resections from the University Hospital Heidelberg. Stem cells containing crypts were isolated following 2 mM EDTA dissociation of tissue sample for 1 hr at 4°C. Crypts were spun and washed in ice-cold PBS. Fractions enriched in crypts were filtered with 70 µM filters and observed under a light microscope. Fractions containing the highest number of crypts were pooled and spun again. The supernatant was removed, and crypts were re-suspended in Matrigel. Crypts were passaged and maintained in basal or differentiated state for 5 days in differentiation culture media.

2.2.4.2 Splitting of cells

The standard cell lines used in this work were passaged when they reached confluency (typically, 2-3 days after seeding). First, the medium was removed, and the cells were washed with 1× PBS. Trypsin-EDTA (0.25%) was added to the cells at a volume sufficient to cover the cells (1-2 mL), depending on the size of the flask. Cells were then incubated at 37°C for 2-10 min (depending on cell type) until all cells detached. Trypsin activity was stopped by adding at least 4 volumes of medium containing FBS. The mixture was pipetted up and down several times to obtain a single cell suspension. Cells were then diluted 1:5 or 1:10 (depending on cell type) in 30 mL (for 175 cm^2 flasks) or 10 mL (75 cm^2 flaks) of growth media. For suspension cultures, addition of trypsin-EDTA was not needed. Instead, cells were suspended well with a pipette and directly diluted in fresh media.

To determine the number of cells per mL, a 10 µL aliquot was taken and mixed with 10 µL Trypan blue, which stains dead cells. Next, 10 µL of the mix were transferred into a Countess cell counting chamber slide and cells were automatically counted using the Countess automated cell counter.

2.2.4.3 Transfection of cells with Lipofectamine 2000

For luciferase assays, HEK293T cells were seeded in 96-well plates at a density of 3×10^4 cells per well. For Western blot analysis, HEK293T cells were grown in 24-well plates at a density of 1.25×10^5 cells per well. Transfections were performed using Lipofectamine 2000 (Thermo Fisher Scientific) following the manufacturer's instructions. Typically, a total of 100 or 800 ng DNA was used per well for 96- or 24-well plates, respectively. From the luciferase reporter constructs (including the SIN variants), 2 or 6 ng of DNA per well were transfected for

96- or 24-well format, respectively. Cells were harvested at 24, 48 or 72 h post-transfection (as indicated in the respective figure legends).

2.2.4.4 Measurement of *Gaussia* luciferase (GLuc) expression

GLuc secretion in the cell medium was determined by using a GloMax96 microplate luminometer equipped with an automatic injector. Twenty µL of the sample medium were pipetted into each well, and the machine injected 100 µL *Renilla* luciferase quenching buffer supplied with coelenterazine at a final dilution of 11.7 µM. All parameters (speed of injection and integration / interval time) were set as default.

2.2.4.5 Measurement of Firefly (FLuc) and *Renilla* (RLuc) luciferase activities

HEK293T cells were cultured in 96-well plates and transfected as described in section 2.2.4.3. After 72 h, cells were lysed in passive lysis buffer (1:5 diluted in H_2O) on a rocking platform for 20 min. From each lysate, 10 µL were pipetted into a Lumitrac 96-well microplate and a dual-luciferase assay was performed in a GloMax96 microplate luminometer. The automatic injectors first pipetted 50 µL of the Firefly luciferase assay buffer (supplemented with luciferin). Subsequently, FLuc activity was estimated by the amount of emitted light units. Next, Firefly luciferase activity was quenched with the *Renilla* luciferase assay buffer (supplemented with coelenterazine) to measure the activity of RLuc.

2.2.5 Virological methods

2.2.5.1 Small-scale virus production

To produce rAAV vectors in small-scale or as “crude lysates”, 5×10^5 HEK293T cells per well were seeded in 6-well plates and cultured in DMEM containing 10% FBS and 1% P/S one day before transfection. After 24 h, the PEI-DNA transfection mixes were prepared as depicted in **Table 16.** After combining the DNA and PEI mixes, the solution was vortexed well and the transfection mix was added dropwise to the cells. The plate was gently shaken and placed at 37°C. After 72 h, the medium was removed, and the cells were harvested in 1 mL PBS by pipetting up and down using a P1000 pipette. Next, the cells were centrifuged in 1.5 mL microcentrifuge tubes at 800×g for 10 min at 4°C. The PBS was then discarded and the cell pellet resuspended in 300-500 µL PBS.

To release the viral particles, the cell suspension was subjected to five freeze-thaw cycles by alternating between liquid nitrogen and a 37°C water bath. The crude cell lysate was centrifuged again (13,200 rpm, 4°C) for 10 min to remove cell debris. The vector-containing supernatant was transferred to a 1.5 mL microcentrifuge tube and used directly to transduce cells or stored at -80°C.

2.2.5.2 Large-scale virus production

For large-scale production of purified AAV or BoV vectors, HEK293T cells were expanded in a sufficient number of 175 cm^2 flasks (7×10^6 cells per flask). After 48 h, the medium was completely removed, and the cells were harvested and counted as described in section 2.2.4.2. Next, 4×10^6 cells were seeded per 150 cm^2 dish in 22 mL medium. The plates were then further incubated for 48 h at 37°C to reach the optimal density (70-80%) for transfection. PEI-DNA transfection mixes were prepared as depicted in **Table 17** and added dropwise to the cells. After 72 h, the cells were harvested in the medium using a cell scraper and were transferred into 50 mL Falcon tubes or 500 mL centrifuge buckets (depending on the final volume of each virus preparation). Next, the cells were pelleted in a Beckman Avanti® centrifuge at 800×g for 15 min. The medium was completely discarded and the cell pellet(s) resuspended in 5 mL benzonase or Tris-HCl buffer for small gradients (up to 10 plates of cells) or in 20 mL for larger gradients (up to 50 plates). The cell suspension was either stored at -80°C or further processed to obtain the crude cell lysate. Therefore, the suspension was vortexed well and subjected to five freeze-thaw cycles by alternating between liquid nitrogen and 37°C. To remove residual and unwanted plasmid DNA, benzonase (in benzonase buffer, 50 U/mL) or DNase I (in Tris-HCl buffer, 270 µg per plate) was added and the mixture was incubated at 37°C for 1 h, while inverting the tubes every 15 min. To remove cell debris, samples were centrifuged twice for 15 min at 4,000×g and 4°C, always saving the virus-containing supernatant. The virus was purified using gradient ultracentrifugation or frozen at -20°C.

Table 17. PEI transfection protocol for large-scale AAV and BoV production.

Plasmid	**rAAV/BoV**		**rAAV/AAV**	**Transfection mix (mL)**
(µg)	3 components	4 components		**DNA mix**
rAAV vector	3.2	2.87	14	0.8 NaCl
Adeno-helper	-	19.97	14	0.8 H_2O
pDGΔVP	24	-	-	**PEI mix**
AAV-helper	-	-	14	0.8 NaCl
AAV2-rep	-	6.07	-	0.4 H_2O
BoV-helper	18	15.19	-	0.4 PEI

2.2.5.3 Chimeric BoV/AAV purification using CsCl density gradients

HEK293T cells were expanded and seeded as described in sections 2.2.4.1 and 2.2.4.2. PEI transfection was performed as described above (section 2.2.5.2). Cells were harvested for virus purification in 10 mM Tris-HCl (pH 8) and subjected to five freeze-thaw cycles. Plasmid DNA was removed by adding 270 µg of DNase I per plate to the cell lysate (for 1 h, 37°C). To optimize cell lysis and to increase virus yields, 1.5 mL 0.25% Trypsin-EDTA and 1.5 mL 10%

sodium deoxycholate were added for 30 min at 37°C. Next, 13.2 g of CsCl were dissolved in the mixture and vortexed well. To assist the dissolving of CsCl, the mix was incubated at 37°C for 15-30 min. Then, the cell debris was removed by centrifugation at 3,000×g for 30 min. The refractive index (RI) of the supernatant was measured and adjusted to 1.372 by adding Tris-HCl or CsCl if the initial value was too high or too low, respectively. The clarified lysate was then transferred to a Beckman ultracentrifugation tube (26×77 mm, 29.9 mL capacity) and centrifuged at 35,000 rpm for 32 h at 4°C. After the completion of the run, the gradients were fixed, and the bottom of each tube was punctured with a 21G needle. To permit air inflow, a second needle was inserted at the top corner of the gradient. With this second needle, the flow of the gradient could be adjusted from fast at the beginning (collection of 2 mL for the first fraction) to slow afterwards (collection of ~30-40 fractions per gradient, each 0.5 mL). The RI of each fraction was measured and the virus-containing fractions (determined by dot blot analysis, section 2.2.3.13) were pooled and pipetted into a Beckman ultracentrifugation tube (16×67 mm, 11.2 mL capacity) to perform a second CsCl gradient by repeating the above-mentioned steps. After completion of the second round, the collected virus-containing fractions are pipetted into a dialysis cassette. The cassettes were then transferred into a beaker and dialyzed against 1 L 1× PBS for 30 min at RT. The PBS was exchanged after 1 h and the dialysis was performed at 4°C on a magnetic stirrer (lowest speed of stir bar spin). After an additional hour, the buffer was changed again and then left O/N. The following day, PBS was changed twice every 2 h, followed directly by the concentration of the sample. Therefore, an Amicon Ultra-15 centrifugal filter unit was washed 2× with 15 mL PBS before the vector-containing solution from the dialysis cassette was applied. Several centrifugation steps at 400-2,000×g allowed for the concentration of the viral vector to ~1,000 µL. The vector was aliquoted and stored at -80°C.

2.2.5.4 Virus purification using iodixanol gradients

Beckman ultracentrifugation tubes (14×89 mm for small gradients or 25×89 mm for large gradients) were arranged in a holder and a sterile Pasteur pipette was inserted into each. The vector cell lysates described in section 2.2.5.2 were pipetted through the Pasteur pipette into the tubes. Next, the different iodixanol solutions were applied in the following order, taking care to avoid air bubbles at any stage:

- 1.5 mL (small) or 7 mL (large) 15% iodixanol solution
- 1.5 mL (small) or 5 mL (large) 25% iodixanol solution
- 1.5 mL (small) or 4 mL (large) 40% iodixanol solution
- 1.5 mL (small) or 4 mL (large) 60% iodixanol solution

In cases where the tubes were not completely full, benzonase buffer was carefully added using a 3 mL syringe and 19G needle to top off the tube. The tubes were closed with metal

caps and sealed using a Beckman tube sealer. Next, the tubes were centrifuged at 4°C for 2 h at 230,000×g in a Beckman 70.1 Ti rotor (small gradients) or at 257,000×g in a Beckman 70 Ti rotor (large gradients). The acceleration was set to “max” and the deceleration to “slow” to allow a gentle stop and prevent the mixing of the phases. After centrifugation, the tubes were fixed well in an appropriate holder and a 21G needle was inserted at the top of the tube to allow inflow of air. A second 21G needle attached to a 3-5 mL syringe was carefully inserted below the 60-40% interface. Then, 1 mL (for small gradient) or 2.5 mL (for large gradients) was collected from the 40% phase, taking care not to touch the 25-40% interface that contains the empty capsids. The resulting vector stocks were either stored in aliquots at -80°C or further processed to exchange the iodixanol with PBS. This was achieved by mixing the virus in iodixanol with 13 mL PBS and applying the mixture onto Amicon Ultra-15 centrifugal filter units. After each centrifugation step, the Amicon tubes had to be inverted several times to mix the iodixanol. Depending on the starting amount of iodixanol (1-2.5 mL), two to three washing and concentration steps were needed as previously described (section 2.2.5.3).

2.2.5.5 TaqMan RT-PCR titration of purified virus

To assess the number of genome-containing particles (=vector titer), quantitative (q)PCR analysis (probe-based) was used with appropriate primers binding in the transgene or promoter region (**Table 18**). The vector stocks were inverted several times before a 10 µL sample was mixed with 10 µL Tris-EDTA (TE) buffer and 20 µL 2 M NaOH, vortexed well, and incubated at 55°C for 30 min. This step denatures the viral particles and releases the encapsidated viral DNA. Next, the solution was neutralized by adding 38 µL 1 M HCl. At this step, the samples were either stored at -80°C or further processed for subsequent qPCR analysis. Therefore, the mixtures were diluted by adding 922 µL H_2O to obtain a 1:100 dilution. For virus stocks in iodixanol, an additional 1:10 dilution was needed because iodixanol can interfere with the absorbance in the qPCR process.

Each sample was pipetted in triplicates. The master mix for each triplet contained: 17.5 µL SensiMix™ II probe kit, 1.4 µL of each forward and reverse primer (10 µM), 0.35 µL probe (10 µM), 9.35 µL H_2O and 5 µL sample (=1:7 dilution). To estimate the viral titer (genome copies/mL), a standard curve with known amounts of appropriate plasmid DNA was used. For this, a stock containing 3.5×10^{11} molecules per mL of the AAV vector plasmid used for transfection was prepared. From this stock, serial 1:10 dilutions ranging from 3.5×10^{11} down to 3.5×10^{6} molecules per mL were made.

Samples and standards were measured on a Corbett RG6000 qPCR instrument using the following conditions: an initial heating step of 10 min at 95°C to activate the polymerase followed by 40 cycles of (i) denaturation for 10 sec at 95°C and (ii) a combined annealing and elongation step for 20 sec at 60°C. To create the standard curve, the Ct values of the standard

samples were plotted on the y-axis and the logarithm of the number of molecules per mL (concentration) on the x-axis. The resulting linear equation of the standard curve and the Ct values of the samples was then used to determine the concentration. As the samples had been diluted 1:100 (or 1:1,000 for iodixanol-purified vectors) with H_2O after neutralization and 1:7 during the preparation of triplicates, the concentration values were multiplied with 700 or 7,000. For the titration of ssAAV constructs, this value was further multiplied by two, because the plasmid used to create the standard curve is double-stranded.

Table 18. Primer and probe sets used for qRT-PCR.

Primer/Probe sets	**Sequence 5'- 3'**
NP1_fwd	GCACAGCCACGTGACGAA
NP1_rev	TGGACTCCCTTTTCTTTTGTAGGA
BoV_NP1_Probe	FAM-TGAGCTCAGGGAATATGAAAGACAAGCATCG-BHQ1
CMVenh_fwd	AACGCCAATAGGGACTTTCC
CMVenh_rev	GGGCGTACTTGGCATATGAT
CMVenh_Probe	FAM-CGGTAAACTGCCCACTTGGCAGT-BHQ1
YFP_fwd	GAGCGCACCATCTTCTTCAAG
YFP rev	TGTCGCCCTCGAACTTCAC
YFP_probe	FAM-ACGACGGCAACTACA-BHQ1
hluc+ fwd	CGCCCGCGACCCTATTTTCG
hluc+ rev	CAGGTAGCCCAGGGTGGTGAAC
hluc+ probe	FAM-AACCAGATCATCCCCGACACCGCTATTCTGAGCGT-BHQ1
SaCas9_fwd	CCGCCCGGAAAGAGATTATT
SaCas9_rev	CGGAGTTCAGATTGGTCAGTT
SaCas9_probe	FAM-AGCTGCTGGATCAGATTGCCAAGA- BHQ1
TTR fwd	TGTTCCGATACTCTAATCTCCC
TTR rev	TATACCCCCTCCTTCCAACC
TTR probe	FAM-TTTGGAGTCAGCTTGGCAGGGATCA-BHQ1
miCMV_fwd	GCACCAAAATCAACGGGAC
miCMV_rev	AGCAGGCTCTTTCGATCAC
miCMV_probe	FAM- TTCCAAAATGTCGTAATAACCCCGCCCCG -BHQ1

2.2.5.6 Transductions

For transduction of cell lines and primary cells in monolayers, cells were seeded in 96-well plates one day prior to transduction at densities of 5×10^3 or 1×10^4 cells per well.

The next day, the different vectors were added at MOIs ranging between 2×10^4 and 2×10^5 for scAAV-YFP/BoV or 1×10^4 to 5×10^4 for scAAV-GLuc/BoV (as indicated in the respective figure legends). Transductions with rAAV/BoV were always performed in the presence of 0.5-1 µM doxorubicin. Viruses and inhibitors were left on the cells O/N and the medium was replaced on the next day. Cell lines were further incubated for a maximum of 72 h and primary cells up to 12 days (as indicated for each experiment).

Likewise, transductions of pHAE with rAAV/BoV were performed in the presence of proteasome inhibitors (5 µM doxorubicin and 40 µM Calpain inhibitor I). Usually transductions were performed for ~16 h by applying the vectors to the apical side of the transwell. For basolateral transduction, the transwell inserts were flipped and scAAV-GLuc/BoV particles were directly added to the surface of the filter at a MOI of 2×10^4. After 1 h at 37°C, the recombinant virus was removed, and inserts were inverted again into the medium. In both cases, cells were further incubated with the two proteasome inhibitors O/N. The next day, the medium was replaced on the basolateral side with fresh medium without any inhibitors and transgene expression was assessed over the following 12 days.

For transduction of lung organoids, luminal access was provided to the viral vectors by either mechanical shearing or direct microinjection. Mechanical shearing was performed by first detaching the organoids from the well and then resuspending them in culture medium using a flame-narrowed glass pipette. Next, 5×10^9 scAAV-GLuc/BoV particles and three volumes of Cultrex® growth factor reduced BME type 2 were added to the culture medium and mixed by pipetting. After solidifying for 30 min, 1 mL of differentiation medium [250] with 1 µM doxorubicin was added. Organoids were then kept in culture for 12 days and assessed for transgene expression.

To transduce intestinal organoids, 5×10^9 scAAV-GLuc/BoV or 1×10^{11} scAAV-YFP/BoV particles were added to each well and incubated overnight in the presence of 0.5 µM doxorubicin. The next day, the media was changed. For transductions with scAAV-GLuc/BoV, aliquots were taken at various time points for luciferase activity measurements.

For transduction experiments with SaCas9-encoding vectors, HEK293T cells were co-transduced with AAV2 particles carrying a full-length or split SaCas9 at a MOI of 1×10^5 per construct and the KS-luciferase or control reporter at a MOI of 1×10^4.

2.2.6 Protein biochemistry methods

2.2.6.1 Protein amount measurement in cell lysates

Protein amounts were measured in cell lysates prior to SDS-polyacrylamide gel electrophoresis using the Pierce BCA Protein Assay Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. The BSA protein standard dilutions were prepared in RIPA

buffer. Samples and standards were measured on a TECAN plate reader. To obtain measurements within the linear range, a 1:5 dilution of the samples was prepared.

2.2.6.2 SDS-Polyacrylamide Gel Electrophoresis (PAGE)

SDS-PAGE was used to separate proteins according to their molecular masses. Mini-PROTEAN precast polyacrylamide gels (Bio-Rad) were used according to the manufacturer's instructions. For the self-made gels used in this work, Bio-Rad cassettes were employed to assemble the SDS-PAGE gels. First, 8-12% resolving gel components were mixed as indicated in **Table 19**. The solution was poured between the two glass plates and then covered with isopropanol until complete polymerization. Next, the stacking gel components were mixed (as shown in **Table 19**) and poured onto the resolving gel (after removal of the isopropanol). A comb was inserted, and the gels were either stored at 4°C or loaded directly with extracted proteins and a protein ladder (PageRuler). Electrophoresis was performed in 1× TBS buffer at 90 V for ~2 h (until the loading dye front leaves the gel).

Table 19. Components for SDS gel electrophoresis.

Resolving Gel (5 mL)	8%	12%	**Stacking Gel (3 mL)**	5%
ddH_2O	2.65 mL	2.15 mL	ddH_2O	2.65 mL
Rotiphorese® Gel 40 (19:1)	1 mL	1.5 mL	Rotiphorese® Gel 40 (19:1)	1 mL
1.5 M Tris HCl (pH 8.8)	1.25 mL	1.25 mL	0.5 M Tris HCl (pH 6.8)	1.25 mL
10% SDS	50 µL	50 µL	10% SDS	50 µL
1% APS	50 µL	50 µL	1% APS	50 µL
TEMED	3 µL	2 µL	TEMED	3 µL

2.2.6.3 Western blotting

After SDS-PAGE electrophoresis, proteins were transferred onto a nitrocellulose membrane using semi-dry Western blot transfer. Therefore, Whatman paper and a nitrocellulose membrane were soaked in transfer buffer and a sandwich was assembled in a Trans-Blot® SD Semi-Dry Transfer Cell in the following order: Three pieces of Whatman paper, one nitrocellulose membrane, polyacrylamide gel, three additional pieces of Whatman paper. A falcon tube was used to roll out any entrapped air bubbles. Blotting was performed at 4°C, 150 mV for 1 h. Next, the blot was disassembled, and the proteins were reversibly stained with Ponceau S to verify proper transfer. The membrane was washed in 1× TBST to remove the Ponceau dye, followed by blocking in 5% milk for 1 h at RT. Primary antibodies were diluted in 5% milk (as indicated in **Table 20**) and incubated with the membrane at 4°C O/N. The next day, the membrane was washed three times with TBST, and an appropriate horseradish peroxidase-conjugated secondary antibody was added for 1 h at RT (see **Table 20** for a list of antibodies used and their dilutions).

To visualize protein bands, Western Lightning Plus-ECL reagent was used and the emitted signal was detected with (i) a chemiluminescent imager (Intas ChemoStar) or (ii) X-ray films, which were developed in an X-OMAT 2000 X-ray film processor.

2.2.6.3.1 Detection of Capsid proteins in purified virus stocks

From each purified virus stock (iodixanol or CsCl), 4-10 µL were mixed with an equal volume of 2× SDS sample loading buffer and boiled for 5 min at 95°C. Then, the samples were separated on 8% SDS-PAGE gels and transferred onto a nitrocellulose membrane using semi-dry transfer. Membranes were blocked with 5% milk for 1 h at RT and incubated O/N with an anti-VP polyclonal primary rabbit antibody recognizing the three bocaviral capsid proteins VP1, VP2 and VP3. For detection, a horseradish peroxidase-conjugated secondary donkey anti-rabbit antibody was used (see **Table 20**).

2.2.6.3.2 Detection of SaCas9 after plasmid transfection or transduction with rAAV vectors

To measure the expression level of SaCas9 under different experimental conditions, HEK293T cells were seeded in 24-well plates at a density of 1.25×10^5 cells per well. After 24 h, cells were either transfected or transduced with full-length or split SaCas9 expression constructs as described in sections 2.2.4.3 and 2.2.5.6, respectively.

In both cases, cells were harvested in PBS at different time points (as indicated in the respective figure legends) and centrifuged at 500×g for 10 min at 4°C. PBS was removed and cells were gently resuspended in 20 µL PBS. Next, RIPA buffer with protease inhibitor cocktail (1:25) was added to the cells (to a finale volume of ~50 µL) and the mixture was incubated on ice for 15 min. Tubes were then centrifuged again at 10,000 rpm to remove remaining cell debris and the supernatants were transferred to new 1.5 mL microcentrifuge tubes. From these supernatants, a small aliquot was diluted 1:5 in RIPA buffer and subsequently used for measurement of protein amounts as previously described (section 2.2.6.1). Next, 1 volume of 2× SDS buffer was added to the lysates and the solution was boiled at 95°C for 5 min. The lysates were then cooled down and either stored at -20°C or loaded on SDS polyacrylamide gels followed by Western blot analysis.

Table 20. List of antibodies and their dilutions.

Primary Antibodies	Dilution	Origin
Anti-VP polyclonal primary rabbit antibody	1:1,000	Kind gift from Maria Söderlund-Venermo, University of Helsinki (Helsinki, Finnland)

S. aureus CRISPR/Cas9 mouse monoclonal antibody	1:4,000	Diagenode (Liège, Belgium) #C15200230
(FITC)-coupled goat polyclonal antibody against GFP/YFP	1:1,000	Novus Biologicals (Littleton, USA) #NB100-1771
Anti-β-Actin mouse monoclonal antibody	1:300	Santa Cruz Biotechnology (Dallas, USA) #C0916
Anti-Cytokeratin 5 mouse monoclonal antibody	1:50	Santa Cruz Biotechnology (Dallas, USA) #RCK103
Anti-Mucin 5AC mouse monoclonal antibody	1:100	Abcam (Cambridge, UK) #ab3649
Anti-CC16 rabbit polyclonal antibody	1:200	BioVendor (Heidelberg, Germany) #RD181022220-01
Anti-β-Tubulin IV mouse monoclonal antibody	1:100	Merck (Darmstadt, Germany) #T7941
Secondary Antibodies		
AF-645 donkey anti-rabbit	1:2,000	Dianova (Hamburg, Germany) #711-605-152
AF-647 goat anti-mouse	1:2,000	Thermo Fisher Scientific (Waltham, USA) #A21235
AF-488 rabbit anti-goat	1:400	Thermo Fisher Scientific (Waltham, USA) #A11078
HRP-conjugated donkey anti-rabbit antibody	1:10,000	GE Healthcare (Chicago, USA) #NA934V
HRP-conjugated goat anti-mouse antibody	1:10,000	Dianova (Hamburg, Germany) #115-035-068

2.2.6.4 Immunostaining and microscopy

Cells transduced with recombinant viruses were fixed with 4% PFA for 15 min at various time points post-transduction (as indicated for each experiment). The cells were then washed three times with PBS and were permeabilized with 0.1% Triton X-100 for 15 min. Next, 3% BSA in PBS was added to the cells, and the plates were blocked for 1 h in the dark. To enhance the YFP signal, 100 to 200 µL of 1:1,000 diluted primary Fluorescein isothiocyanate (FITC)-coupled anti-GFP antibody were added to each well. After O/N incubation at 4°C, plates were washed three times with PBS before a secondary AF-488 rabbit anti-goat antibody was added for 1 h at RT (see **Table 20** for a list of antibodies used). Plates were washed again three times with PBS and Hoechst was added at a 1:3,000 dilution to the PBS from the last wash. Finally, plates were stored with 200 µL PBS per well in the dark at 4°C until microscopy analysis.

Microscopy pictures were taken with an Olympus inverted fluorescence microscope IX-81 and processed using Fiji.

2.2.6.5 Flow cytometry analysis

The apical surface of pHAE cells was washed with PBS for 20 min at 37°C to remove cell-associated mucus. Cells were then detached by adding 0.25% Trypsin/EDTA to the apical side for 10-15 min at 37°C. Trypsinization was stopped by adding 1% BSA/PBS, and four transwells (from different donors) were pooled and centrifuged at 400×g for 15 min. Next, cells were washed twice with PBS and resuspended in PBS to a final concentration of 1×10^6 cells/mL. To distinguish live from dead cells, 1 µL of Fixable Aqua Dead Cell Stain was added to the cell suspension (for 1×10^6 cells) for 30 min in the dark at 4°C. Cells were washed twice with PBS and then fixed with 4% PFA for 15 min at RT, followed by permeabilization with 1% Triton X-100 for 20 min. Cells were washed twice again with 0.5% BSA/PBS, and a FITC-coupled antibody against GFP/YFP was added to all samples (including the negative controls) to enhance the signal in YFP-expressing cells or to estimate the background (in the controls).

To stain the different cell types in the airway epithelia, primary antibodies against the following markers were used: (i) goblet cell marker MUC5A/C, (ii) ciliated cell marker β-Tubulin IV, (iii) basal cell marker Cytokeratin5, or (iv) Clara (club) cell marker CC16 at the dilutions listed in **Table 20**. The primary antibodies were incubated for 1 h at 4°C, followed by treatment with secondary AF-647 goat anti-mouse or AF-645 donkey anti-rabbit antibody for 30 min at RT. Cells were measured in a FACSVerse and analysis was performed using Flowing Software (section 2.1.7.4). Only living cells were used for the analysis and gates were placed according to the following control conditions: (i) cells transduced with scAAV-GLuc/HBoV1 and then stained with FITC-coupled anti-GFP antibody, the different cell-specific primary antibody and the respective secondary antibody, (ii) cells transduced with scAAV-YFP/HBoV1 and then stained with each secondary antibody alone and FITC-coupled anti-GFP antibody, or (iii) cells transduced with scAAV-YFP/HBoV1 and only stained with FITC-coupled anti-GFP antibody.

3 RESULTS

3.1 Packaging of rAAV genomes into the HBoV1 capsid

3.1.1 Development of a new streamlined protocol for bocaviral vector production

In the original report of rAAV2/HBoV1 vectors by Yan *et al.* [10], rAAV genomes were pseudotyped with the HBoV1 capsid by co-transfecting four plasmids into HEK293T cells (**Figure 8A**, plasmids 1 to 4). Proteins that were expressed from these plasmids include the Rep proteins from AAV2 and the structural/non-structural proteins from HBoV1 (NS, NP1 and VP1-3). Moreover, a plasmid encoding the adenovirus helper genes was supplied, which is strictly required for rAAV replication and protein expression. The resulting viral vectors were purified from the cell lysates using two rounds of CsCl density ultracentrifugation. This process results in stocks of high purity but is laborious and time-consuming (2-3 days). Therefore, we tested whether the faster process of iodixanol gradient centrifugation (2.5 h) that is commonly used for rAAV purification irrespective of serotype can also be applied to rAAV/BoV vectors (**Figure 8A**, right). Indeed, we found that the majority of viral particles (up to ~70%) accumulated in the 40% iodixanol phase, as detected by TaqMan qPCR analysis of the different iodixanol fractions (**Figure 8B**).

Two of the plasmids used for the generation of pseudotyped rAAV2/HBoV1 particles (plasmids 2 and 3 in **Figure 8A**) are major components of the rAAV production system. In essence, the standard AAV helpers combine the AAV *rep* and *cap* ORF on one plasmid, which is then co-transfected with the rAAV vector and adenoviral helper plasmids. Notably, the Kleinschmidt lab reported a series of different AAV helper plasmids that express the AAV and adenoviral helper functions from a single backbone. This includes pDG [135], a plasmid harboring the whole AAV2 genome except for the AAV2 ITRs, as well as pDGΔVP that was derived from pDG by deleting the AAV2 *cap* ORF [137]. The latter was interesting for the chimeric rAAV2/HBoV1 production system because it combines two plasmids (2+3) in one and hence reduces the total number of plasmids needed. Moreover, in pDGΔVP, the AAV2 p5 promoter was replaced with the weaker MMTV-LTR promoter, which results in reduced levels of Rep expression. This was found to result in increased rAAV titers [135] and to favor encapsidation of intact scAAV genomes [251]. These advantages tempted us to perform a side-by-side comparison of the 1-plasmid (pDGΔVP, plasmid 2+3) and 2-plasmid (plasmids 2 and 3) AAV helpers for rAAV2/HBoV1 production. Accordingly, we packaged a scAAV-YFP genome into the HBoV1 capsid using either the two separate helpers (standard 4-plasmid system) or the

combined helper pDGΔVP (3-plasmid system) and purified the resulting vectors using iodixanol gradient centrifugation. The two approaches resulted in largely comparable viral yields in a range of 5×10^9 to 1×10^{10} vector genomes per mL from five 15 cm plates of HEK293T cells (**Figure 8C**). These titers are consistent with previous reports that the original 4-plasmid production system gives yields reaching 10% of rAAV vectors [10].

Moreover, we tested the infectivity of iodixanol-purified scAAV-YFP/HBoV1 stocks. Therefore, pHAE were transduced at a MOI (multiplicity of infection, *i.e.*, viral particle number per cell) of 2×10^4 and transgene expression was monitored over time (**Figure 8D**). YFP expression became visible two days post-transduction and the efficiency was largely comparable or even higher than the one obtained after transduction with CsCl-purified vectors (see section 3.2.5).

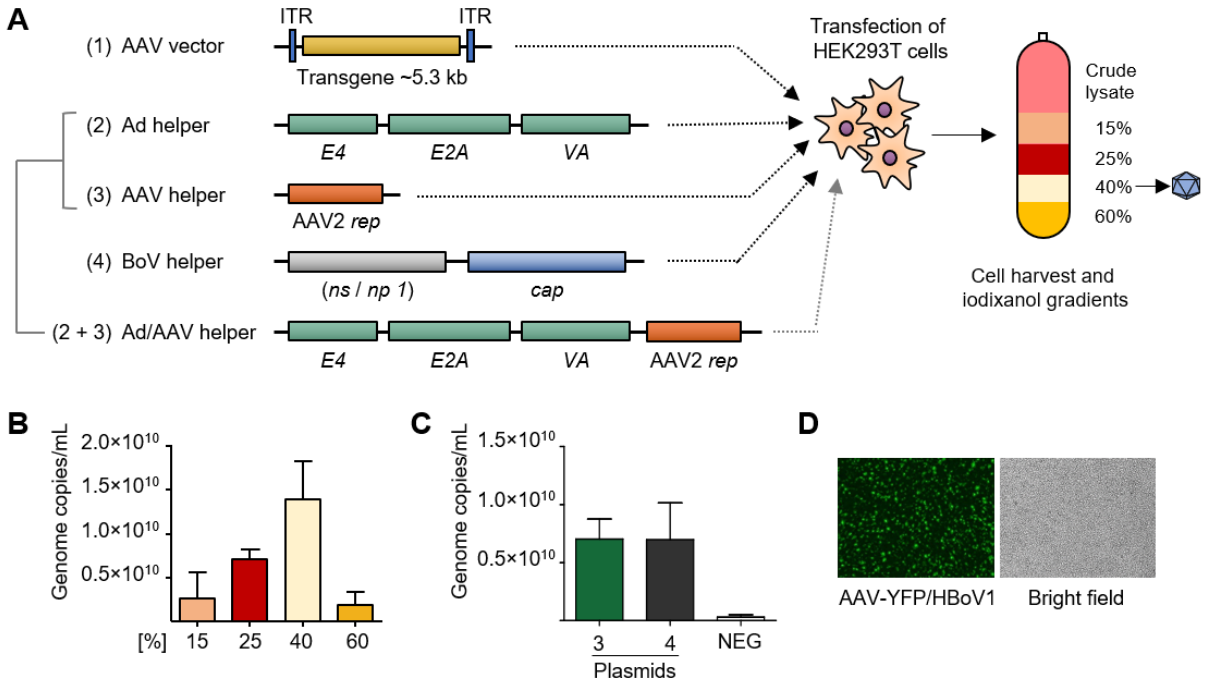


Figure 8. Triple-plasmid transfection and iodixanol purification of rAAV/BoV vectors.

(A) Plasmids for rAAV2/HBoV1 vector production using the 4-plasmid (1 to 4) or 3-plasmid (1, 2+3, 4) production systems. Plasmids are co-transfected into HEK293T cells and the viral vectors in the cell lysates are purified with iodixanol gradient centrifugation. The 40% iodixanol phase contains the full (*i.e.*, DNA-containing) viral particles. **(B)** Different iodixanol fractions were analyzed for the presence of genome-containing particles (mean ± SD; n = 3) using TaqMan qPCR analysis. **(C)** Comparison of the 3- and 4-plasmid transfection protocols. Viruses were purified using iodixanol gradient centrifugation and the amount of genome copies per mL (mean ± SD; n = 3) was determined by TaqMan qPCR analysis. NEG = H_2O control. **(D)** Transduction of pHAE with 2×10^{10} genome copies of scAAV-YFP/HBoV1 purified using iodixanol gradient centrifugation. Transduction was performed in the presence of LLnL and doxorubicin at concentrations of 40 µM and 5 µM, respectively. Images were taken at day 15 post-transduction. Magnification = 10×. Panels A-C were published and adapted from Fakhiri *et al.* [252].

3.1.2 Analysis of rAAV2/HBoV1 packaging capacity using single-stranded or self-complementary rAAV vectors

One therapeutically relevant feature of HBoV1 is the larger genome size of ~5.5 kb as compared to AAV, which only packages genomes of ~4.7 kb in size. Yan *et al.* [10] harnessed this property in the rAAV2/HBoV1 vector system to package an oversized ssAAV genome carrying a full-length CFTR cassette under a strong chicken β-actin (CBA) promoter. In addition, other previous reports have shown that many autonomous parvoviruses can package genomes that exceed the wt length to a certain level [83, 253].

Here, we aimed (i) to validate the reported packaging capacity of HBoV1 and (ii) to determine the maximal genome size that allows efficient packaging with the fewest truncations possible. To this end, we designed a set of oversized ssAAV genomes encoding all-in-one CRISPR constructs. They included two components, namely, the SpCas9 from *Streptococcus pyogenes* and a gRNA, which were expressed from different RNA Pol II and Pol III promoters, respectively (**Figure 9A**). The resulting genome sizes varied between 5.1 and 6.1 kb, corresponding to 92-110% of the 5.5 kb wtHBoV1 genome, or 109-130% of the 4.7 kb wtAAV2 genome, with gradual increases between 300 and 400 bp. These rAAV-CRISPR constructs are very attractive tools for gene editing but have been shown to represent a challenge for standard rAAV vectors, due to the limited packaging capacity of AAV [209, 254]. Indeed, packaging of these genomes into AAV2 particles revealed encapsidation only of the smallest genome tested (5.1 kb), consistent with previous reports [254-256] (**Figure 9B**, upper gel). All the other, larger genomes did not result in distinct bands at the expected heights, but produced a smear of low-molecular-weight DNA, most likely representing truncated genomes, again congruent with prior reports [254, 255]. To exclude any defects in the rescue and replication of the viral genomes during vector production, we performed and analyzed Hirt DNA extracts (**Figure 9B**, lower gel). As hoped for, monomeric (M) and dimeric (D) ssAAV replication forms could be detected for all the tested constructs, ruling out deficiencies at the steps prior to encapsidation.

In parallel, we analyzed the capacity of AAV2 to package oversized scAAV genomes. Therefore, we cloned stuffer sequences of various lengths, *i.e.* 100-300 bp (in 100 bp increments) and 500-1600 bp (in 300-400 bp increments), into the C-terminal end of the *yfp* ORF in our standard scAAV backbone (**Figure 9C**). The resulting genome lengths varied between 2.0 and 3.6 kb, corresponding to 72-131% of the wtHBoV1 genome, or 85-153% of the wtAAV2 genome. Then, we again packaged all the different genomes into AAV2 capsids and analyzed the encapsidated viral DNA using neutral gel electrophoresis followed by Southern blotting (**Figure 9D,** upper gel). Our results show that scAAV genomes larger than 2.8 kb could not be packaged into AAV2 capsids. Notably, it was shown that the maximum packaging size of scAAVs can be further pushed to 3.2 kb, by optimizing the genome sequence

and structure [251], as discussed later in more detail (section 4.1.1) We also confirmed proper rescue and replication of all oversized scAAV genomes (**Figure 9D,** lower gel).

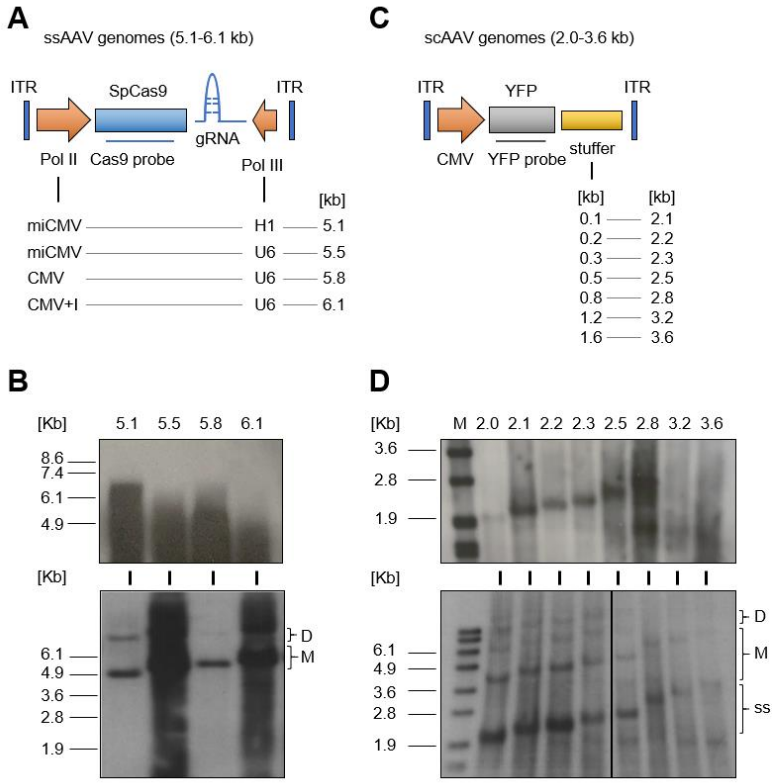


Figure 9. Packaging of oversized ssAAV and scAAV genomes into the AAV2 capsid.

(A) Schematic representation of the ssAAV-CRISPR genomes used in this work. Different combinations of the indicated RNA Pol II and Pol III promoters were used, which resulted in the total genome sizes listed in the third column. **(B)** Upper gel: Alkaline gel electrophoresis and Southern blot analysis of the indicated ssAAV-CRISPR genomes, which were packaged into, and isolated from, AAV2 particles. Lower gel: Southern blot analysis of low-molecular-weight (Hirt) extracts corresponding to the indicated constructs. **(C)** Illustration of the scAAV genomes used in this work. Stuffer sequences of different lengths were cloned downstream of the *yfp* cassette to result in the indicated genome sizes. **(D)** Upper gel: Neutral gel electrophoresis and Southern blot analysis of the indicated scAAV-YFP genomes, which were packaged into, and isolated from, AAV2 particles. Lower gel: Southern blot analysis of low-molecular-weight (Hirt) extracts corresponding to the indicated constructs. Brackets indicate monomeric (M) and dimeric (D) AAV replicative forms. ss, single-stranded. Parts of this figure were published and adapted from Fakhiri *et al.* [252].

The same experiments were repeated using the HBoV1 capsid (**Figure 10**). In contrast to AAV2, HBoV1 capsids could package oversized ssAAV and scAAV genomes up to 6.1 (**Figure 10A,** upper gel) and 3.2 kb (**Figure 10B,** upper gel), respectively. This implies that the HBoV1 capsid can package 0.5 kb scAAV DNA and 1 kb ssAAV DNA more than AAV2 capsids. Importantly, proper rescue and replication of the rAAV genome occurred during packaging into the HBoV1 capsid for all the constructs tested (**Figure 10A-B**, lower gels). Notably, the fraction of monomeric to dimeric replication forms seemed to be equal in the rAAV2/HBoV1 system, whereas we observed a preference for the monomeric form in the rAAV production system (compare lower gel pictures in **Figure 9B-D** and **Figure 10A-B**).

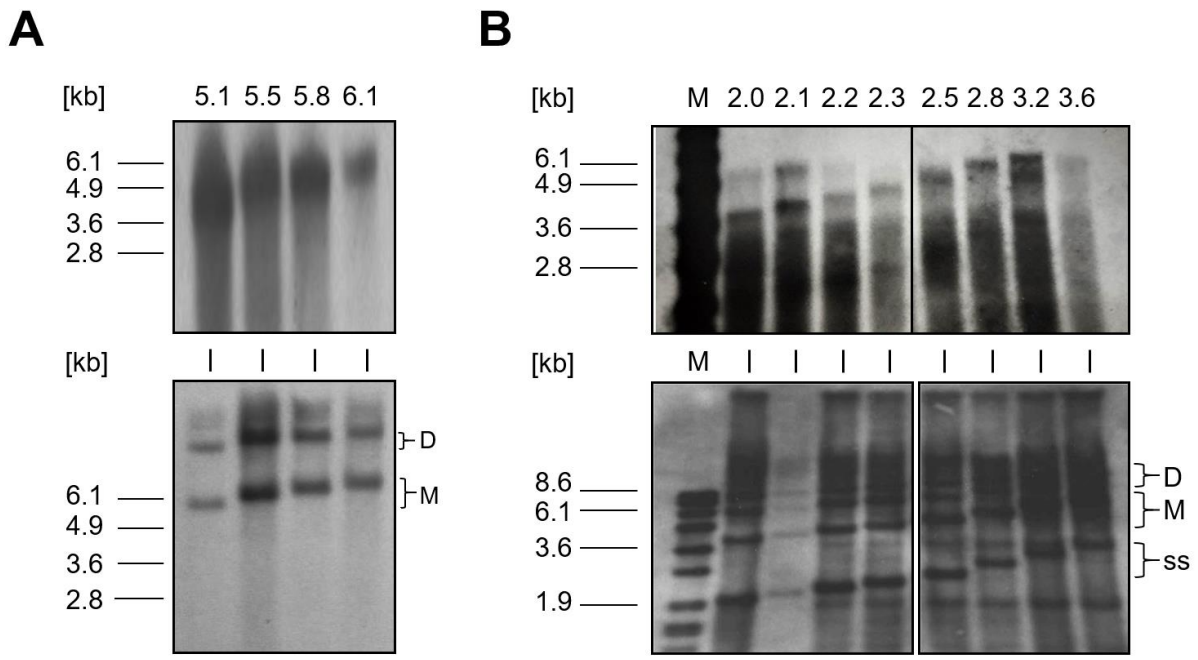


Figure 10. Packaging of oversized ssAAV and scAAV genomes into the HBoV1 capsid.

(A) Upper gel: Alkaline gel electrophoresis and Southern blot analysis of the ssAAV-CRISPR genomes shown in **Figure 9A**, which were packaged into, and isolated from, HBoV1 particles. Lower gel: Southern blot analysis of low-molecular-weight (Hirt) extracts corresponding to the indicated constructs. **(B)** Upper gel: Alkaline gel electrophoresis and Southern blot analysis of the indicated scAAV-YFP genomes, which were packaged into, and isolated from, HBoV1 particles. Lower gel: Southern blot analysis of low-molecular-weight (Hirt) extracts corresponding to the indicated constructs. Brackets indicate monomeric (M) and dimeric (D) AAV replicative forms. ss, single-stranded. Panel A and a part of panel B were published and adapted from Fakhiri *et al.* [252].

3.1.3 HBoV1 tyrosine-phenylalanine mutants show distinct transduction abilities

One limitation of rAAV2/HBoV1 vectors is their low transduction ability in the absence of proteasome inhibitors (PIs) (shown in **Figure 11** and also reported elsewhere [10]). This dependency represents a major hurdle in the clinical application of these vectors for gene therapy. Interesting in this context is the observation that phosphorylation of surface-exposed tyrosine residues in the capsids of various AAV serotypes targets the viral particles for ubiquitination and proteasome-mediated degradation [257]. Thus, mutations of these tyrosines (Y) to phenylalanines (F) reduced capsid degradation and enhanced transduction ability *in vitro* and *in vivo* (as for example shown for residues Y700 and Y730 in the AAV2 capsid) [258]. Interestingly, when the same two residues were converted to alanines, capsid assembly no longer occurred [259], which reflects their biological relevance for different steps in the viral life cycle.

Based on these findings, we wondered whether surface-exposed tyrosines in the HBoV1 capsid might play a similar role. Therefore, we cooperated with the group of Mavis Agbandje-McKenna (University of Florida, USA), which predicted these Y residues based on the 3D

capsid structure of the HBoV1 capsid that the group has determined by cryo-electron microscopy [260]. Accordingly, we conducted site-directed mutagenesis of six predicted surface tyrosines to phenylalanines, as outlined in **Figure 11A**.

To study the effect of each mutation on particle assembly and transduction ability, we first packaged a scAAV-YFP vector into the different HBoV1 capsid mutants and purified the resulting particles using iodixanol gradient centrifugation. Analysis of the 40% phase via qPCR revealed the presence of DNase I-resistant particles at levels comparable to HBoV1 (data not shown), implying that virus assembly was not compromised.

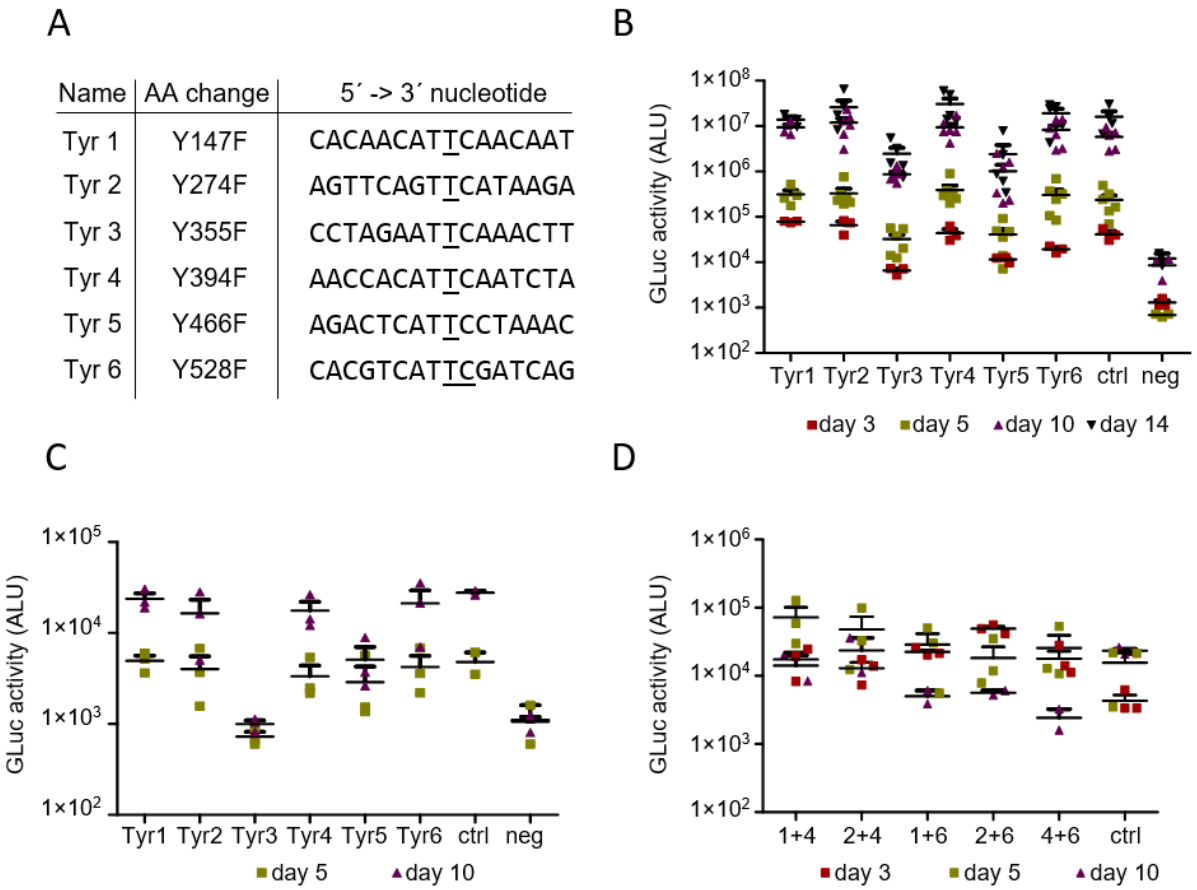


Figure 11. Single and double HBoV1 tyrosine-to-phenylalanine mutants.

(A) Tyrosine (Y) to phenylalanine (F) mutations in the HBoV1 *cap* ORF. Numbers refer to the amino acid positions in VP2. The corresponding nucleotide sequence is shown with the changed nucleotide underlined. **(B-C)** Transduction of pHAE with the indicated Y-to-F single mutants in the presence (B) or absence (C) of proteasome inhibitors. **(D)** Transduction of pHAE with the indicated Y-to-F double mutants in the absence of proteasome inhibitors. GLuc activity in the medium was measured at the indicated time points (mean + SEM, n ≥ 3 independent transwells) and plotted on the y-axis as arbitrary light units (ALU). For all tested constructs, a MOI of 2×10^4 was used. ctrl = HBoV1 *cap* sequence without mutations; neg = lysates from untransduced cells.

Next, to test whether the Y-to-F mutations had alleviated the dependency on proteasome inhibition, we transduced pHAE from the apical side at a MOI of 2×10^4 (data of this pre-screen are not shown). In the presence of PIs, transgene expression could be detected for Y147, Y274, Y394 and Y528, but not for Y355 and Y466. In the absence of PIs, only a low number

of YFP+ cells could be detected (except for Y355 and Y466), which hampered the quantification of these events. Nonetheless, the best-performing candidates from this experiment were juxtaposed in double mutants, to test whether this might further augment transduction in the absence of PIs (see **Table 2** for a list of mutants).

To allow for a better quantification of transduction efficiencies, we packaged the GLuc reporter that has a higher sensitivity and a larger dynamic range of activity than the YFP reporter used in the pre-screens into all single and double HBoV1 Y-to-F mutants. Again, all the tested single and double mutants resulted in titers comparable to the HBoV1 control (ctrl), as estimated from TaqMan qPCR analysis of viral stocks (data not shown).

To investigate the impact of the single and double Y-to-F mutants, we transduced pHAE with all the different variants or the ctrl in the presence or absence of PIs (**Figure 11B-D**). In concordance with our results of the pre-screen using the YFP reporter (data not shown), the single Y-to-F mutants Y147, Y274, Y394 and Y528 resulted in highest transgene expression, which was slightly better or comparable to the ctrl (**Figure 11B-C**). In contrast, the two other mutants Y355 and Y466 were impaired in their transduction ability at all time points (6.8- and 4.5-fold reduction, as compared to the ctrl). The double Y-to-F mutants showed a transduction ability similar to the underlying single mutants and ctrl in the absence of PIs (**Figure 11D**). Also, among each other, the double Y-to-F mutants gave a similar transduction in the presence of proteasome inhibition (data not shown).

Collectively, all tyrosine mutants (single and double; except for Y355 and Y466) were comparable in their transduction ability in the presence or absence of PIs, respectively. The tested Y-to-F mutations could, however, not improve transduction in the absence of PIs and displayed a similar reduction in transgene expression to the ctrl (~200-fold).

Of note, after completion of these experiments, a refinement of the HBoV1 capsid structure (personal communication by M.A.-M.) revealed that only three of the studied tyrosine residues, namely, Y147, Y274 and Y466, are indeed located on the capsid surface. Interestingly, Y355, whose mutation resulted in the strongest phenotype, is buried inside the capsid, which implies a unique function beyond assembly and receptor binding. Moreover, based on this new structure, other surface-exposed tyrosine residues were predicted: Y218, Y360, Y387 and Y510. The site-directed mutagenesis of these residues is currently ongoing and was beyond the time frame of this thesis.

3.2 Creation of new BoV helper plasmids for rAAV/BoV vector production

3.2.1 Phylogenetic and sequence analysis of primate BoVs

Due to their higher packaging capacity and their potent transduction of primary airway epithelia, pseudotyped rAAV2/HBoV1 vectors are presently being developed as attractive tools for gene delivery into the lung [10, 261]. Importantly, HBoV1 is not the only BoV variant that can be harnessed for gene transfer, as many other serotypes and variants were detected in both, primates and non-primates. Interestingly, this broad distribution has also been observed for AAVs. The first rAAV vector was derived from the AAV serotype 2, which was isolated from human tissue [240]. Later, 12 other serotypes and hundreds of variants were described, many of which have been engineered as vectors and tested *in vitro* and *in vivo* [118, 119]. Thus, we here aimed to test whether these promising findings with AAV vectors can be recapitulated with BoV, and therefore focused our attention on four additional primate BoVs, namely, HBoV2-4 and GBoV and on their as-of-yet unknown tropism.

First, we determined the degree of sequence identity in the non-structural and structural ORFs by performing a sequence alignment using the Clustal Omega online tool [236] (**Figure 12A**). The results revealed a generally high degree of conservation and at least 70% of sequence identity. Especially HBoV2 and HBoV4 share 92, 94 and 87% of sequence identity in the *np1*, *ns1* and *cap* ORFs, respectively. Interestingly, HBoV1 is mostly related to GBoV, sharing >80% of sequence identity in all non-structural and structural genes.

On the other hand, HBoV3 shares the highest level of sequence identity with HBoV1 and GBoV in the *np* and *ns* region (>84%) but less in *cap* (77% and 80%, respectively). HBoV1 *cap* seems to be more closely related to that of HBoV2 and HBoV4, with 87% and 88% identity, respectively. This discrepancy in loci association has been observed before [262] and was linked to possible recombination events between HBoV1 and HBoV2 that led to the emergence of HBoV3.

To gain a better understanding of the evolutionary relationship of these primate BoVs, the alignment was used to conduct a phylogenetic analysis (**Figure 12B**). Here, we were especially interested in the viral *cap* ORF, which determines the cell and tissue tropism. To this end, the *cap* ORFs used in this study (black circles in **Figure 12B**) were compared to different representative animal BoV sequences obtained from GenBank.

A

%	HB1	HB2	HB3	HB4	GB
HB1	-	76	86	77	88
HB2	75	-	75	92	76
HB3	87	75	-	75	87
HB4	73	94	74	-	77
GB	87	75	85	75	-

%	HB1	HB2	HB3	HB4	GB
HB1	-				
HB2	78	-			
HB3	77	87	-		
HB4	77	87	88	-	
GB	82	79	80	80	-

ns *np1* *cap*

B

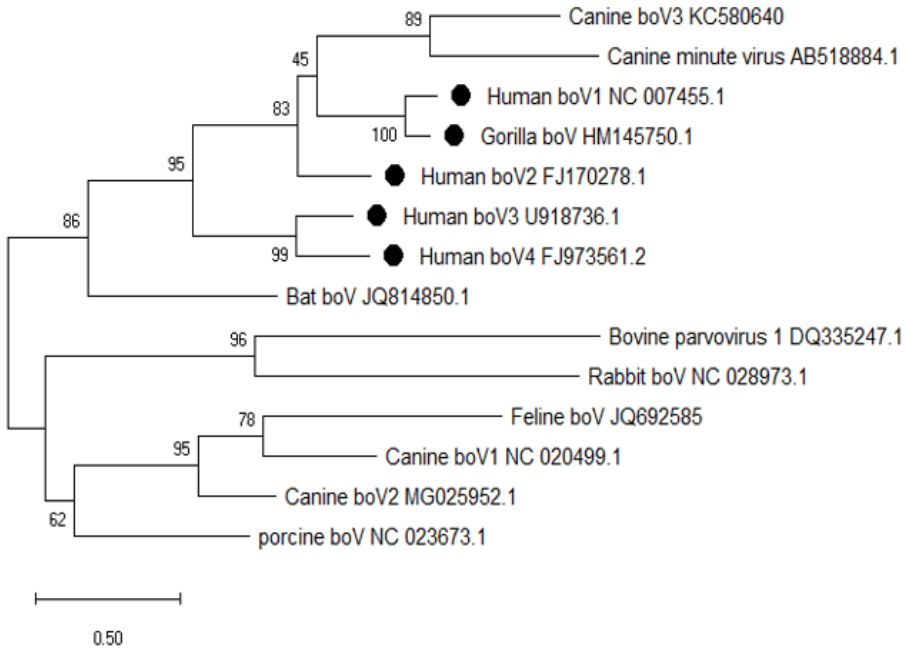


Figure 12. Sequence alignment and phylogenetic analysis of primate BoVs.

(A) Percentage of sequence identity between the different indicated primate BoV in the *ns*, *np1* and *cap* ORFs. **(B)** Phylogenetic analysis of BoV *cap* sequences used in this work (marked with black circles). The evolutionary history was inferred in MEGAX by using the Maximum Likelihood method based on the Tamura-Nei model [263]. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test (1,000 replicates) are shown next to the branches. The scale bar represents the average number of nucleotide substitutions per site. Panel B was adapted from Fakhiri *et al.* [252].

As expected from the high sequence identity, HBoV1/GBoV and HBoV3/HBoV4 clustered together in one clade each. The analysis also revealed an interesting relationship of all primate BoVs to (i) the canine minute virus (GenBank: AB518884), a pathogenic BoV linked to gastrointestinal and lung diseases in fetal and young dogs [67], (ii) the canine BoV3 (GenBank: KC580640), a newly identified canine BoV detected in the liver [264], and (iii) a bat BoV (GenBank: JQ814850) with unknown tropism and properties. From this, we concluded that

cells from the lung, liver or intestine might represent interesting targets for HBoV2-4 or GBoV vectors (see also later chapter 3.2.5).

3.2.2 Assembly and iodixanol-based purification of five chimeric rAAV/BoV vectors

Based on the high conservation of the *ns* and *np1* ORFs between primate BoVs, we speculated that these proteins may complement each other and support the packaging of rAAV genomes into capsids from different BoV serotypes. This has previously been shown for rAAVs, where the Rep proteins from AAV2 supported the packaging of not only AAV2 genomes but also genomes derived from all other AAV serotypes, except for AAV5 [118]. Accordingly, we only replaced the *cap* ORF of HBoV1 in a seamless manner by the *cap* sequences from HBoV2-4 and GBoV (**Figure 13A**).

Notably, the small part of *np1* (14 nt) that overlaps with the *cap* ORF is highly conserved among primate BoVs and thus remains intact after insertion of the new *cap* sequences. This resulted in the new helpers pCMVNS*NP1Cap$_x$, where x denotes the BoV serotype used. Importantly, all BoV helpers used in this work only encode the smaller NS proteins (NS3-4 and NP1*)* from HBoV1. The knockout of NS1-2 (NS*) resulted in higher viral titers than in the original rAAV/HBoV1 production system (personal communication by Ziying Yan [University of Iowa, USA] at the time when the work was performed). A recent report by Yan *et al.* in 2018 provided a comparison of both systems and showed that viral titers were indeed negatively affected by high expression of NS1/NS2 but not NS3/NS4 [138].

Next, we applied the triple-transfection protocol described in section 3.1 to test whether intact viral particles could be produced. Briefly, the helper constructs were transfected into HEK293T cells along with pDGΔVP and a scAAV-YFP plasmid. For purification of the viral particles from the cell lysate, we employed iodixanol gradient ultracentrifugation as described for rAAV2/HBoV1 vectors (section 3.1.1, **Figure 8** and methods sections 2.2.5.2 and 2.2.5.4).

TaqMan qPCR analysis of the 40% iodixanol fractions revealed the presence of DNase I-resistant particles for all rAAV/BoV preparations in comparable numbers of ~5×10^{11} genome copies per mL (**Figure 13B**). Moreover, correct expression of the three Cap proteins VP1-3 was confirmed by Western blot analysis of the 40% virus-containing fractions (**Figure 13C**).

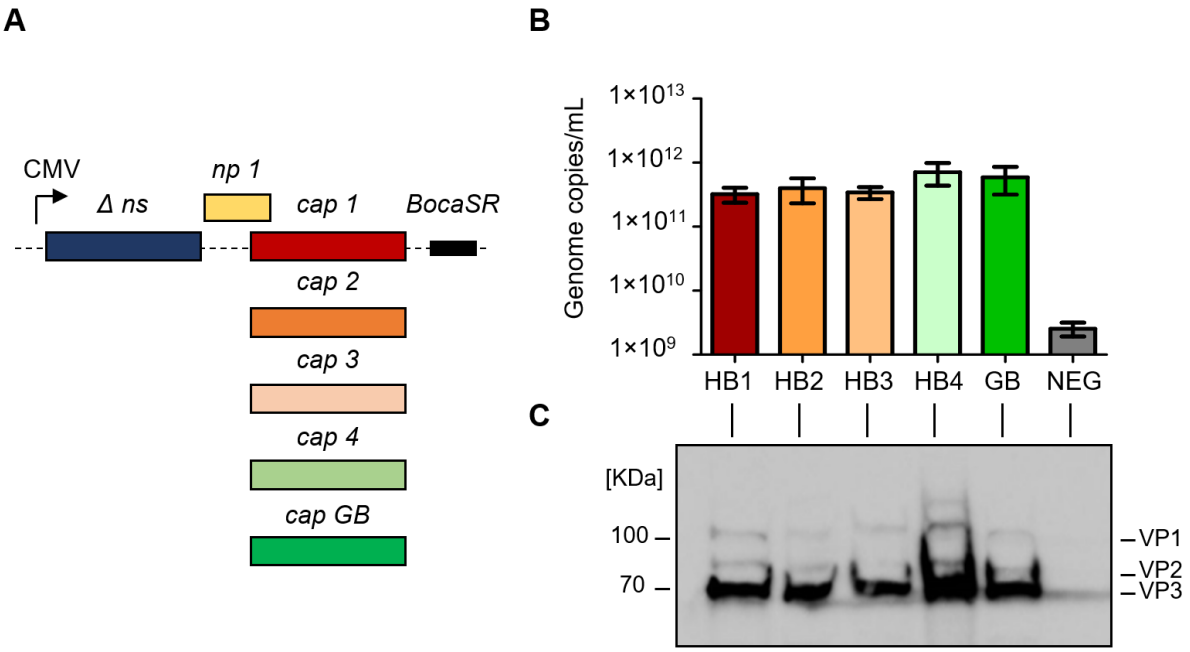


Figure 13. Construction and production of rAAV/BoV vectors.

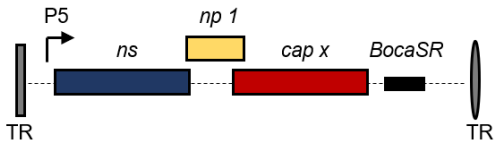
(A) Schematic representation of BoV helper constructs. The *cap1* ORF from HBoV1 was replaced by the *cap* sequences from HBoV2-4 and GBoV. The CMV promoter drives the expression of all shown ORFs. BocaSR = bocaviral noncoding small RNA. Δ *ns* = encodes only the NS3 and NS4 proteins. **(B)** TaqMan qPCR analysis of iodixanol-purified scAAV-YFP/BoV vectors. The mean genome copy number per mL (± SEM, n = 4) of the 40% iodixanol fraction is plotted on the y-axis. **(C)** Western blot analysis of the viral stocks shown in B. Detected are the three capsid proteins VP1-3. NEG = iodixanol gradient from untransfected cells (negative control). Panels B-C from this figure were adapted from Fakhiri *et al.* [252].

After we had succeeded to produce all five bocaviral vectors from the hybrid helpers, we asked whether this approach can also be used to produce replication-competent BoVs. As benchmarks, we performed side-by-side comparisons to the BoV vectors to test whether and to which extent the vectors reflect the behaviour of the wt viruses. Thus, we transferred the *cap* ORFs to the HBoV1 infectious clone [72] as shown in **Figure 14A**. In this context, the original viral p5 promoter drives the expression of transcripts encoding the NS, NP1 and Cap proteins. Moreover, in contrast to the helper plasmids for recombinant virus production, all NS proteins are expressed, which has been shown to be important for wtHBoV1 replication in pHAE [78]. Finally, the genomes are flanked by the HBoV1 asymmetrical terminal repeats (TRs), which are required for replication and packaging of the wtBoV genome [72, 265].

For virus production, we transfected the infectious clones into HEK293T cells and purified the viral particles using iodixanol gradients. Next, we quantified the amount of encapsidated viral genomes using TaqMan qPCR analysis (**Figure 14B**). Intriguingly, the wt viruses produced at efficiencies comparable to the pseudotyped viral vectors, except for HBoV2 and HBoV4, which gave 10-fold lower titers. The comparable efficiencies obtained with HBoV3 and GBoV support the previously observed advantage of the pseudotyped cross-

genera approach (rAAV/HBoV1) as compared to a system using recombinant genomes based on HBoV1 TRs (rHBoV1/HBoV1). The latter was tested in the above-mentioned study by Yan *et al.* and has been shown to produce less efficiently that wtHBoV1 (viral yields were ~20% of wtHBoV1) [10].

A



B

Type	Titer (gc/mL)
HBoV1	3.0×10^{11}
HBoV2	3.0×10^{10}
HBoV3	2.4×10^{11}
HBoV4	1.4×10^{10}
GBoV	4.5×10^{11}

Figure 14. Construction of wt and replication-competent bocaviral plasmids. **(A)** Genome organization of the HBoV1-4 and GBoV infectious clones. All viral elements except for the *cap* ORF are derived from HBoV1. *cap x* is derived from one of the BoV serotypes, *i.e.* x = HBoV1 to 4, or GBoV. **(B)** Wild-type bocaviral stocks purified by iodixanol gradient ultracentrifugation. The number of viral genome copies (gc) per mL was determined by TaqMan qPCR analysis (n = 1).

3.2.3 rAAV/BoV vector purification using cesium chloride (CsCl) density ultracentrifugation

CsCl density ultracentrifugation is one of the standard methods used for rAAV vector purification [266, 267]. Despite being more laborious and time-consuming than standard iodixanol centrifugation, rAAV stocks purified by CsCl gradients were shown to contain fewer empty particles (<1% *versus* ~20% for iodixanol [268]). Empty particles represent an inadvertent contaminant and interfere with efficient *in vivo* applications [269]. Therefore, we studied the migration of the different bocaviral capsids in a CsCl gradient. To this end, we packaged a scAAV-YFP construct into the different BoV capsids or an AAV2 control and purified the resulting viral vectors as described earlier [10, 138] with two rounds of CsCl density centrifugation (for more details see methods section 2.2.5.3).

Concurrent with previous studies [10], rAAV/AAV2 vectors banded at a density of ~1.421 g/mL. By contrast, all rAAV/BoV vectors migrated at higher densities ranging from 1.438 to 1.489 g/mL (**Figure 15**). This shift compared to AAVs is in line with previous reports of rAAV2/HBoV1 vector preparations [10] and can be explained by the larger size and thus different density of the BoV capsids. To increase the purity and further concentrate the virus, the collected virus-containing fractions were subjected to a second round of CsCl density centrifugation using a smaller volume for the gradient (as exemplified for AAV2 in **Figure 16A-B**). After dialysis and concentration of the viral preparations, the number of viral genomes per mL was determined using TaqMan qPCR analysis. From these values, the number of genome copies per cell was estimated, to allow for a comparison between the two purification methods

(iodixanol *versus* CsCl) in cases where different numbers of 15 cm dishes had been used. Our results showed a slightly but consistently higher yield of viral genomes per cell following iodixanol-based purification (**Figure 16C**). This is, however, expected considering the typical loss of viral particles in the second round of CsCl centrifugation and during the concentration step using Amicon columns.

In conclusion, iodixanol purification was chosen as a method of choice for the majority of vector preparations in this work because of its many advantages: (i) the comparable or higher viral yields (**Figure 16C**) and infectivity (**Figure 8D**), (ii) the ease of handling and the short run time of 2 h (compared to 24 h for CsCl), (iii) the inert properties of iodixanol, which allow for direct application in cells [270, 271], and (iv) the broad applicability for purification of different BoV serotypes that, in contrast to CsCl (**Figure 15**), all accumulate in the same 40% phase (**Figure 13B**). This eliminates the need for extensive fractionation and refractometry.

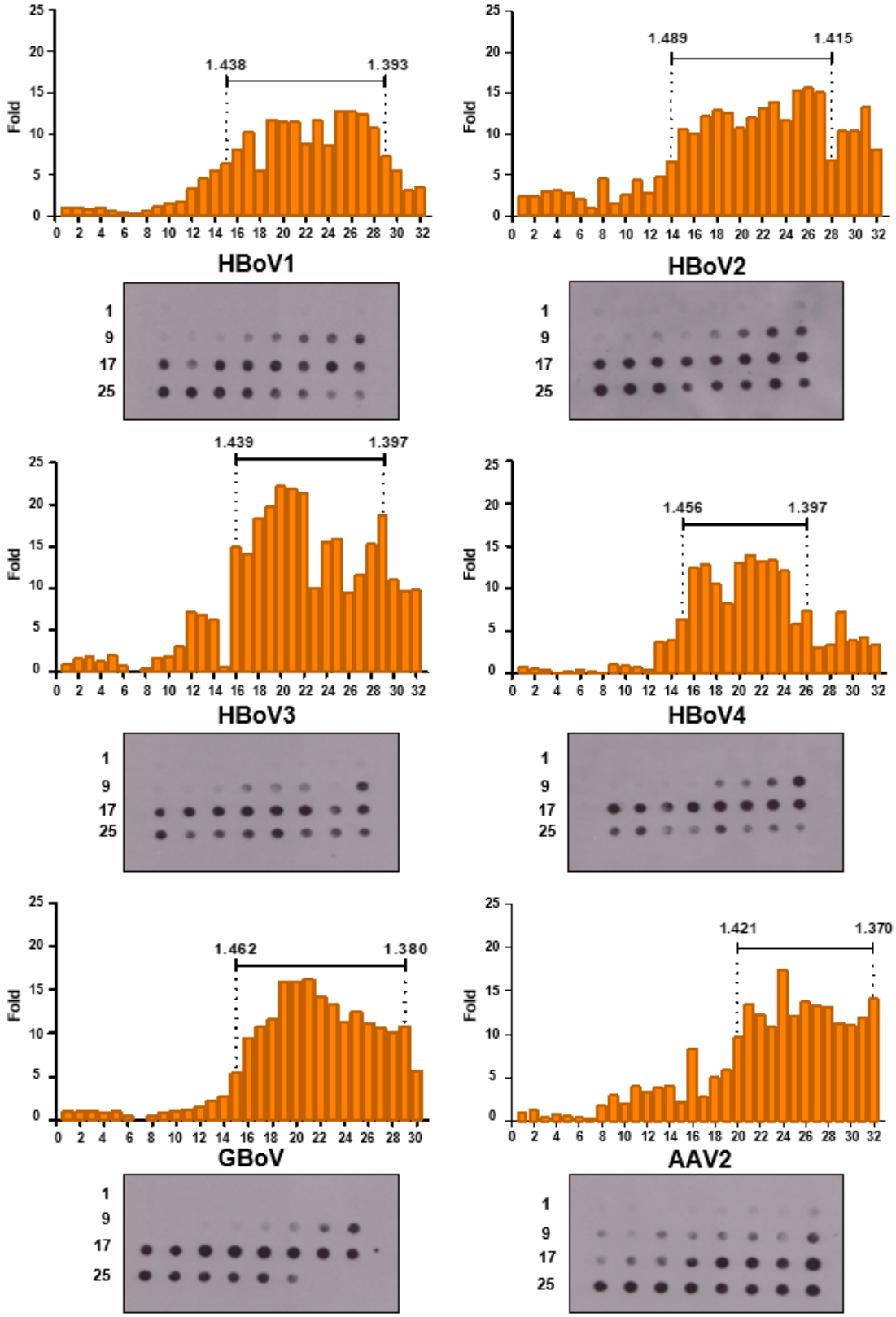
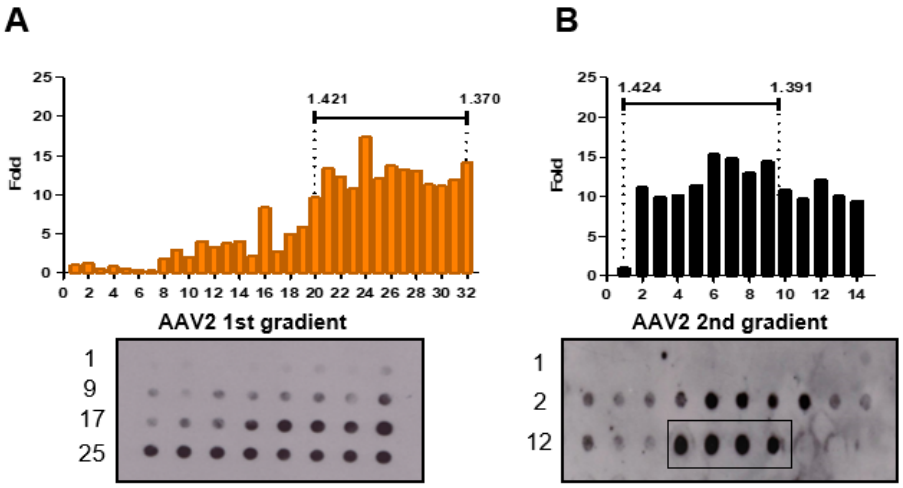


Figure 15. Purification of pseudotyped rAAV/HBoV1 vectors using CsCl density ultracentrifugation.

From each gradient, 30 to 32 fractions were collected and analyzed for the presence of packaged viral genomes (scAAV-YFP) using DNA dot blot analysis with a probe against the *yfp* cassette. The intensity of each dot was measured using ImageJ after background subtraction and the fold increase to fraction

1 was plotted on the y-axis. The density range (g/mL) in which the full viral particles accumulated is denoted in each graph. This data set was published and adapted from Fakhiri *et al.* [252].



C

Purification	Iodixanol	CsCl
Variant	Titer (gc/cell)	Titer (gc/cell)
HBoV1	4.6×10^3	1.5×10^3
HBoV2	5.7×10^3	2.0×10^3
HBoV3	4.9×10^3	2.6×10^3
HBoV4	1.0×10^4	3.9×10^3
GBoV	8.4×10^3	5.4×10^3
AAV2	2.4×10^4	1.1×10^4

Figure 16. Two-step purification of pseudotyped rAAV/BoV and rAAV/AAV2 vectors using CsCl density ultracentrifugation.

(A-B) Representative example of a two-step purification of a rAAV/AAV2 control vector. Gradients 1 (A) and 2 (B) were fractionated and analyzed for the presence of viral genomes (scAAV-YFP) using DNA dot blot analysis with a probe against the *yfp* cassette. For the second gradient (B), 14 fractions were collected. The last four indicated dots in the corresponding dot blot are dilutions of a positive control (linearized CMV-YFP plasmid). **(C)** Titers of rAAV/BoV and rAAV/AAV2 vectors in genome copies (gc) per cell after one-step iodixanol (average; n = 4, except for AAV2, n = 2) or two-step CsCl ultracentrifugation (average; n = 2). Panel A and C were adapted from Fakhiri *et al.* [252].

3.2.4 Electron microscopy analysis of rAAV/BoV vectors

To further characterize our rAAV/BoV vectors and wtBoV stocks, we used electron microscopy analysis to study virus morphology (**Figure 17**). This was done in collaboration with Robin Burk from the group of Hans-Georg Kräusslich (Virology Department, University Hospital Heidelberg, Heidelberg, Germany). As hoped for, all rAAV/BoV stocks were composed of particles displaying the typical icosahedral structure observed before for HBoV1 [10]. Intriguingly, for all rAAV/BoV vectors, we mainly detected full viral particles, *i.e.* capsids

containing a viral genome (**Figure 17,** upper panel). Note that because the HBoV1 wt and vector preparations were of low quality, respective images are not shown; however, exemplary images can be found in references 10, 72, 272. In contrast to the recombinant vectors, we found that the stocks of hybrid replication-competent (wt) HBoV2-4 predominantly contained empty particles (**Figure 17**, lower panel). Surprisingly, the opposite was true for the wtGBoV stock, which contained more full than empty capsids. This might indicate a better compatibility of the GBoV Cap with the HBoV1 NS proteins and terminal repeats (TRs). Notable in this context are also the consistently higher titers obtained with the rAAV/GBoV vector (**Figure 16C**).

Interestingly, we noticed smaller particles in our viral preparations (wt and recombinant), which most probably correspond to proteasome 20S subunits [273]. Proteasomes were also found as low-level contaminants in AAV preparations but to a much lower extent [272].

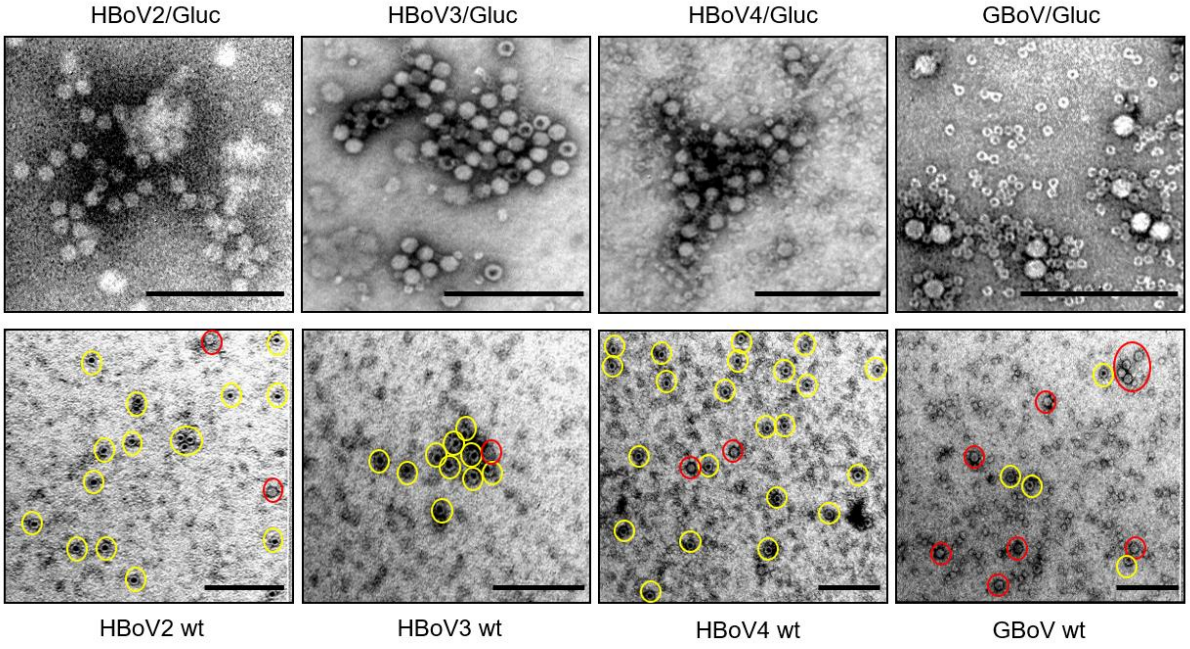


Figure 17. Negatively stained electron micrographs of recombinant and wild-type BoVs.

The indicated rAAV/BoV vectors or hybrid wild-type constructs were purified using iodixanol gradient centrifugation. Particles were negatively stained with 1% uranyl acetate and examined by electron microscopy. Empty particles are stained and evidenced by a dark core. The numerous smaller structures visible in the GBoV sample are most likely 20S proteasomes that co-purified with the viral particles. In the lower panel, empty particles are marked with yellow circles, while genome-containing particles are highlighted with a red circle. Scale bar = 200 nm. Images were taken by Robin Burk.

3.2.5 Assessment of BoV transduction in pHAE and CuFi-8 cell line

The ability to efficiently produce all the five pseudotyped rAAV/BoV vectors laid the foundation for the second aim of this part of the thesis, which was to test the functionality and to define the as-of-yet unknown cell tropism of the new HBoV2-4 and GBoV vectors/viruses.

Initially, the incomplete knowledge of the natural tropism of these BoV isolates prevented us from rationally selecting specific cells for testing and for use as benchmarks. Interestingly, though, animal BoVs have been extensively studied and linked to many diseases (see section 1.1.2.1). Thus, we determined the closest relatives to primate BoVs in the animal kingdom to indirectly predict promising target cells. As shown above, this was achieved by performing a phylogenetic analysis of the *cap* ORF, which contains the hypervariable regions, *i.e.*, the most likely determinants of the viral tropism (**Figure 12**). One interesting target revealed by the analysis was the lung, which was further supported by the sporadic detection of some of the primate BoVs in nasopharyngeal aspirates [75].

Accordingly, we tested whether our established pHAE cell culture system was permissive to transduction with any of the newly generated viral vectors and/or to infection with, and replication of, the corresponding wild-type viruses, respectively. To this end, scAAV-YFP or scAAV-GLuc genomes were packaged into BoV capsids or AAV2 as control. The use of two different reporters allowed for the visualization (YFP) or consecutive monitoring (GLuc) of transgene expression over time. Transductions were always performed in the presence of two proteasome inhibitors (LLnL and doxorubicin), which was shown to increase the transduction with rAAV/HBoV1 by a factor of 1,000 [10].

Bronchial sections from different donors were used to cultivate pHAE on transwells as a polarized, pseudostratified epithelium. The transwells were inoculated from the apical side with equal amounts of viral particles. Transgene expression became detectable at 48-72 h post-transduction (data not shown) and sustained over the course of the experiment (day 14, **Figure 18A**). Interestingly, we observed that not only HBoV1 but also HBoV4 and GBoV transduced pHAE to different extents, whereas HBoV2 and HBoV3 remained at background level. Since GLuc is an excreted protein, we were able to monitor the increase in expression over time (**Figure 18B**) and found that it reached a peak after 12-14 days in culture (not shown).

Moreover, we could confirm our findings in CuFi-8 cells, which is a cell line derived from a cystic fibrosis patient and was shown to be permissive for HBoV1 infection, albeit to a lesser extent than pHAE [10] (**Figure 18C**).

Previous studies have shown that HBoV1 displays a preference for the apical side of the airway epithelia [10]. Thus, we asked whether this polarity also applies to HBoV4/GBoV and may even have hampered transduction with HBoV2/HBoV3 in the previous experiment. To test this, transwells were incubated with equal amounts of viral particles either from the apical or basolateral side, as described in methods section 2.2.5.6. In accordance with prior observations [10], HBoV1 displayed the expected polarity of transduction (apical >> basolateral, 100-fold). Likewise, GBoV was 10-fold more efficient at transducing from the apical than from the basolateral side (**Figure 18D**). This is intriguing in view of the close phylogenetic

relationship of these two viruses. Overall, delivery from the basolateral side could not improve the transduction of any of the tested viruses.

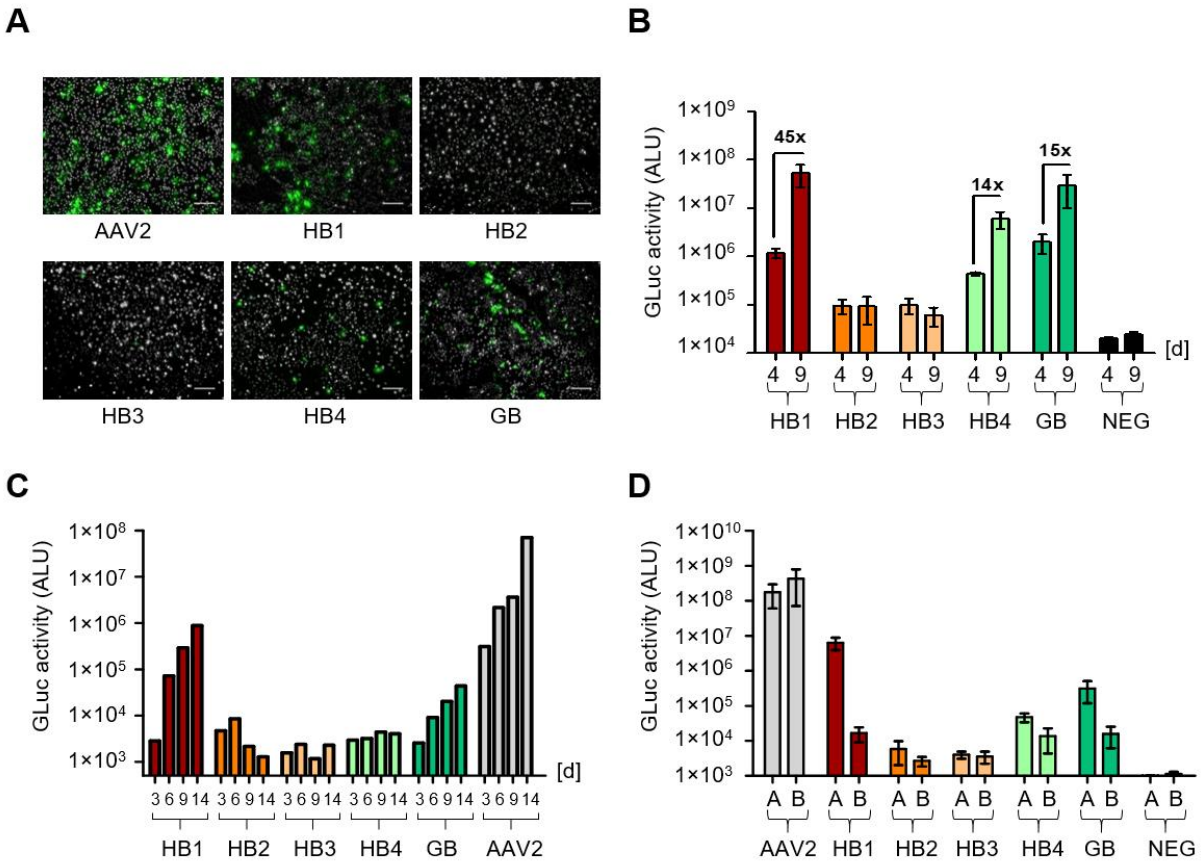


Figure 18. Transduction of pHAE and CuFi-8 cells with rAAV/BoV vectors.

(A) Transduction of pHAE from the apical side with scAAV-YFP/BoV vectors or AAV2-YFP as control (CsCl purified) at a MOI of 4×10^4. Images were taken 14 days post-transduction. Gray = Hoechst staining of nuclei. Scale bar = 50 µm. **(B)** Transduction of pHAE with scAAV-GLuc/BoV vectors (iodixanol purified) at a MOI of 2×10^4. Aliquots of medium were collected 4 and 9 days post-transduction. GLuc activity was measured and plotted on the y-axis (mean ± SEM, n = 3) as arbitrary light units (ALU). **(C)** Transduction of CuFi-8 cells with the indicated scAAV-GLuc/BoV vectors or AAV2-GLuc as control. A total of 2×10^{10} viral genomes was added to the apical side of the transwell and transduction was performed O/N in the presence of 1 µM doxorubicin and 8 nM LLnL. GLuc activity was measured at the indicated time points and plotted on the y-axis (average of two independent transwells). **(D)** Transduction of pHAE from the apical (A) or basolateral side (B) with the indicated scAAV-GLuc/BoV vectors at a MOI of 2×10^4. GLuc activity was measured at day 9 post-transduction and plotted on the y-axis (mean ± SEM, n = 5). Transduction of pHAE was always performed in the presence of LLnL and doxorubicin at concentrations of 40 µM and 5 µM, respectively. NEG = untransduced cells (negative control). n = number of independent transwells per condition. Transduction of CuFi-8 cells in panel C was performed by Verena Schildgen (Institute for Pathology, Kliniken der Stadt Köln gGmbH, Hospital of the Private University Witten/Herdecke, Cologne, Germany). This data set was published and adapted from Fakhiri *et al.* [252].

Notably, in contrast to previous observations [10], transduction with rAAV2 was more efficient than with HBoV1 from both, the apical and basolateral side of the transwell. This discrepancy can be explained by differences in the experimental set-up, such as the incubation time of 1 h here compared to 16 h in the previous study. Moreover, the extent of polarization

and tight-junction formation in pHAE may also have especially influenced transduction with rAAV2, which can transduce from both sides of the membrane and is supposedly more efficient from the basolateral side. Finally, the application of proteasome inhibitors during the transduction can particularly boost transduction of rAAV2 from the apical side [274] and hence lead to the loss of the expected basolateral polarity [214].

Next, we asked whether the hybrid wild-type viruses would recapitulate the transduction patterns seen with the BoV vectors. Therefore, we infected pHAE with the replication-competent BoVs described in section 3.2.2 at a MOI of 1×10^4. Apical washes were collected over different time points to detect the presence of released viral particles. As expected, the highest number of viral genomes could be detected for HBoV1 16 days post-infection (~7×10^8 gc/µL), followed by HBoV4 and GBoV with an average of ~1×10^6 and 2×10^6 gc/µL, respectively (**Figure 19A**). By contrast, HBoV2 and HBoV3 were only detected at very low levels, which were close to the background of the assay. The trend obtained with the wtBoV infection of pHAE is similar to what we have previously observed using the BoV vector system (**Figure 18A-C**). Notable, however, is the ratio to HBoV1, which differs between transduction (vectors) and infection (wt viruses). For example, wtGBoV is 300-fold less efficient than wtHBoV1 but only ~10-fold in the vector system.

We furthermore assessed viral VP protein expression, which was shown to be high during a productive infection of pHAE with wtHBoV1 [72]. Therefore, the cells were detached at day 16 and proteins were extracted for Western blot analysis (**Figure 19B**). Congruent with previous observations [72], wtHBoV1 replication led to a high expression of VP1-3 proteins. By contrast, protein expression was below the detection limit of the Western blot for all other isolates, including the two that had yielded detectable genomic titers (HBoV4 and GBoV).

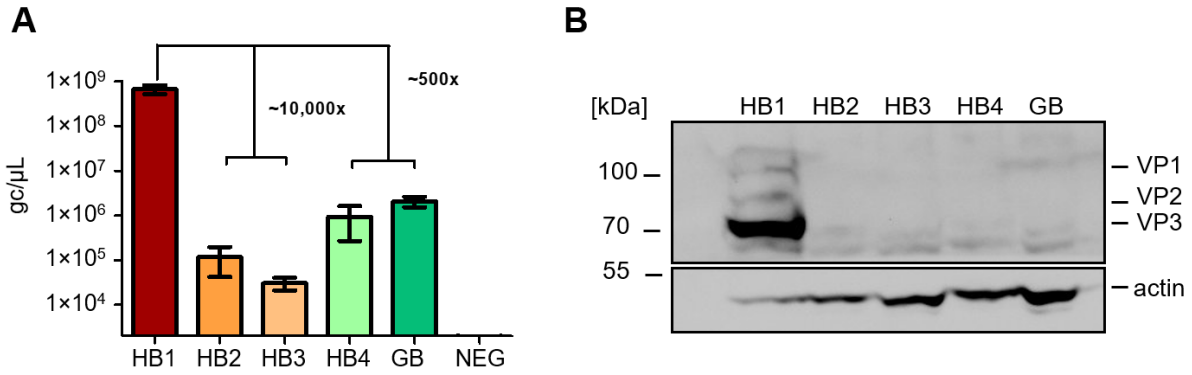


Figure 19. Infection of primary HAE with hybrid wild-type BoVs.

(A) pHAE were infected from the apical side with the indicated wtBoVs at a MOI of 1×10^4. Apical washes (in PBS) were collected 16 days post-infection and the viral DNA was extracted and quantified using qPCR analysis (mean; n ≥ 3). **(B)** Western blot analysis of cells infected with the indicated wtBoVs. Detected are the three BoV capsid proteins VP1, VP2 and VP3. Actin served as a loading control. gc = genome copies. Infections were performed in the absence of PIs.

3.2.6 Transduction of primary lung organoids reproduces findings in pHAE

Over the last years, many organoid systems have been established as models that represent the *in vivo* situation better than 2D cell cultures [275]. These organoids are usually composed of different organ-specific cell types, which arrange to form a three-dimensional architecture.

For this reason, we started a collaboration with the group of Hans Clevers (Hubrecht Institute, Utrecht, the Netherlands), who has established multiple types of primary organoids derived from a variety of tissues, such as liver, pancreas, lung and intestine [276]. We focused our attention on lung organoids and tested whether we can reproduce the data observed in pHAE in this alternative type of cell culture system. Therefore, lung organoids were transduced with equal amounts of the different scAAV-YFP/BoV or scAAV-GLuc/BoV vectors. Since we knew that two of the BoVs that we had functionally characterized prefer the apical side of the organoids, which lies on the inside, luminal access was provided by using two different methods: (i) breaking (br) of the organoids using mechanical shearing or (ii) direct injection (in) of vectors into the organoids (**Figure 20A**). Individual YFP+ cells were detected after three days in culture for HBoV1, HBoV4 and GBoV (**Figure 20B**), which agrees with our previous observations in pHAE (**Figure 18B**).

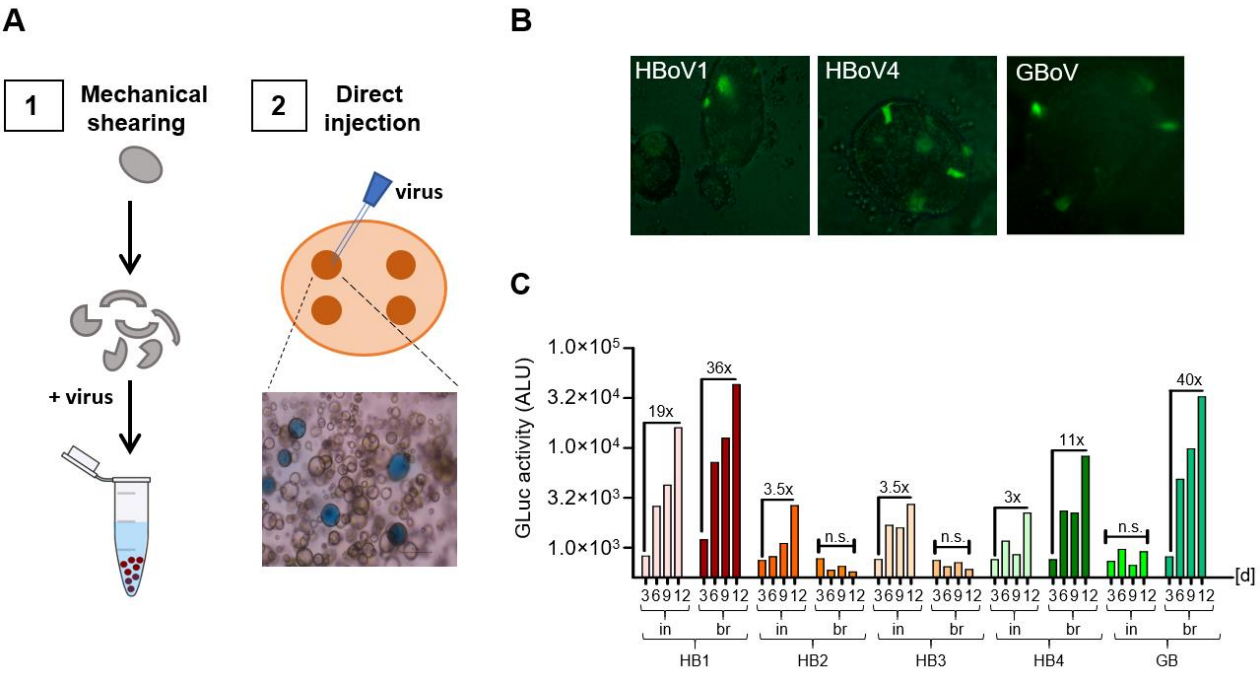


Figure 20. Transduction of primary lung organoids using rAAV/BoV vectors.

(A) Illustration of the two methods used for transduction of primary lung organoids: (1) In mechanical shearing, organoids are gently suspended and broken with a pipet tip to expose the apical surface. The broken organoids are then incubated with the virus solution O/N. (2) In direct microinjection, a small volume of virus is mixed with a green dye and the mix is directly injected into the lumen of the organoids. Successfully injected organoids are marked with a green color. **(B)** Transduction of primary lung organoids using mechanical shearing with 6.6×10^9 viral particles of scAAV-YFP/BoV. Pictures were taken by Jens Puschhof at day 5 for GBoV and day 6 for HBoV1 and HBoV4. **(C)** Transduction of primary

lung organoids with the indicated scAAV-GLuc/BoV vectors. Multiple organoids from one donor were either mechanically broken (br) and incubated with 5×10^9 viral genomes or microinjected (in) with 5×10^8 to 1×10^9 viral particles. GLuc activity was measured at different time points post-transduction and plotted on the y-axis as ALU. Transductions were performed in the presence of 1 µM doxorubicin. This data set was published and adapted from Fakhiri *et al.* [252].

These data were further supported using the second reporter GLuc, which allowed us to follow transgene expression over a period of 12 days (**Figure 20C**). Likewise, HBoV1, HBoV4 and GBoV displayed the highest levels of transgene expression, which increased with time. Noteworthy, we saw striking differences between the two application methods, which reflects the complexity of virus-cell interactions. For example, HBoV2 and HBoV3 preferred the injection method, whereas GBoV transduction was exclusively observed with broken organoids (compare "in" with "br" in **Figure 20C**).

During our transduction experiments in pHAE and primary lung organoids, we observed that only a few cells were targeted by the BoV vectors as compared to a broader transduction with rAAV2. Thus, we asked whether BoVs prefer a certain cell type in the airway. Therefore, we collaborated with Marc Schneider (Thoraxklink, University Hospital Heidelberg, Heidelberg, Germany) to perform cryosections and immunohistochemistry of transwells transduced with scAAV-YFP/HBoV1. Sections were stained using antibodies against cell type-specific markers, namely, MUC5AC for goblet cells, CC10 for Clara (club) cells, ß-Tubulin IV for ciliated cells and KRT5 for basal cells (**Figure 21**)**.** Most of the YFP+ (hence HBoV1+) cells displayed markers either typical for Clara or ciliated cells. Arguably, a cell-membrane staining (*e.g.* against E-cadherin) would substantially increase the quality of the data and allow for a more solid assignment of cell types. Here, however, we did not optimize the staining because the low rate of transduction and the significant loss of sample during sectioning further hampered a quantitative readout. Instead, to overcome these limitations, we validated our preliminary assumption of a specificity of HBoV1 for Clara and ciliated cells with a more sensitive method, namely, flow cytometry analysis (**Figure 22-23**)**.**

For the generation of pHAE, around 9×10^4 undifferentiated epithelial cells are seeded per transwell. From this population, a small number of progenitor cells (**Figure 22A**) start to differentiate into the four major cell types: ciliated, Clara (club), goblet and basal cells. This results in an invasion of the filter and the formation of a polarized, multi-layered structure. The nature of these progenitor cells and the process of lung tissue regeneration are not well understood and differ between mouse and human tissue. In humans, a subset of basal cells (KRT5+ KRT14+) are believed to be early progenitors, which give rise to many different cell types [277]. Moreover, some differentiated cells such as Clara cells have also been reported to self-renew or to generate ciliated cells [278]. Here, the unravelling of all the subtypes was beyond the scope of this work. Instead, we focused on the establishment of a flow cytometry protocol that allows the detection of the four major cell types in pHAE (see methods section 2.2.6.5).

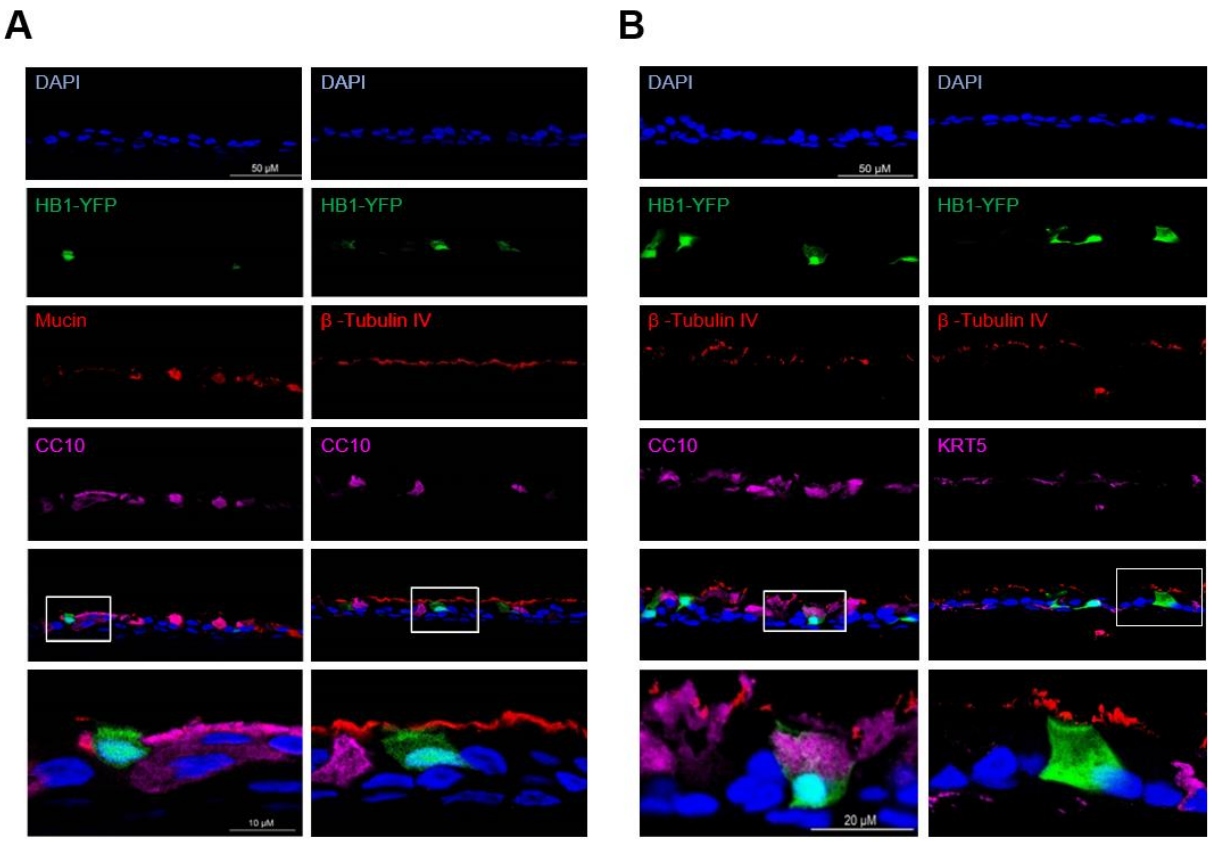


Figure 21. Immunohistochemical staining of scAAV-YFP/HBoV1–transduced transwells.

Polarized pHAE were transduced with scAAV-YFP/HBoV1 at a MOI of 6×10^4. Cryosections and stainings were performed 10 days after transduction by Marc Schneider. **(A)** and **(B)** represent exemplary immunofluorescence staining of transwells derived from two different patients. Shown in (A) are immunofluorescence co-stainings with anti-CC10 and anti-MUC5AC (Mucin) or anti-ß-Tubulin IV antibodies. In (B), a co-staining was performed with anti-ß-Tubulin IV and anti-CC10 or anti-KRT5 antibodies. Images at the bottom are merged images (two different magnifications) of the indicated cell-type-specific markers and the YFP+ (and hence HBoV1+) cells.

To this end, we transduced pHAE with scAAV-YFP/HBoV1 or scAAV-GLuc/HBoV1 as control and analyzed the cells after 10 days in culture. This time point was chosen because it combined good transgene expression with acceptable cell viability. Importantly, scAAV-GLuc/HBoV1 served as a negative control because our preliminary data consistently showed a change in the characteristics of the cell population after transduction, which was evidenced by a shift in the side and forward scattering (data not shown). Consequently, untransduced cells represented an unsuitable negative control for the analysis.

The graph in **Figure 22B** shows the low background of the assay when no primary antibodies are used. Staining with primary antibodies against the different cell type-specific molecules (**Figure 22C**) revealed two major populations, which were KRT5+ (~41-55%) and ß-Tubulin IV+ (~31-40%). These were followed by CC10+ cells (~30%) and finally MUC5AC+ cells (~13-20%). Notably, these values sum up to more than 100%, which is, however, expected given the fact that a pool of transwells from different donors was used for the analysis

and that some cells may express more than one marker, which impedes their proper identification and unanimous classification (**Figure 22A**).

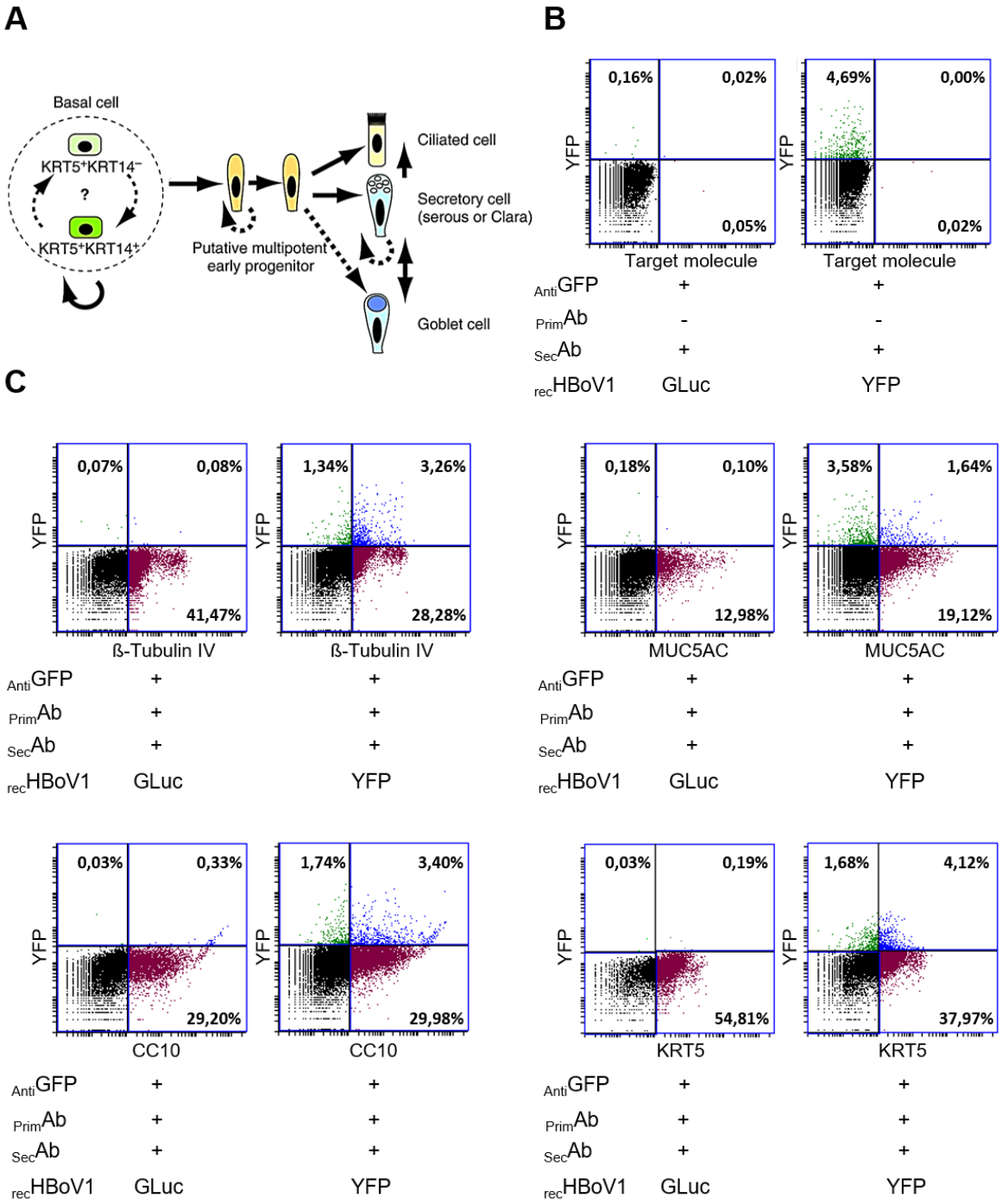


Figure 22. Flow cytometry analysis of pHAE transduced with rAAV/HBoV1.

(A) Model of basal stem cell self-renewal or differentiation into the indicated cell types (adapted from reference 277). **(B)** Two control graphs to set the gates used in all experiments. Cells were transduced with either scAAV-YFP/HBoV1 or scAAV-GLuc/HBoV1 (both at a MOI of 5×10^4) and fixed 10 days post-transduction. Next, the cell suspension was incubated with a FITC-coupled anti-GFP antibody and the Alexa-coupled secondary antibodies. **(C)** Cells were co-stained for YFP and the primary antibodies against the indicated cell type-specific markers: ß-Tubulin IV (ciliated cells), MUC5AC (goblet cells),

CC10 (club cells) or KRT5 (basal cells). The percentage of double-positive cells is shown in the upper right quarter. For all conditions, four transwells from different donors were pooled. Parts of this data set were published and adapted from Fakhiri *et al.* [252].

A FITC-coupled anti-GFP primary antibody was used to enhance the signal from virus-transduced cells (or to estimate the background in the negative controls) and co-stainings with the respective primary antibodies were performed (**Figure 22C**). The overall transduction efficiency with HBoV1 was between 5% (in this experiment) and 15% (**Figure 8D**, flow cytometry data not shown). Overall, KRT5+, CC10+ and ß-Tubulin IV+ cells were transduced with scAAV-YFP/HBoV1 to comparable efficiencies, whereas MUC5AC+ cells were less amenable to transduction (compare the upper right quarters, which show the double-positive cells).

Next, we asked whether a different preference may be detected for HBoV4 and GBoV, which would explain the lower total transduction efficiencies as compared to HBoV1 (**Figure 23**). As expected, the overall transduction efficiencies of GBoV and HBoV4 were less than those of HBoV1 with around 3% and 2% YFP+ cells, respectively. Interestingly, GBoV showed a pattern similar to HBoV1, whereas HBoV4 seemed to prefer KRT5+ cells and was less efficient at transducing ß-Tubulin IV + cells.

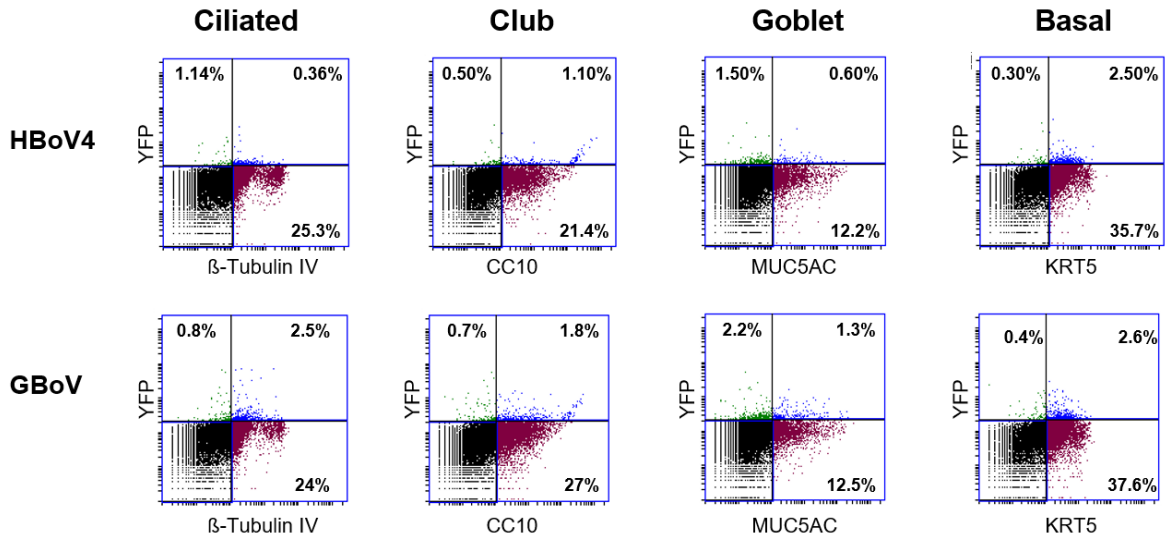


Figure 23. Flow cytometry analysis of pHAE transduced with rAAV/HBoV4 and rAAV/GBoV.

pHAE were transduced with scAAV-YFP/HBoV4 or scAAV-YFP/GBoV at a MOI of 5×10^4. The gates in the graphs were set according to a scAAV-GLuc/HBoV1 negative control shown in **Figure 22B**. Four transwells from different donors were pooled and fixed 10 days post-transduction. Next, the cell suspension was co-stained for YFP and the indicated cell type-specific markers: ß-Tubulin IV (ciliated cells), MUC5AC (goblet cells), CC10 (club cells) or KRT5 (basal cells). The percentage of double-positive cells is shown in the upper right quarter. This data set was published and adapted from Fakhiri *et al.* [252].

3.2.7 Differential effect of IVIg on pHAE transduction with rAAV/BoV

One hurdle in the application of viral vectors is the immune response of the host or the pre-existence of neutralizing antibodies (NAbs). The latter are particular problematic for gene therapies where the vectors are delivered via the blood, as pre-existing Nabs will neutralize the viral capsid and hence negatively influence the delivery and transduction efficiency of a viral vector. For example, ~70% of the human population have detectable antibodies against the most studied rAAV vector serotype 2 (AAV2) [110] and may thus be refractory to rAAV2 gene therapies. HBoV infections have also been detected globally and the amount of NAbs is estimated to be around 59%, 34%, 15% and 2% for HBoV1, HBoV2, HBoV3 and HBoV4, respectively [279].

In this part of the project, we asked whether the least seroprevalent HBoV4 and the non-human bocavirus GBoV display less reactivity than HBoV1 with human antibodies. Therefore, we adapted a previously described *in vitro* neutralization assay [280]. Briefly, virus particles were mixed with different dilutions of human serum (**Figure 24A**) or commercially available human intravenous immunoglobulins (IVIg), which is a pool of human IgG (**Figure 24B**). Next, the samples were incubated for 1 h at 37°C and were subsequently added to the apical surface of pHAE. Importantly, we used a serum sample from a hospitalized child, who suffered from a HBoV1 infection [281] and whose serum contained HBoV1-specific IgM and IgG antibodies, indicating an acute phase of the infection. An aliquot of this serum sample was kindly provided by Paul Schnitzler from the Diagnostic section of the Virology Department, University Hospital Heidelberg, Heidelberg, Germany.

The undiluted serum completely inhibited transduction with all three BoV vectors tested (HBoV1, HBoV4 and GBoV). As expected, increasing dilutions of the serum resulted in less inhibition of BoV transduction, as reflected in the increase of GLuc expression (**Figure 24A**). Notably, 100% and 68% of HBoV4 and GBoV transduction, respectively, were detected at the 1:3200 dilution of serum, compared to only 12.8% for HBoV1. These preliminary results imply a lower reactivity of HBoV4 and GBoV towards IgG-specific HBoV1 antibodies, which is in concordance with previous enzyme-linked immunosorbent assay (ELISA) data showing a certain degree of cross-reactivity of human antibodies to HBoV1-4 [279]. However, the impact of these interactions on functionality had not been shown prior to this doctoral work.

Next, we used our established assay to investigate the reactivity of HBoV1, HBoV4 and GBoV to IVIg at a concentration of 25 mg/dl, which represents a ~1:40 dilution of the normal average IgG concentration in adults. Due to the high variability of pHAE, a dilution of IVIg was chosen that results in a significant log-change in HBoV1 transduction, while remaining over background level (results of the pre-test are not shown). We observed a strong reduction of HBoV1 transduction (>10-fold) after incubation with IVIg. In contrast, HBoV4 and GBoV remained entirely unaffected (**Figure 24B**).

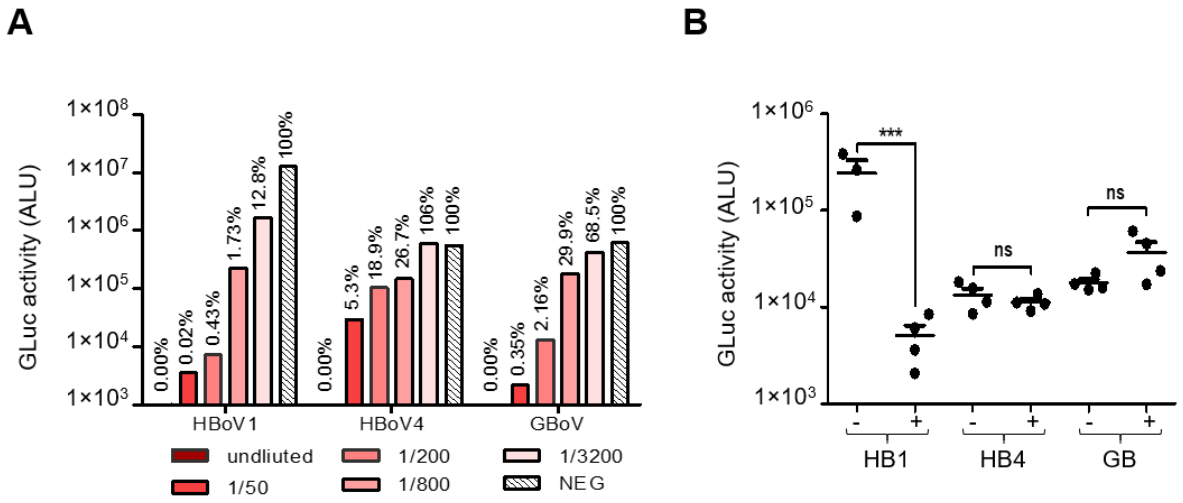


Figure 24. Effect of neutralizing antibodies on transduction efficiencies of HBoV1, HBoV4 and GBoV.

(A) A serum sample from a child with acute HBoV1 infection was mixed with PBS at the indicated dilutions and incubated with equal amounts (1×10^9) of viral particles. A negative control included virus only mixed with PBS (NEG). *In vitro* neutralization assays were performed as described above. GLuc activity was measured after nine days in culture and plotted on the y-axis as ALU. Numbers over the columns represent the mean percent transduction of the NEG control (n = 1). **(B)** *In vitro* neutralization assay using IVIg. Equal amounts of viral particles (5×10^9) were mixed with a 25 mg/dl IVIg solution or PBS as control. GLuc activity in the medium was measured after five days. Shown is the mean GLuc activity + SEM, n = 4; except for HB1 (-), n = 3. For statistical analysis, a one-way ANOVA with Tukey's multiple comparison test was used. Significance at $p<0.001$ is indicated by a triple asterisk. ns, non-significant. Panel B was adapted from Fakhiri *et al.* [252].

3.2.8 High susceptibility of primary human hepatocytes to BoV infection

The interesting connection of the BoV isolates studied here to the canine BoV3 revealed by the phylogenetic analysis of the *cap* ORFs (**Figure 12B**) tempted us to test primary human hepatocytes (pHeps) for their susceptibility to chimeric rAAV/BoV transduction. Therefore, monolayers of pHeps were transduced with equal amounts of viral particles at three different MOIs (2×10^4, 6×10^4 or 1.2×10^5). YFP expression started two days post-transduction (not shown) and was dependent on the applied MOIs (**Figure 25A**). Using GLuc as a reporter, we detected an increase in transgene expression from day 3 to day 6 in culture (**Figure 25B**), which is consistent with our previous data in pHAE (**Figure 18B**). Moreover, we found equal transduction efficiencies for all tested BoV vectors. This is especially intriguing for HBoV2 and HBoV3, which showed little to no expression in pHAE, implying different cellular specificities of the five BoV isolates.

3.2.9 Robust transduction of CD4+ cells by rAAV/BoV vectors

The hematopoietic system plays an important role in the dissemination of viruses in the body. This transport is usually mediated in the form of free virions or virus-infected cells. For

example, some arboviruses transiently infect blood cells in the acute phase of infection [282], whereas HIV causes a persistent infection [283].

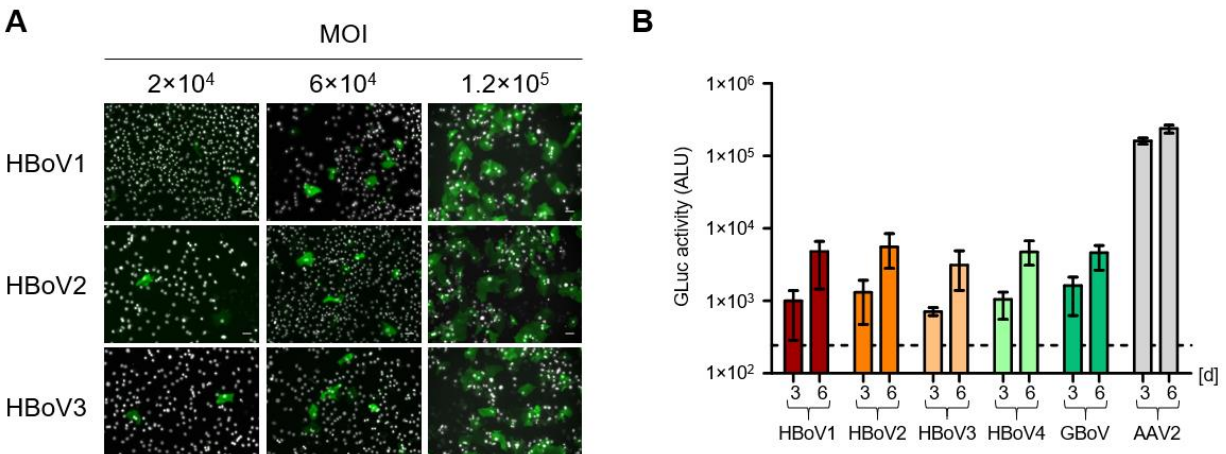


Figure 25. Transduction of pHeps with rAAV/BoV.

(A) Transduction of pHeps with scAAV-YFP at the indicated MOIs. Pictures were taken at day 5 post-transduction (n = 3 donors). Cells were fixed and stained with a FITC-coupled anti-GFP antibody. Gray = Hoechst staining of nuclei. Scale bar = 50 µm. **(B)** GLuc activity measured in the medium of cells transduced with the indicated scAAV-GLuc vectors (MOI of 1×10^4) at the specified time points after transduction. Data represent the mean and range of two independent measurements (n = 2 donors). This data set was published and adapted from Fakhiri *et al.* [219].

The detection of virus particles in the bloodstream is called “viremia” and can be either primary or secondary, depending on the course of infection. Several studies reported the presence of HBoV1 in the blood of symptomatic and asymptomatic individuals [284]. Based on this, we tested the ability of rAAV/BoV vectors to transduce three cell types in the blood (macrophages, PBMCs and CD4+ T-cells), which are common targets for virus infection. Measurements of GLuc activity at different time points post-transduction showed that macrophages were not susceptible to infection and PBMCs displayed low to no GLuc expression (**Figure 26A-B**).

By contrast, CD4+ T-cells were highly permissive to BoV transduction. In particular HBoV4 and GBoV resulted in high transgene expression levels comparable to AAV2 (**Figure 26C**). Importantly, in concordance with our previous observations in pHAE and pHeps (**Figure 18B and Figure 25B**), transgene expression increased over time in culture (compare day 3 with day 12 in **Figure 26C**).

We next aimed to validate this striking pattern observed in primary T-cells by transducing the cells with scAAV-YFP/BoV vectors. As hoped for and in line with the GLuc assay, YFP+

cells could be detected for all vectors. Particularly notable is the performance of HBoV4 and GBoV, which resulted in the highest number of YFP+ cells (13-15%) and nearly matched the efficiency of rAAV2 (16%) (**Figure 26D**).

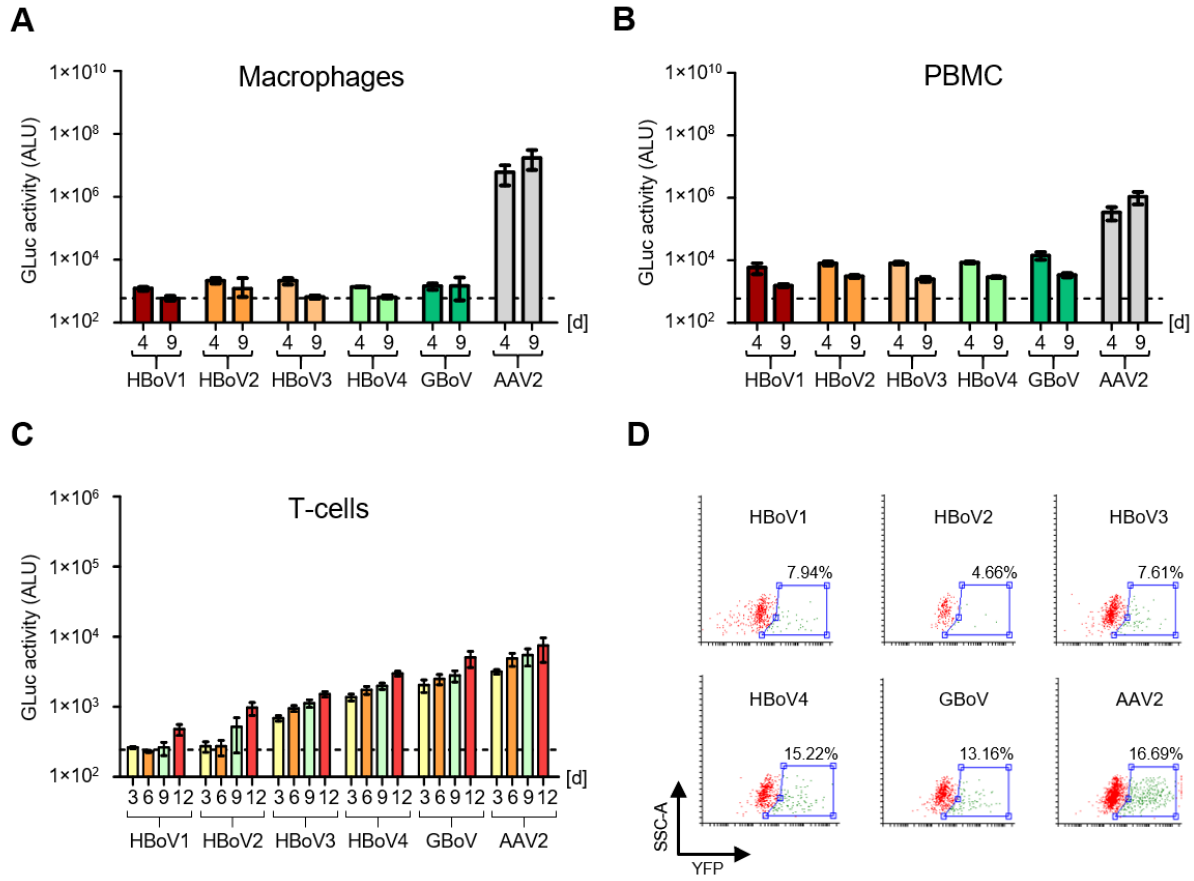


Figure 26. Transduction of primary blood cells with BoV vectors.

(A-C) Transduction of the indicated primary cells with scAAV-GLuc/BoV vectors or rAAV2 as control. Macrophages were incubated with a total of 5×10^8 viral genomes (n = 2 donors). PBMCs or T-cells were transduced at MOIs of 6×10^4 (n = 2 donors) or 1×10^4 (n = 3 donors), respectively. GLuc activity was measured in the medium at the indicated time points post-transduction. Plotted is the mean GLuc activity (± range for n = 2 or ± SEM for n = 3). **(D)** Flow cytometry analysis of primary T-cells transduced with the indicated scAAV-YFP/BoV vectors or rAAV2 as control at a MOI of 6×10^4. Transductions were performed in the presence of 1 µM doxorubicin. This data set was published and adapted from Fakhiri *et al.* [252].

3.2.10 Transduction of colon-derived cells and organoids with rAAV/BoV vectors

HBoV1 was frequently detected in surgically excised colorectal cancer specimen, which suggests a shedding of the virus from its primary site of replication (the lower respiratory tract) to the intestine [285, 286].

This assumption was strengthend by the ability of wtHBoV1 to infect and replicate in two colon cancer cell lines, namely, T84 [287] and Caco-2 [288]. However, the replication and amount of released progeny virus was less than in pHAE and the infectivity of the released progeny virus was not tested. In contrast to HBoV1, the other three human bocaviruses 2 to 4 were

commonly linked to gastroenteritis and thus believed to primarily infect the gastrointestinal tract [65, 262, 289]. In view of all these previous findings, we tested T84 cells and primary organoids from the intestine for their susceptibility to rAAV/BoV transduction.

In contrast to the study performed by Schildgen *et al.* [287], T84 cells were not polarized but maintained as monolayers under standard culturing conditions. Cells were incubated with equal amounts of scAAV-GLuc/BoV or AAV2 as control (MOI = 5×10^4). All vectors resulted in GLuc expression levels over background, with HBoV2-4 showing a slightly higher transduction ability than HBoV1 and the best results being obtained with GBoV.

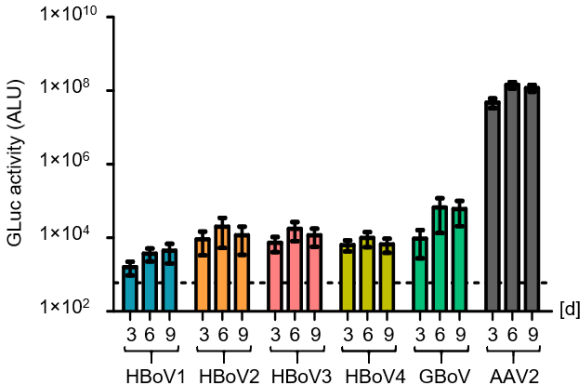


Figure 27. Transduction of T84 cell line with rAAV/BoV vectors.
T84 cells were cultured as monolayers and transduced with the indicated scAAV-GLuc/BoV vectors at a MOI of 5×10^4. GLuc activity in the medium was measured at the defined time points post-transduction. Plotted is the mean GLuc activity (± SEM for n = 3). This data set was published and adapted from Fakhiri *et al.* [252].

Yan *et al.* showed that rAAV/HBoV1 vectors result in 1,000-fold higher transduction efficiencies in pHAE, compared to the polarized CuFi-8 cell line [10]. Therefore, we tested whether the observed transduction in T84 cells could be enhanced in a primary organoid culture of polarized intestinal epithelium. Accordingly, we collaborated with the group of Steeve Boulant (Department of Infectious Diseases, University Hospital Heidelberg, Heidelberg, Germany), who has established a protocol for cultivation and transduction of intestinal organoids. Three different types of organoid systems were tested: (i) undifferentiated ileum, (ii) differentiated ileum, and (iii) colon organoids.

Interestingly, undifferentiated and differentiated ileum showed the lowest levels of GLuc activity for all rAAV/BoV vectors tested. Moreover, in contrast to our previous observations, transgene expression was decreasing and reaching background levels after 12 days in culture (**Figure 28A-B**). Colon organoids also showed this decrease in GLuc expression between day 3 and day 6 but resulted in a more robust and sustained expression after the initial drop (**Figure 28C**). This result was further extended by using YFP as reporter, which showed YFP+ cells in all colon organoids (**Figure 28D**) but not in the ileum organoids (data not shown).

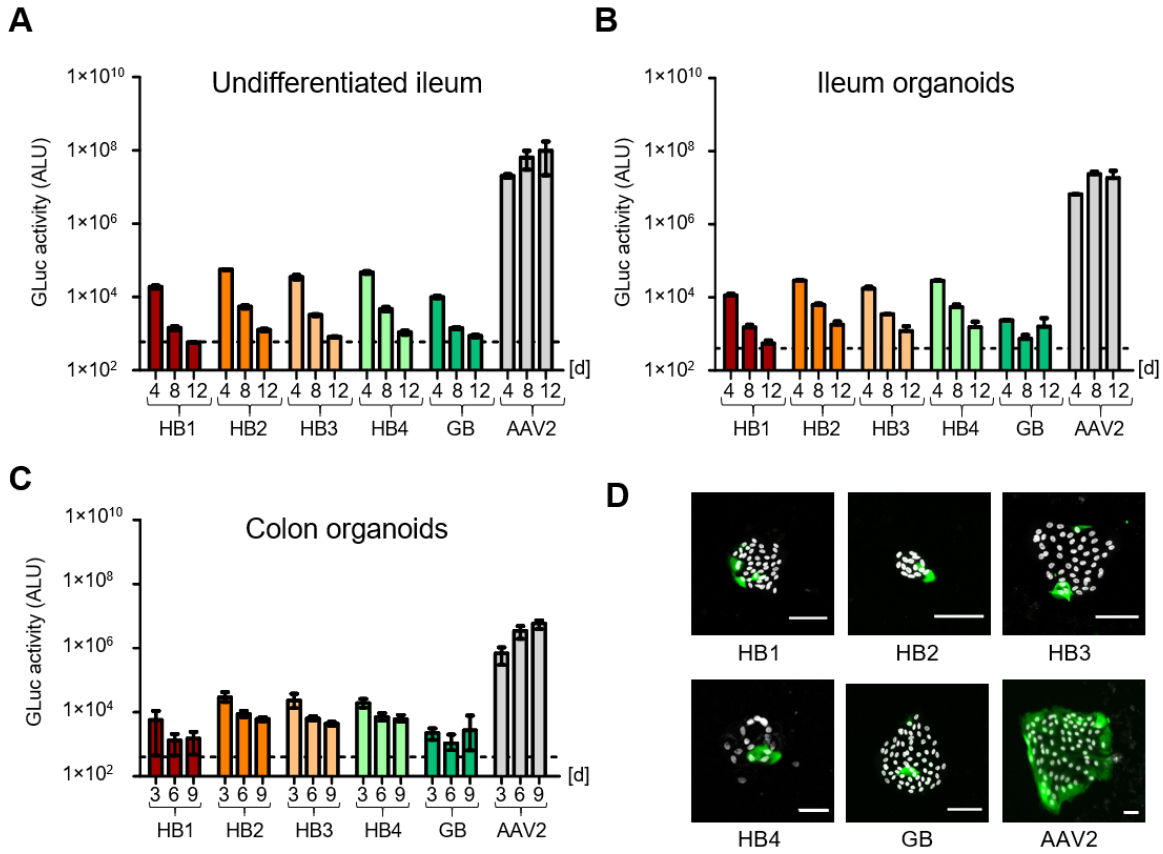


Figure 28. Transduction of undifferentiated and differentiated intestinal organoids.

(A-C) Transduction of the indicated intestinal organoids with 5×10^9 genome copies of scAAV-GLuc/BoV or rAAV2 as control. Plotted is the mean GLuc activity (± range for n = 2 donors in A and B or ± SEM for n = 3 donors in C). **(D)** Transduction of colon organoids with the indicated scAAV-YFP/BoV vectors or AAV2 as control (1×10^{11} genome copies per well). Cells were fixed at day nine post-transduction and then stained with a FITC-coupled anti-GFP antibody (n=2). Gray = Nuclei stained with Hoechst. Scale bar = 50 µm. The exposure was reduced three times for AAV2 to avoid saturation of the signal. Transductions were performed in the presence of 1 µM doxorubicin. Dashed lines indicate the assay background. This data set was published and adapted from Fakhiri *et al.* [252].

3.2.11 Detection of a broad cell tropism of primate BoV vectors *in vitro*

3.2.11.1 Transduction of an additional panel of primary cell types

Motivated by our previous finding of three primary cell types (pHAE, pHep and T-cells) that were efficiently and distinctly transduced with rAAV/BoV vectors, we further extended our screen for target cell types. Consequently, four additional cell types were tested, namely: (i) primary skeletal myoblasts and myotubes, (ii) cardiomyocytes, (iii) pulmonary fibroblasts, and (iv) vein endothelial cells.

Interestingly, undifferentiated muscle cells (myoblasts) showed the highest permissiveness for scAAV-GLuc/BoV transduction, especially for HBoV2 and GBoV (**Figure 29A**). Differentiated muscle cells (myotubes) were also highly susceptible to scAAV-YFP/BoV transduction (**Figure 29B**). This is intriguing in view of the fact that BoV infection has so far

never been associated with myopathies, in contrast to other muscle-infecting viruses such as rabies [290] and influenza [291]. By contrast, in two recent studies, HBoV1 DNA has been detected in heart tissue and linked to myocarditis [292, 293]. In our screening, all scAAV-GLuc/BoV vectors transduced cardiomyocytes to a comparable extent (except for HBoV3). However, the efficiencies were 10- to 200-fold lower than in skeletal muscle cells (**Figure 29A**).

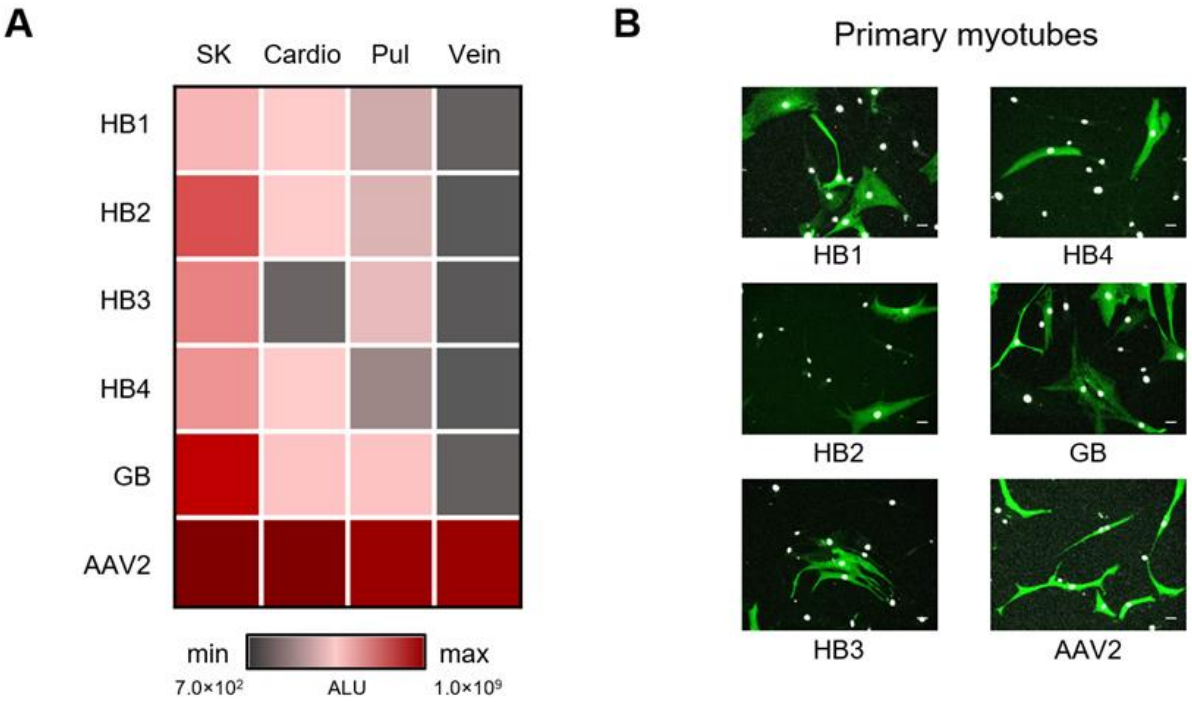


Figure 29. Transduction of numerous primary cells with rAAV/BoV vectors.

(A) Heat map showing the transduction of the indicated primary cells with the different scAAV-GLuc/BoVs at a MOI of 5×10^4. The values used for the generation of the heat map correspond to the measured GLuc activity in the medium nine days post-transduction (n=1 donor). **(B)** Transduction of primary myotubes with 1×10^9 viral genomes of scAAV-YFP/BoV vectors. Cells were fixed five days post-transduction and stained with a FITC-coupled anti-GFP antibody. Gray = Hoechst staining of nuclei. Scale bar = 50 µm. SK, skeletal muscle cells; Cardio, cardiomyocytes; Pul, pulmonary fibroblasts; Vein, saphenous vein endothelial cells. Transductions were performed in the presence of 0.5 µM doxorubicin. This data set was published and adapted from Fakhiri *et al.* [252].

We moreover tested non-polarized pulmonary fibroblasts, which produce the extracellular matrix and thus play an important role in the physical support and function of the lung [294]. In our initial experiments, only three out of the five tested vectors, namely, HBoV1, HBoV4 and GboV, were able to transduce pHAE (see **Figure 18A-B**)**.** In primary fibroblasts, however, this pattern was not maintained, and a nearly equal transduction efficiency was detected for all rAAV/BoV vectors.

Finally, we tested vein endothelial cells, which are a known target cell type for the replication of many viruses [295]. These cells were, however, the hardest to transduce with little to no measured GLuc activity.

3.2.11.2 Transduction of cell lines

One hurdle in the application and pre-clinical study of rAAV/BoV vectors is their preference for primary cells, which are limiting, expensive and usually difficult to culture. Thus, we tested several immortalized cell lines from different tissues for their susceptibility to rAAV/BoV transduction (**Figure 30**).

Consistent with our data in primary hepatocytes, the two liver-derived cell lines Huh 7 (hepatocytes) and LX-2 (stellate cells) were permissive to rAAV/BoV vectors, albeit at lower efficiencies than the primary cells. In contrast, PancI (pancreatic cancer cells) and Raw 267.4 (murine macrophages) cells were largely resistant to BoV infection. Notably, transduction of the latter mouse cell line could have been hampered by the pronounced species-specificity of BoV infection.

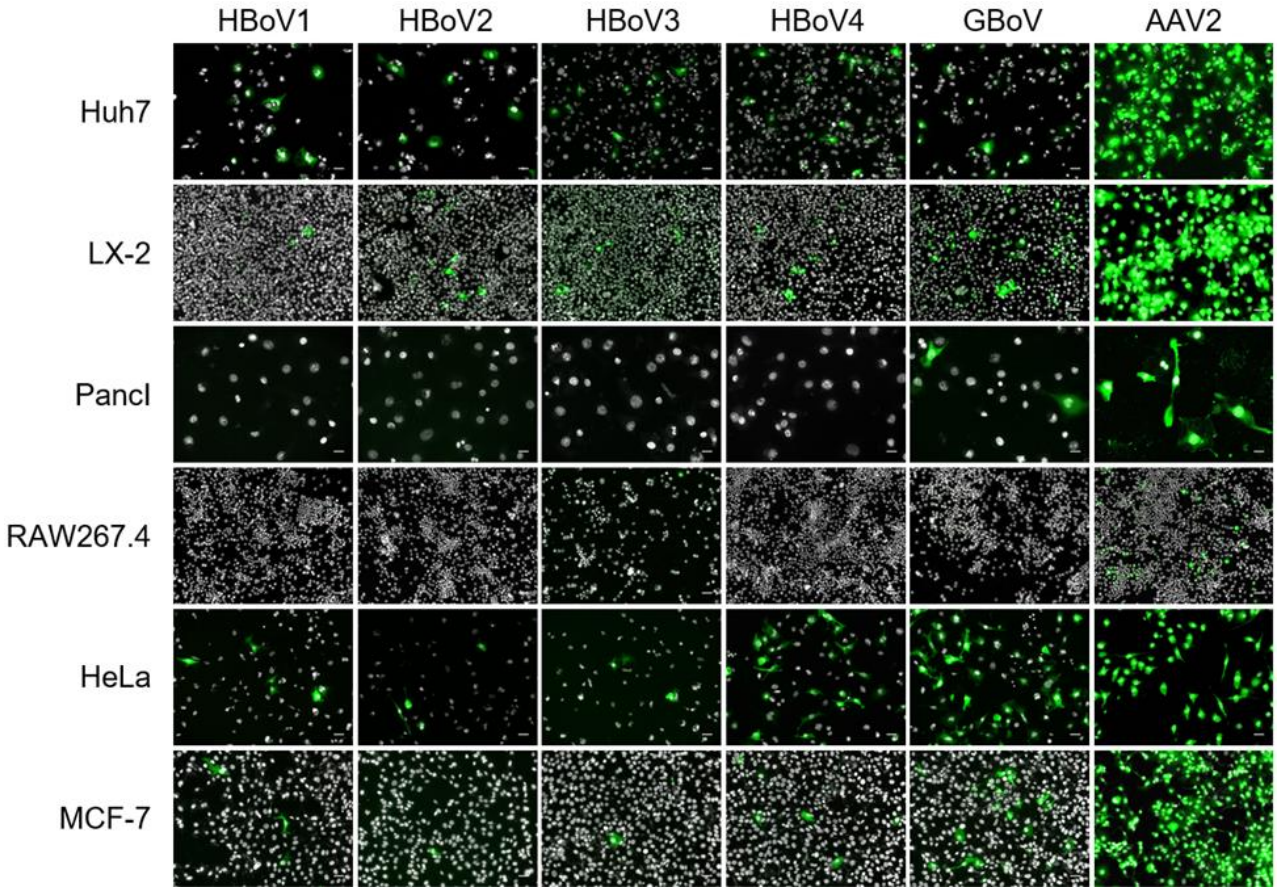


Figure 30. Transduction of cell lines with rAAV/BoV vectors.

The indicated cell lines were transduced with scAAV-YFP/BoV vector or rAAV2 as control at a MOI of 2×10^5. Cells were fixed five days post-transduction and stained with a FITC-coupled anti-GFP antibody (n=3). Nuclei were stained with Hoechst (gray). Scale bar = 50 µm. This data set was published and adapted from Fakhiri *et al.* [252].

Surprisingly, two additional human cell lines, MCF-7 (breast cancer) and HeLa (cervix carcinoma), displayed many YFP+ cells, especially for HBoV4 and GBoV. This efficiency in HeLa cells is of note in view of an ongoing debate about the association of BoV infection with miscarriage based on sporadic detection of HBoV DNA in placenta and aborted tissue [296].

3.2.12 Development of a shuffled BoV library

The tropism of a virus is defined as its specificity for particular cells within a host. During the course of evolution, viruses can change their tropism, host range and virulence to adapt to new environments. It has been shown that these evolutionary mechanisms can be mimicked and accelerated by harnessing the power of directed evolution [111, 297]. This process starts with the creation of a highly diverse pool of viruses, which is then subjected to a negative and/or positive selective pressure to enrich viral particles with a specific feature. These pools (libraries) can be generated either (i) from single serotypes, for example, by random mutagenesis of the *cap* gene [298], or (ii) through combining different serotypes using *in vitro* recombination [111, 297] (DFS, DNA family shuffling). Both approaches have been shown to be highly promising for the generation of novel AAV vectors for gene therapy [299].

Here, we tested whether DFS can be applied to the primate BoVs used in this work akin to AAVs. To this end, we adopted a protocol previously reported for AAV family shuffling [111]. Briefly, the BoV *cap* sequences were PCR-amplified and then fragmented using a DNase I digest. The fragments were loaded on a 1% agarose gel and a range of fragment sizes between 100 bp and 1 kb was excised, to serve as input for a second assembly PCR that results in chimeric, full-length fragments of ~2 kb (**Figure 31A**). This part was performed together with Stefan Holderbach, and details about the optimization of the process can be found in his BSc. thesis as well as in Fakhiri *et al.* [252].

Next, the chimeric *cap* sequences were cloned into the acceptor plasmid pAAVNSΔVP-1×BsmBI, which resulted in a chimeric BoV plasmid library (hereafter referred to as 5-component library to indicate the five underlying parental serotypes), and 10 randomly picked clones were sent for sequencing analysis. The obtained sequences were then aligned in MEGA-X (using MUSCLE) and saved as a FASTA file. This file was subsequently uploaded into our in-house shuffling alignment analysis tool (SALANTO) [239] to perform a quantitative and qualitative analysis of the library (**Figure 31A**).

As hoped for, the high sequence identities between the primate BoV *cap* sequences (see **Figure 12A** and **Figure 31B**) enabled frequent recombination events (on average, nine per clone). Importantly, we observed a similar distribution of HBoV2, 3, 4 and GBoV (23%, 21%, 28% and 18%, respectively), whereas HBoV1 was less frequent with 8%. Moreover, congruent with previous observations with AAV *cap* shuffling [300], the numbers of cross-over events positively correlated with the degree of sequence identity and was thus highest among HBoV2-4, which share 87-88% identity (**Figure 31B**).

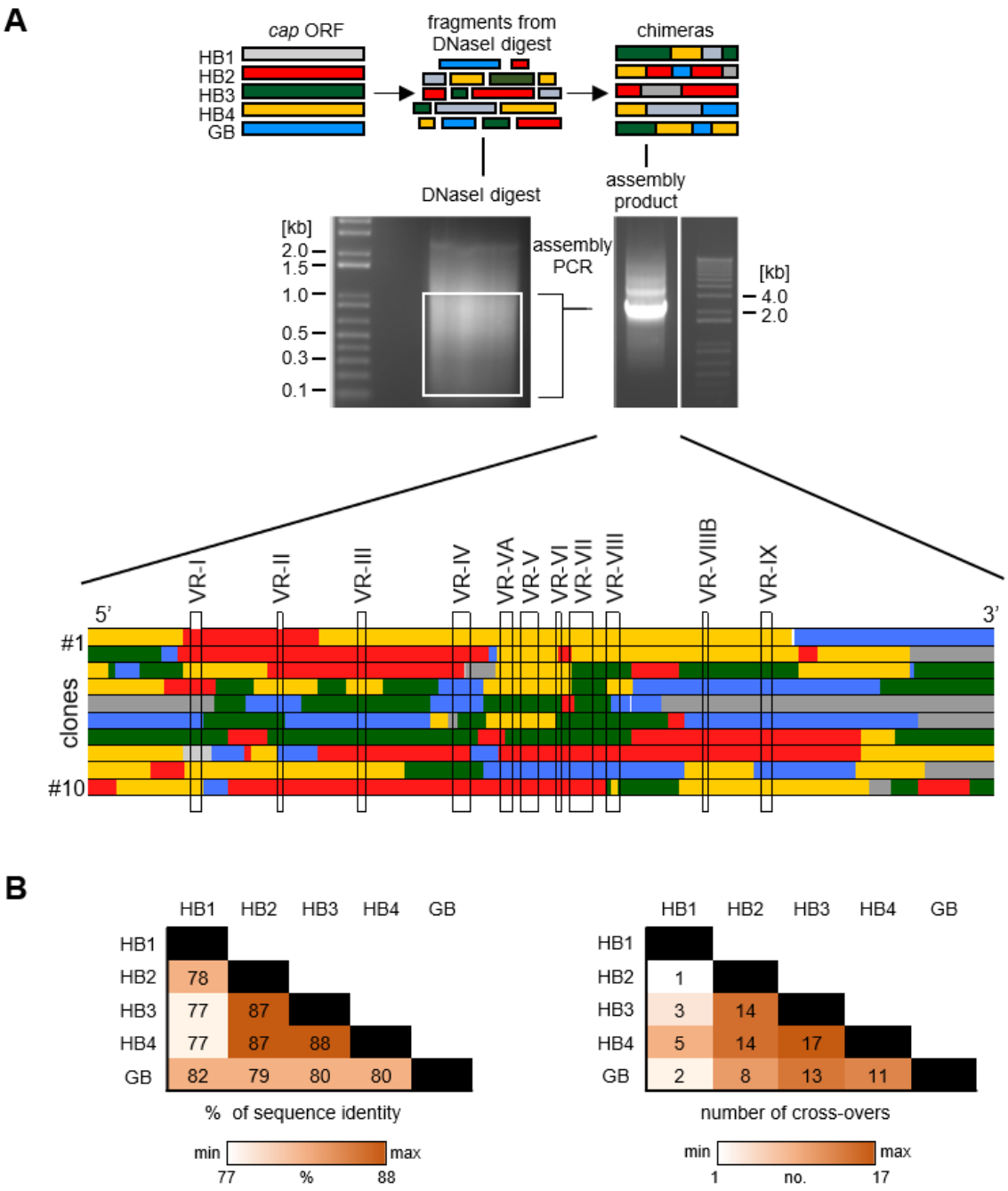


Figure 31. DNA family shuffling of BoV *cap* sequences and construction of 5-component library.

(A) The indicated BoV *cap* ORFs were PCR-amplified and fragmented using a DNase I digest. Then, the fragments were loaded on a 1% agarose gel and excised, to serve as a template for the assembly PCR. The ~2 kb *cap* product was used for the generation of the displayed chimeric library. Shown are the sequence compositions of 10 clones (5' to 3') with the BoV variable regions (VR) indicated. The assignment of sequences to the most likely parental sequence was done in SALANTO. **(B)** The first heat map shows the DNA sequence identity of the different BoV *cap* sequences. The second heat map shows the pairwise crossover events calculated from the displayed representative clones. Gel pictures origin: Stefan Holderbach, BSc. thesis. Light gray = ambiguous sequences. This data set was published and adapted from Fakhiri *et al.* [252].

The acceptor plasmid pAAVNSΔVP-1×BsmBI harbours a modified HBoV1 viral genome flanked by AAV2 ITRs and encodes the HBoV1 NS and NP1 proteins. Seamless insertion of

the chimeric BoV *cap* ORFs resulted in a chimeric plasmid library. When this library was transfected along with pDGΔVP into HEK293T cells, DNase I-resistant particles (*i.e.* encapsidated viral genomes) could be detected in a density range of 1.47 to 1.35 g/mL, as revealed by dot blot analysis of different fractions of the CsCl gradient (**Figure 32**, upper panel). This matches the density range previously observed for the different single BoV capsids (**Figure 15**). To study the diversity of the library after virus production, viral DNA was isolated from the virus stock and used as a template for amplification of the *cap* ORF. Several PCR reactions were needed to gain enough PCR product for cloning into an acceptor plasmid. Five positive clones could be identified and were sent for sequencing analysis (**Figure 32**, lower panel). Despite the low number of clones, which hampers a quantitative analysis, we noted a site-specific accumulation of both, HBoV4 and HBoV3, especially HBoV4-derived sequences in the variable region 3 (VR III).

After confirming the diversity of the viral library, we performed several selection rounds in pHAE to enrich for candidates with novel properties like enhanced transduction and/or packaging ability (**Figure 32**, lower panel).

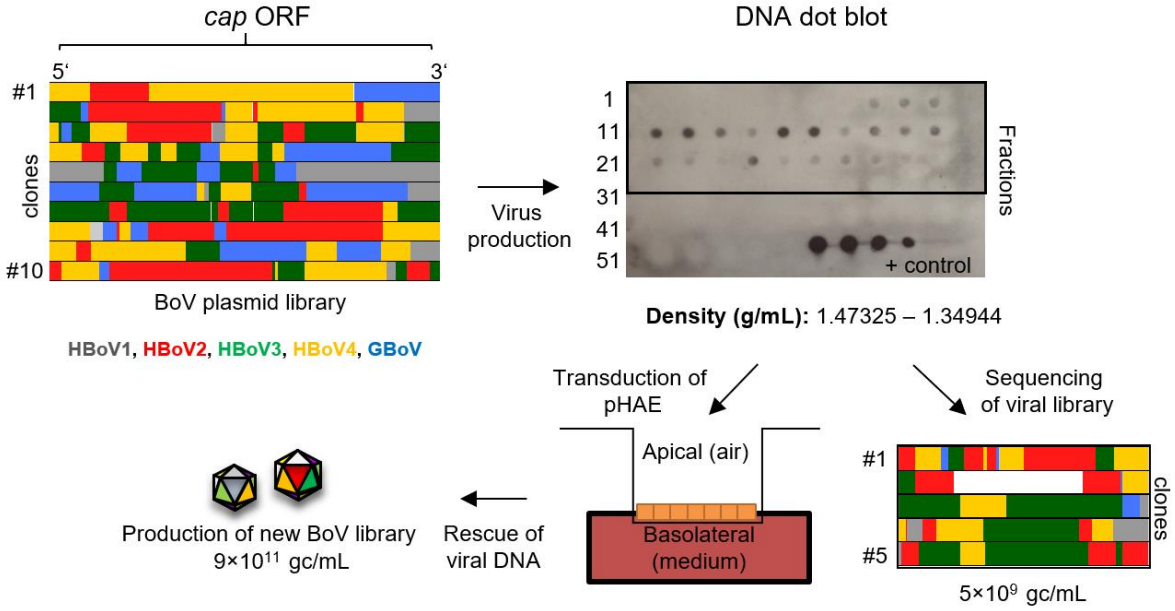


Figure 32. Production and screening of chimeric BoV libraries.

The chimeric BoV plasmid library (represented here by 10 clones) was transfected into HEK293T cells for virus production followed by CsCl density centrifugation. Fifty-two fractions were collected and tested for the presence of DNase I-resistant particles using DNA dot blot analysis. The viral library was then analyzed for its clonal composition, which is exemplified here with five clones. Concurrently, pHAE were incubated with ~5×10^8 viral genomes O/N. The next day, cells were washed to remove unbound virus and lysed for extraction (rescue) of viral DNA. This served as a template for PCR amplification of enriched *cap* sequences followed by cloning of these sequences into an acceptor plasmid. The resulting plasmid library was then used to generate a second viral library. Gray = HBoV1, red = HBoV2, green = HBoV3, yellow = HBoV4, blue = GBoV. White = unsequenced areas and light gray = ambiguous sequences.

3.3 Cycling of a recombinant BoV library in primary airway epithelia

3.3.1 Selection in the presence or absence of proteasome inhibitors

Transduction of pHAE with the chimeric BoV library described in 3.2.12 was done in the presence or absence of the two PIs, doxorubicin and LLnL (**Figure 33A-B**). These were typically added to ensure and enhance transduction with rAAV/BoV vectors (**Figure 11A-B**). Interestingly, in both conditions, the observed site-specific accumulation of HBoV4 in the VR III after packaging (**Figure 32**, lower panel) was retained in >50% of the clones. By contrast, HBoV3 became less abundant and was mostly replaced by HBoV4-derived sequences (VR-IV to VR-VI). The only difference between the two conditions was observed in VR-VIIIB and VR-IX. In the presence of PIs, we noted an enrichment of HBoV2 sequences at the C-terminus, which was not evident in the absence of PIs. In the next experiments, we thus focused on selections in the presence of PIs as this gave the most striking pattern, with the aim to validate and further extend these results.

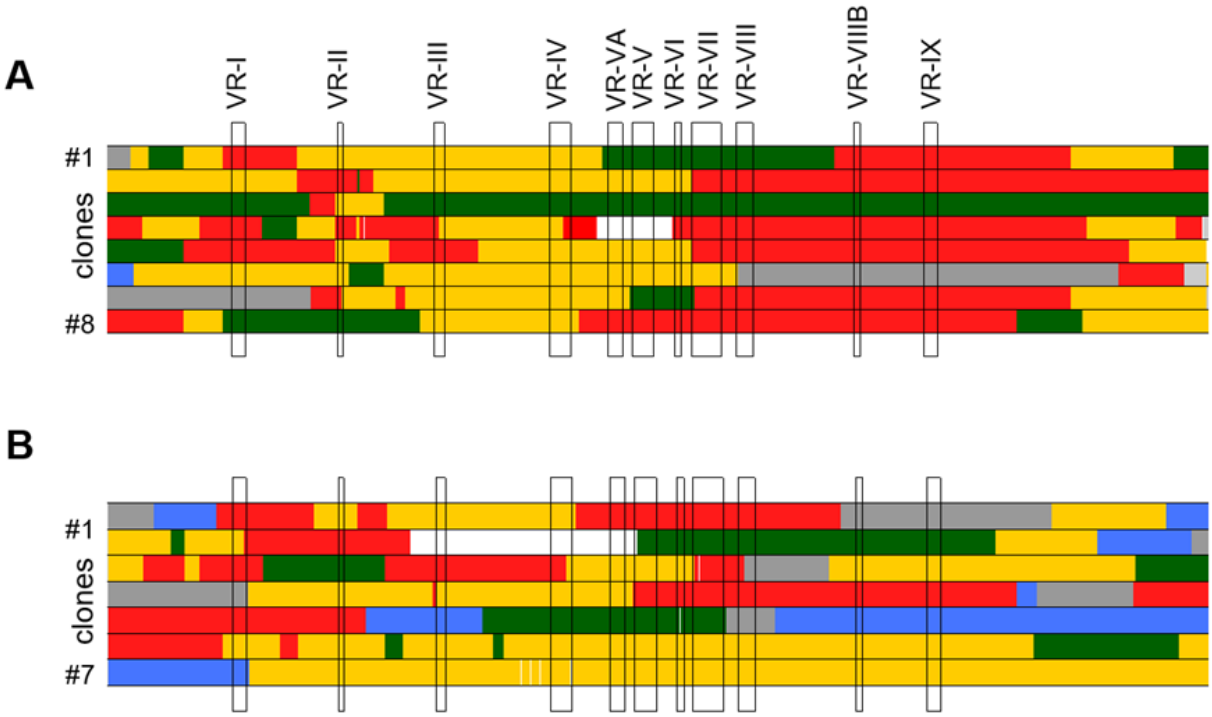


Figure 33. Analysis of chimeric BoV *cap* sequences after the first round of selection in pHAE.

Transductions of pHAE with the 5-component BoV library were performed either in the presence **(A)** or absence **(B)** of LLnL and doxorubicin at 40 µM and 5 µM concentrations, respectively. Alignments were performed in MEGA-X, while assignment of sequences to the underlying BoV serotypes was done in SALANTO. Gray = HBoV1, red = HBoV2, green = HBoV3, yellow = HBoV4 and blue = GBoV. White = unsequenced areas, light gray = ambiguous sequences.

3.3.2 Additional rounds of molecular evolution revealed new emerging patterns in the BoV capsid

Already in the unselected, viral BoV library, we had observed a first layer of selection that has probably occurred during production of the packaged viral library and has resulted in a bias towards sequences derived from HBoV3 and HBoV4 (**Figure 32,** "viral library" in lower panel). Over the following three selection rounds in pHAE (**Figure 33A** and **Figure 34**), the selective accumulation of HBoV4 in VR III was retained and culminated in a dominant presence in all sequenced clones after selection round 3. Moreover, the overall percentage of HBoV4-derived sequences increased from 18% in selection round 1 to 34% after round 3 and included VR-VI and VR-VII.

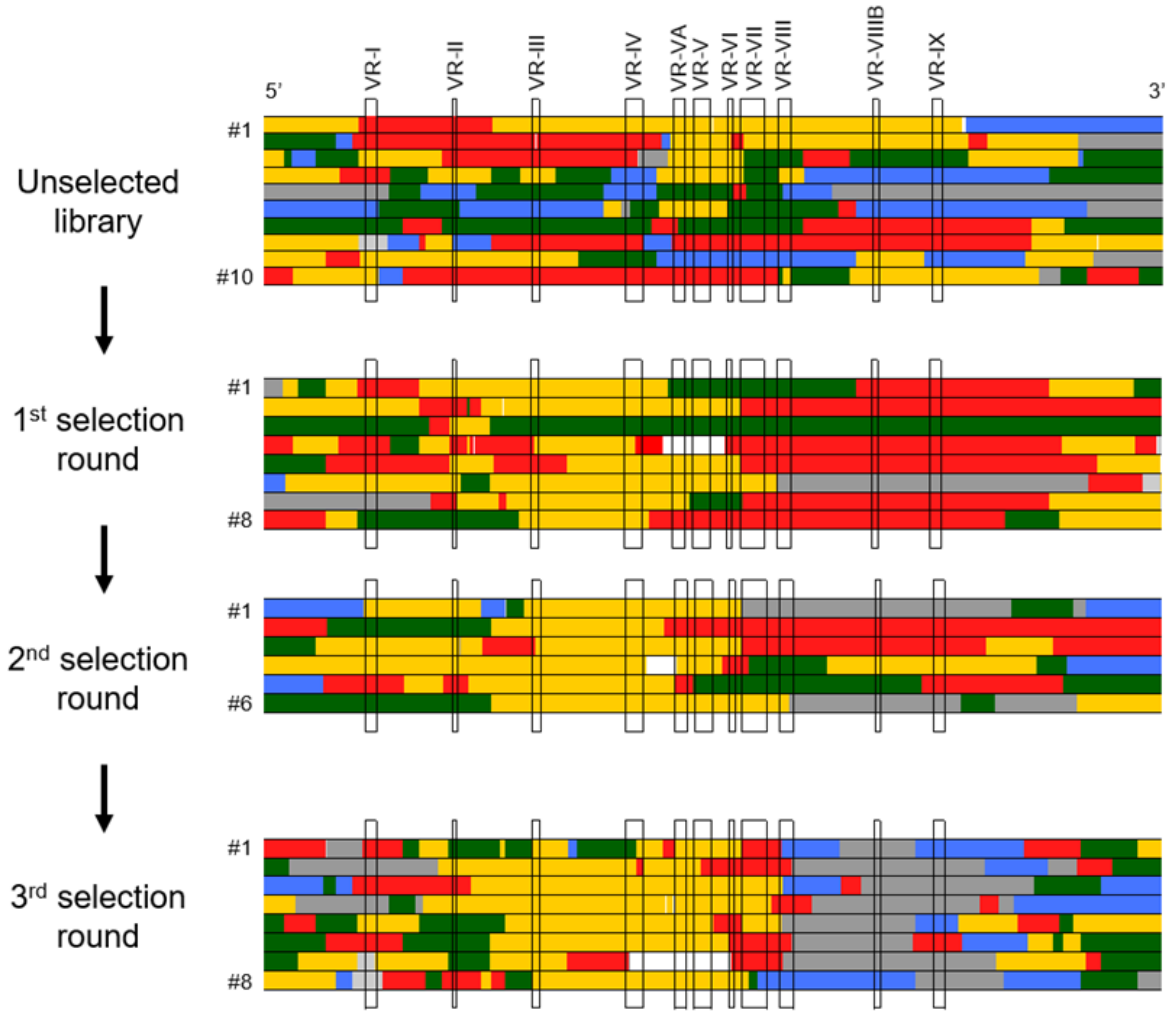


Figure 34. Cycling of chimeric BoV library in pHAE.

The unselected library refers to the BoV plasmid library before virus production and has already been shown in **Figure 32**. Likewise, the library composition after round 1 has already been shown in **Figure 33A**. Both are reprinted here for direct comparison with the results after rounds 2 and 3. Each "selection round" included three steps: (i) inoculation of pHAE with the virus O/N, (ii) harvest of the cells and isolation of viral DNA, and (iii) amplification of rescued *cap* sequences and preparation of a new virus library. Alignments were performed in MEGA-X. Regions were assigned using SALANTO. HBoV1 = gray, HBoV2 = red, HBoV3 = green, HBoV4 = yellow, GBoV = blue. White = unsequenced areas, light gray = ambiguous sequences.

Another interesting observation is the emergence of HBoV1 and GBoV at the C-terminal part of the *cap* ORF, involving VR-VIIIB and VR-IX, while these two serotypes were rather underrepresented in the original library. By contrast, the N-terminal part of *cap* including VR-I and VR-II remained diverse with no clear serotype or sequence preference. This region involves the highly conserved C-terminal end of NP1 and the phospholipase 2A domain [301]. The VR-II showed a slight bias towards HBoV3 (which has a V instead of a T at amino acid position 146, VP2 numbering) but it is highly conserved in sequence and structure among primate BoVs. This is in line with its important function in externalization of the VP1u domain and packaging of the viral genome [302]. In contrast to VR-II, VR-I is highly heterogenous among BoVs.

Moreover, we noted a 40- to 180-fold increase in the overall viral titer, from 5×10^{9} gc/mL of the unselected library to 2-9×10^{11} for the libraries generated after rounds 1 and 2 (qPCR data not shown).

3.3.3 Chimeric BoV capsids show improved packaging efficiency but less potent transduction

One aim and hope underlying our shuffling and selection approach was to select for viruses that are more infectious. To test whether this applies to the variants we had gained after selection round 3, we randomly chose five candidates and cloned the respective *cap* sequences into our BoV helper (pCMVNS1*HBoV1) that lacks AAV or BoV TRs. Next, a scAAV-GLuc reporter was packaged into each capsid and the number of viral genomes per mL was determined using qPCR analysis of the 40% fraction after iodixanol purification of viral particles (**Figure 35A**).

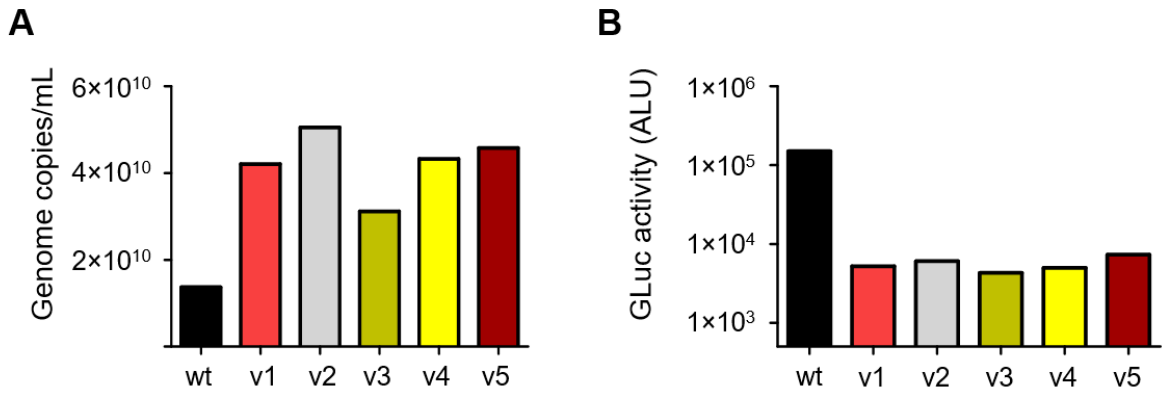


Figure 35. Packaging and transduction efficiency of single chimeric BoV capsids.

(A) Five randomly selected BoV *cap* variants were used to package scAAV-GLuc and the vector stocks were titrated using TaqMan qPCR (n=1). **(B)** Transduction of pHAE at a MOI of 1×10^{4}. GLuc activity was measured in the medium five days post-transduction and plotted on the y-axis as ALU. wt = benchmark vector based on wild-type HBoV1 *cap* (n=1 donor).

Interestingly, all the tested chimeras showed an enhanced, 2- to 4-fold higher packaging efficiency, as compared to the wild-type counterpart (scAAV-GLuc/HBoV1, denoted as wt in **Figure 35A**). This is congruent with the overall increase in library titers after the first selection round. By contrast, transduction of pHAE with equal amounts of viral particles revealed a lower transduction ability of the chimeric capsids as compared to the HBoV1 capsid, which gave the highest transduction (**Figure 35B**). Possible reasons for this unexpected but informative and thus important result will be discussed later (chapter 4).

3.4 Development of CRISPR-based approaches to control rAAV vector persistence

One of the major hurdles in clinical gene therapy using rAAV or other viral vectors with the capacity to persist in cells is the limited ability to control transgene expression at will, including the option of a complete shutdown, if needed. Thus, a major goal in this second part of the work was to develop regulatory genetic circuits that would permit to terminate an ongoing therapy or to restrict the duration of transgene expression from rAAV vectors. Importantly, because these technologies act on the level of the vector genome, they are independent of the capsid and thus fully compatible with the rAAV/BoV vectors developed in the first part of this doctoral work.

The particular approach we pursued here was to take advantage of the CRISPR/Cas9 technology, by developing two types of Cas9-inactivatable rAAV vectors: (i) a so-called "KS" (kill switch) rAAV vector that expresses a transgene and a gRNA towards the vector itself, thus allowing transgene expression to be shut off by introducing Cas9 *in trans*, or (ii) "SIN" (self-inactivating) rAAV-CRISPR constructs that simultaneously target the transgene and the Cas9. The latter concurrently allows to control gene expression from rAAV vectors and prevents long-term Cas9 expression, which is crucial since the persistent presence of Cas9 has been shown to increase the risk of adverse off-target activity [303] and to provoke a cellular immune response [304].

3.4.1 Screening for functional gRNAs against Firefly luciferase

It has been shown that rAAV vectors predominantly persist in transduced cells as episomes or large concatemers [107]. Thus, to first test whether these structures can be targeted by Cas9, we chose Firefly luciferase (FLuc) as a “transgene” because of its high sensitivity and broad dynamic range as compared to other reporters.

Next, we designed seven gRNAs against different regions of the FLuc ORF or the poly A tail by using the CRISPR MIT tool (https://zlab.bio/guide-design-resources). Targeting a region outside the transgene (here, the poly A signal) whose disruption leads to an efficient

knockout of associated transgene expression would allow the design of universal gRNAs that control different transgenes sharing the same regulatory elements.

The functionality of the selected gRNAs was tested in HEK293T cells by using triple plasmid transfection of (i) a gRNA, (ii) psiCheck-2™ (a dual-luciferase reporter plasmid expressing both, FLuc and *Renilla* luciferase (RLuc)), and (iii) SpCas9. A luciferase assay was performed three days after transfection, and FLuc activity was divided by that of RLuc (which remained unaffected by the gRNA and could thus be used for normalization) to account for differences in transfection efficiencies between the wells. Next, these values were further normalized to a control gRNA targeting exon 10 of the *CFTR* locus to calculate the knockout efficiency (normalized Firefly activity, **Figure 36A-B**).

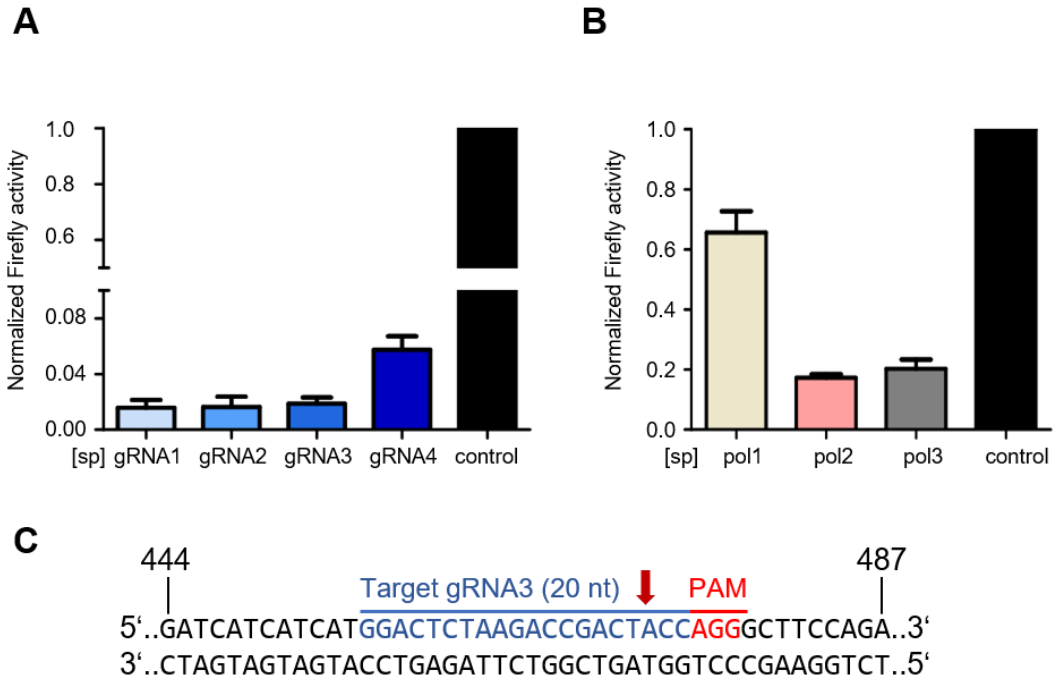


Figure 36. Functional validation of Sp gRNA cleavage *in vitro*.

(A-B) HEK293T cells were seeded in 96-well plates at a density of 3×10^4 cells/well. After 24 h, the cells were triple-transfected with (i) a gRNA (gRNA1-4, pol1-3 or control), 75 ng/well, (ii) the SpCas9, 125 ng/well and (iii) psiCheck-2, 10 ng/well. SpCas9 cleavage efficiencies with the seven anti-luciferase gRNAs targeting the FLuc ORF in (A) or the poly A signal in (B) were evaluated by dual-luciferase assay. FLuc and RLuc activities were measured in the cell lysates three days post-transfection. Data represent the mean relative FLuc activity per well (mean ± SD, n = 3) normalized to RLuc activity as well as to the negative control gRNA (set to 1). **(C)** DNA target region of gRNA3 in the FLuc ORF. The target sequence is 20 nt in length and is adjacent to a 5'-NGG PAM sequence (N can be any nucleotide; here it was A). Cleavage occurs three nt upstream of the PAM, as indicated by the arrow. Numbers refer to the nucleotide position within the FLuc ORF.

Intriguingly, all gRNAs were functional and resulted in a robust decrease in FLuc activity of ~40-95%. Notably, the gRNAs targeting the FLuc ORF were on average 10-fold more efficient than those targeting the poly A signal. One of the best performing gRNAs (gRNA3, **Figure 36C**) was cloned together with the H1 promoter into a ssAAV vector expressing both,

FLuc and RLuc flanked by AAV2 ITRs (ssAAV-FLuc-RLuc), resulting in construct ssAAV-FLuc-RLuc-H1-g3-SP (below referred to as KS Sp reporter; KS for kill switch).

During the course of this study, an alternative Cas9 ortholog from *Staphylococcus aureus* (SaCas9) was described, which is ~1.1 kb smaller than SpCas9 and thus offers a size advantage in the AAV context, while displaying a similar level of activity [12]. Consequently, we implemented and tested this system in parallel, by designing four compatible gRNAs targeting the same FLuc ORF. As observed with the SpCas9 system, all the tested gRNAs were functional and resulted in varying degrees of FLuc knockout efficiencies (**Figure 37A**). Also here, one of the best performing gRNAs (gRNA1, **Figure 37B**) was cloned under the H1 promoter and the entire cassette was subcloned into the ssAAV-FLuc-RLuc plasmid, resulting in construct ssAAV-FLuc-RLuc-H1 g1 Sa (below referred to as KS Sa reporter).

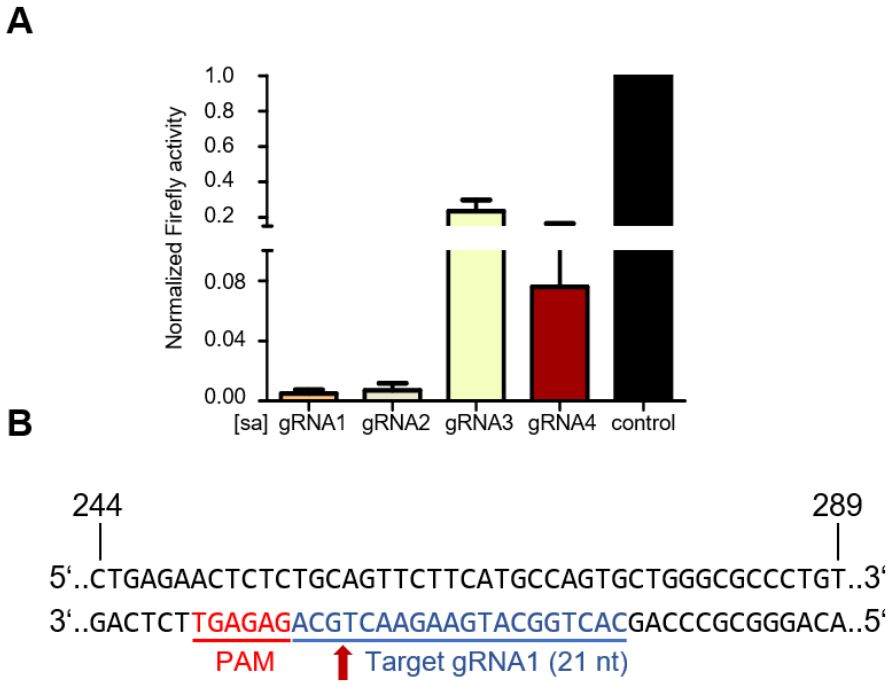


Figure 37. Functional validation of Sa gRNA cleavage *in vitro*.

(A) HEK293T cells were seeded in 96-well plates at a density of 3×10^4 cells/well. After 24 h, the cells were triple-transfected with (i) a gRNA (gRNA1-4 or a control gRNA), 75 ng/well, (ii) the SaCas9, 125 ng/well and (iii) psiCheck-2, 25 ng/well. SaCas9 cleavage efficiencies with four anti-luciferase gRNAs targeting the FLuc ORF were evaluated by dual-luciferase assay. FLuc and RLuc activities were measured in the cell lysates three days post-transfection. Data represent the mean relative luciferase activity per well (mean ± SD, n = 3) normalized to RLuc activity as well as to the negative control gRNA (set to 1). **(B)** DNA target region of gRNA1 in the FLuc ORF. The target sequence is 21 nt in length and is adjacent to a 5'-NNGRRT PAM sequence (note that it is shown in reverse here, since the 3'-5' strand is targeted). Cleavage occurs at the indicated site, three nt upstream of the PAM. Numbers refer to the nucleotide position within the FLuc ORF.

One observation we made during these initial experiments is that RLuc activity was also significantly decreased under our Cas9/gRNA cleavage conditions, as compared to the negative control (**Figure 38A**). As none of the gRNAs was predicted to bind in the RLuc ORF, we speculated that the loss in RLuc activity might reflect the complete degradation of the FLuc/RLuc co-encoding plasmid after Cas9 cleavage. To test this hypothesis, we designed two short-hairpin (sh)RNAs that target the FLuc mRNA and tested their activity in HEK293T cells (**Figure 38B**). Our underlying rationale was that this would allow us to specifically suppress FLuc expression using an independent trigger and to then assess whether RLuc expression would again be co-affected, which would have suggested their coupling (by unknown means) on the mRNA or protein level. One of the two shRNAs (sh2) significantly reduced FLuc activity by ~70%, without affecting the expression of RLuc. Accordingly, we concluded that the expression of FLuc is not coupled to RLuc and that the concurrent downregulation of FLuc and RLuc observed with CRISPR but not RNAi is indeed related to the loss of the common plasmid DNA. Thus, having learned that RLuc expression is automatically affected in the Cas9 cleavage assays, we omitted the RLuc normalization step in all subsequent experiments.

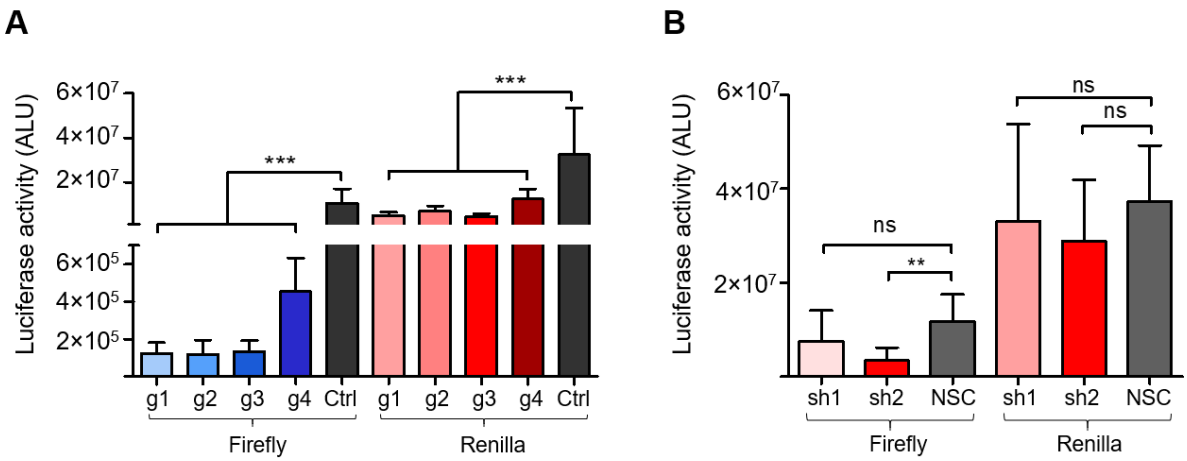


Figure 38. Comparison of knockout and knockdown experiments in HEK293T cells using the dual-luciferase reporter.

HEK293T cells were seeded in a 96-well at a density of 3×10^4 cells/well. After 24 h, cells were transfected for Cas9 knockout experiments in **(A)** or shRNA-mediated knockdown in **(B)**. For the FLuc knockout experiments, a triple transfection was performed (gRNA or control 75 ng/well, SpCas9 125 ng/well and psiCheck-2 10 ng/well). Ctrl = Control non-targeting gRNA against *CFTR* exon 10. For the knockdown experiments in **(B)**, 40 ng/well of the indicated shRNAs (sh1-2) or non-silencing control (NSC) was transfected. FLuc and RLuc activities were measured in the cell lysates three days post-transfection. Plotted is the mean luciferase activity as ALU (mean ± SD, n = 3). For statistical analysis, a one-way ANOVA with Dunnett's multiple comparison test was used. Significance is indicated by: **, p<0.01, ***, p<0.001, ns, non-significant.

3.4.2 *In vitro* validation of the rAAV KS reporter

The above-mentioned KS Sa and KS Sp reporters were called kill-switch vectors because they constitutively express a gRNA complementary to a sequence in the FLuc ORF. Importantly, the whole system can be deliberately “switched on” at any given time by expressing or providing Cas9 *in trans*, which then, together with the gRNA, triggers the degradation of the vector and thus ablates transgene expression (**Figure 39A**).

The knockout efficiencies of both reporters were measured and compared in a dual plasmid transfection of reporter and cognate Cas9 (Sp or Sa) in HEK293T cells (**Figure 39B-C**). As hoped for, we detected high and comparable knockout efficiencies of 161- and 218-fold for KS Sp and KS Sa, respectively. Moreover, the coupling of the gRNA cassette and target gene in the KS reporter resulted in a 10-fold higher knockout as compared to the standard triple transfection, where target and gRNA were provided by separate plasmids (compare H1 gRNA3 with KS Sp in **Figure 39B**).

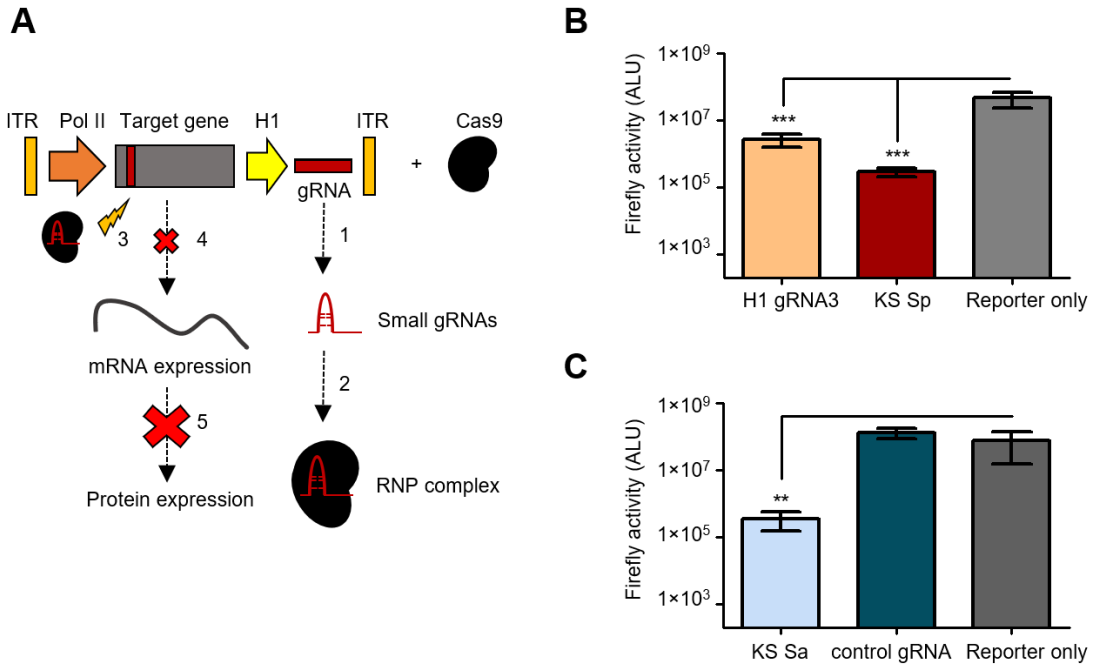


Figure 39. Functional validation of KS reporters *in vitro*.

(A) Schematic representation of the KS circuit**.** Shown is the KS vector, which is a rAAV genome that encodes a transgene of interest (in this work, FLuc) and a self-targeting gRNA. The gRNA is constitutively expressed from the vector (1), but an active ribonucleoprotein complex (RNP) is formed only in the presence of the Cas9 (2). This RNP then directs the Cas9 to the complementary target DNA sequence (shown as a red line within the “Target gene”), thereby promoting cleavage of the vector (3). Depending on the location of the binding site and the resolution of the cleavage by error-prone NHEJ, this can lead to debilitating perturbations in mRNA (4) or protein expression (5) and thus ultimately result in vector inactivation. **(B-C)** *In vitro* validation of the indicated KS reporters. HEK293T cells were either transfected with two plasmids, *i.e.*, the KS reporter and Cas9 (KS Sp and KS Sa conditions), or three plasmids: (i) encoding a gRNA against FLuc (or control), (ii) the psiCheck-2 reporter, and (iii) expressing the Cas9 (conditions H1 gRNA3 and control gRNA). In the "Reporter only" control, the KS reporter was co-transfected with stuffer DNA. All gRNA and reporter constructs were used at an amount of 25 ng/well.

The Cas9 expression plasmid (Sa or Sp) was transfected at 125 ng/well. FLuc and RLuc activities were measured in the cell lysates three days post-transfection (mean ± SD, n = 3). For statistical analysis, one-way ANOVA with Dunnett's post-test was used. Significance is indicated by: **, p<0.01, ***, p<0.001.

3.4.3 Potent reduction of Firefly luciferase expression from the rAAV KS vector *in vivo*

To test the activity and inactivation of our rAAV KS vectors in a physiologically more relevant setting, we performed a pilot *in vivo* experiment in mice, using the AAV8 capsid to direct transgene expression to the liver. FLuc activity was monitored in living animals after intraperitoneal injection of the D-Luciferin substrate. To be able to perform these experiments, we cooperated with the group of Michael Boutros (DKFZ, Heidelberg, Germany) who has access to Cre-dependent SpCas9-expressing C57BL/6 mice. While this setting where the Cas9 gene is already embedded in the host genome does not genuinely reflect the *in vivo* situation in a human patient, the rationale behind this first experiment was to test whether Cas9 can target the KS vector *in vivo* at all and, if so, at which efficiencies. To this end, mice were injected with either 2×10^{11} gc of the AAV KS Sp vector alone (Cre- mice) or with a combination of this vector with 1×10^{11} gc of scAAV-TTR-Cre vector (TTR [transthyretin] is a strong liver-specific promoter), to activate the endogenous Cas9 expression (denoted as Cre+ mice). FLuc expression was followed over time using the Xenogen IVIS100 imaging system (**Figure 40A**).

At different time points (two days as well as one and two weeks post-injection), we measured the total number of photons/second/cm^2/radian (*i.e.* total photon flux) emitted from each mouse (**Figure 40B-C**). Intriguingly, FLuc expression from the ssAAV KS vector was already observed two days post-injection (at 1×10^6 - 1×10^7 light units) and further increased after one week in both, Cre- and Cre+ animals. After two weeks, the signal continued to increase in the Cre- animals (**Figure 40B**), whereas it declined significantly in the Cre+ mice (**Figure 40C**).

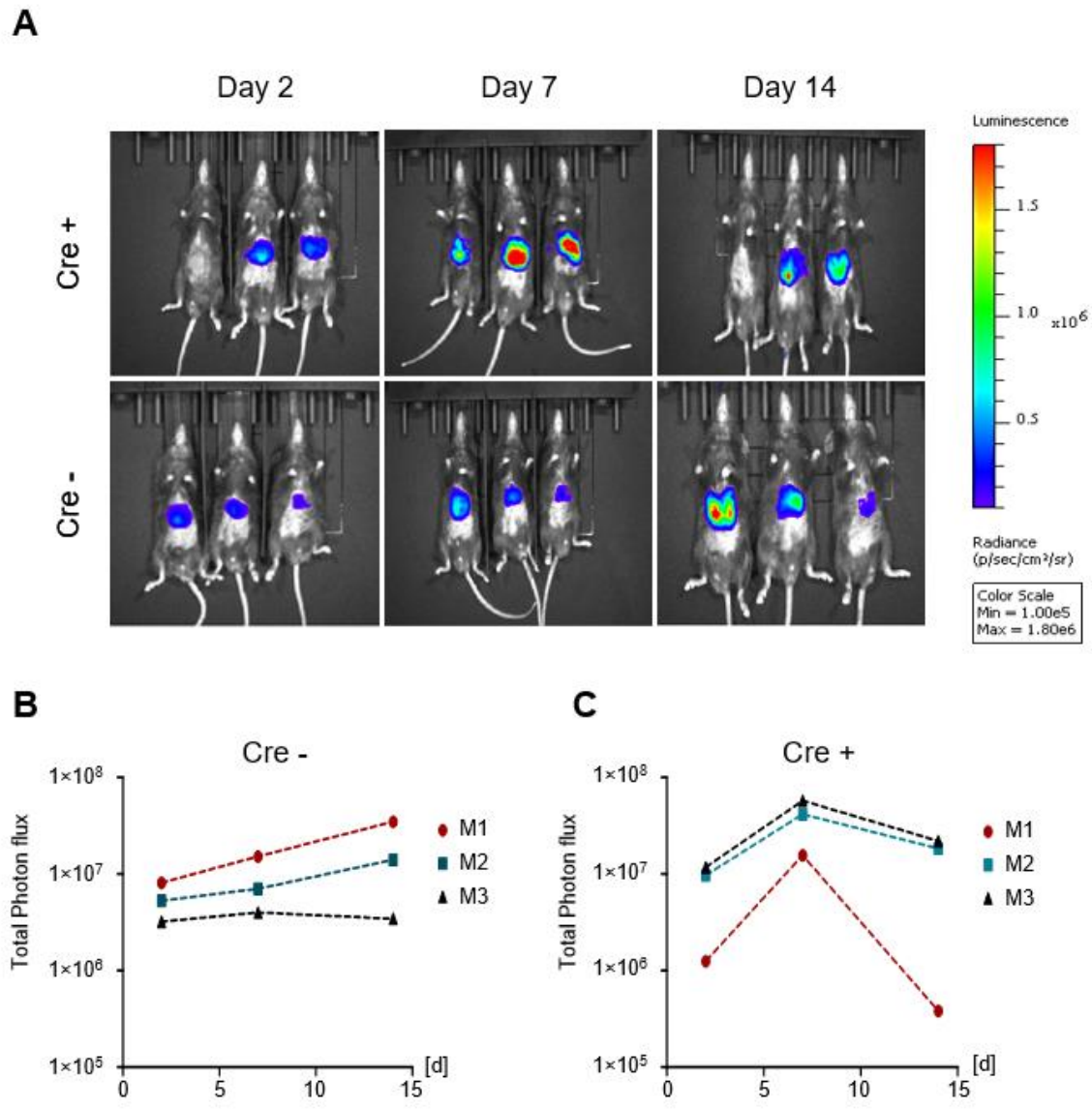


Figure 40. *In vivo* imaging of FLuc expression after tail vein injection of rAAV vectors.

(A) C57BL/6 mice encoding a Cre recombinase-dependent SpCas9 were co-injected with the KS Sp vector and a scAAV-TTR-Cre vector (upper panel; Cre+) or the KS Sp vector alone as control (lower panel; Cre-). Images were taken by Iris Augustin (Boutros lab, DKFZ, Heidelberg, Germany) at the indicated time points post-injection. **(B-C)** Total photon flux induced by FLuc activity at the depicted time points for the Cre- control animals in (B) and the Cre+ animals in (C). M1-3: mice 1-3, d = days.

This decrease in FLuc activity may well correspond to the activation of SpCas9 expression through rAAV vector-encoded Cre recombinase, followed by loading of the (Sp) gRNA3 into SpCas9 and cleavage of the FLuc-encoding gene. To clarify whether the sharper and earlier decrease in signal after week 2 in the Cre+ animals was a direct consequence of SpCas9 expression (as intended) or rather an effect of uneven vector doses, as previously reported for hFIX [305], the mice were sacrificed after week 5 and the livers isolated for DNA extraction and sequencing.

From the total liver DNA, the FLuc target region was PCR-amplified and cloned into an acceptor plasmid (**Figure 41A**). Next, 13 positive clones were sent for Sanger sequencing analysis. As hoped for, we detected mutations in eight clones at the expected gRNA cleavage site, which ranged from a few nucleotides to large deletions (see **Figure 41B** for a representative sequencing result from one mouse). By contrast, no mutations were detected in the Cre- control group (data not shown).

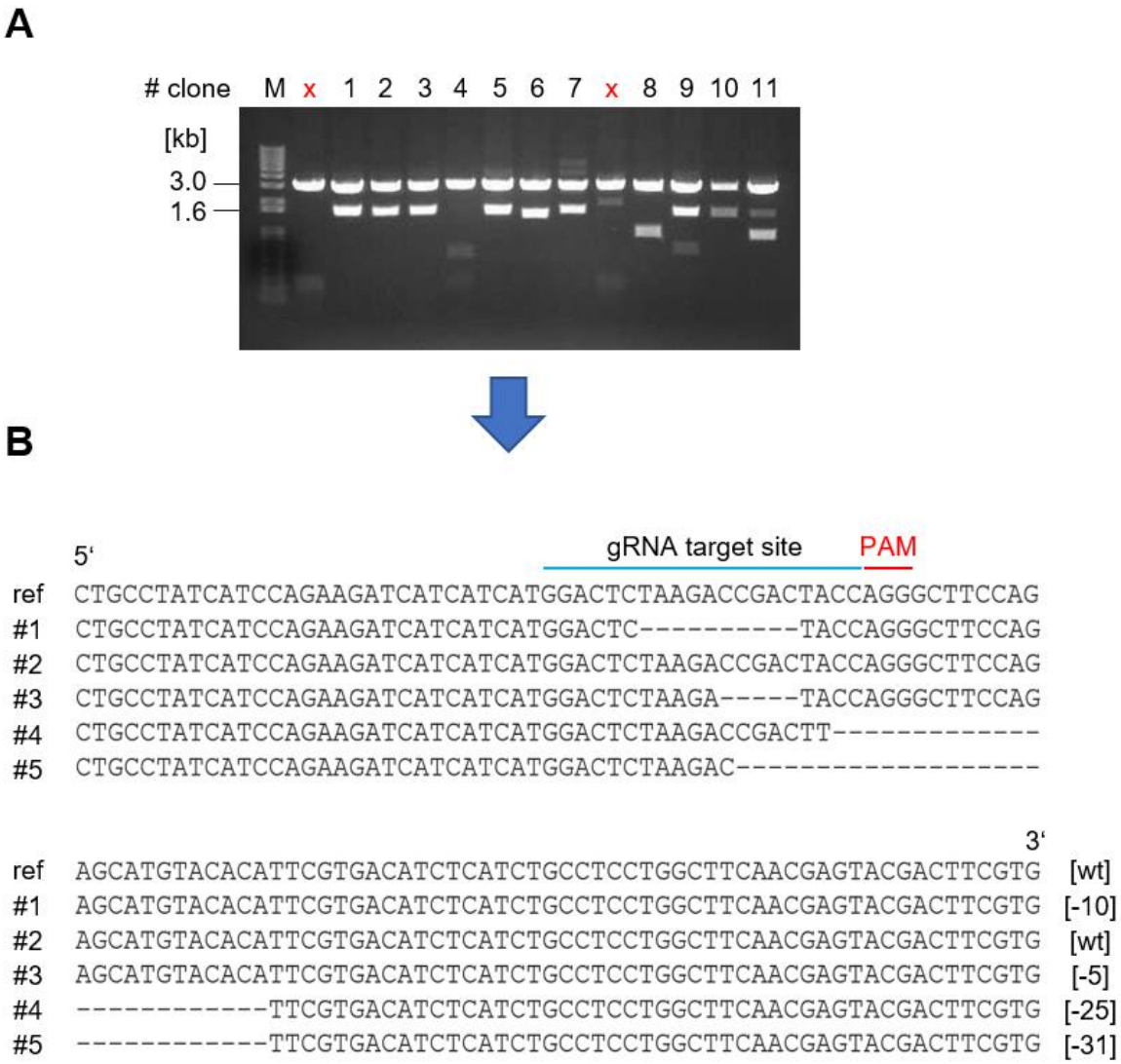


Figure 41. Analysis of Cas9-mediated cleavage of rAAV-KS Sp vector *in vivo*.

(A) The FLuc ORF was PCR-amplified from liver DNA and cloned into a pBS KS (+) acceptor plasmid. Several clones were picked from the plate and tested for the presence of an insert using BamHI / HindIII digest. The expected band is 1.6 kb in length and corresponds to the complete FLuc ORF. Cas9 cleavage can result in large deletions leading to smaller bands (*e.g.*, clone 6 and 8). M = 1 kb plus DNA ladder, x = negative clones devoid of insert. **(B)** Positive clones were sent for Sanger sequencing and aligned to the reference (ref) sequence. Numbers in brackets denote the number of deleted nucleotides at the target site. wt = wild-type reference sequence.

3.4.4 Construction and initial validation of rAAV SIN split SaCas9 vectors

One additional advantage of SaCas9 is the ability to split this ortholog into two halves, each of which is small enough to be packaged into, and delivered by, scAAV vectors. The latter are modifications of original ssAAV vectors that typically yield much faster and often also more robust transgene expression [226]. The split SaCas9 variant used in this work was originally constructed and provided by a current PhD student in our lab (Carolin Schmelas). The two parts of the split SaCas9 (Cas9 N and C) are expressed from two separate vectors and form a full-length protein upon co-expression and intein-mediated splicing.

To create a self-inactivating (SIN) variant of this vector design, we added a gRNA cassette targeting either the N- or C-terminal part of the split SaCas9 (denoted as N1-2 and C1-2, respectively) into the vector encoding the split Cas9 C-terminal part (**Figure 42A**). The resulting constructs were validated by transfection into HEK293T cells along with the KS Sa vector. In this experiment we measured: (i) FLuc knockout efficiencies (**Figure 42B-C**) using the SIN Cas9 system compared to the controls (split Cas9 or full-length Cas9 vector; both without inactivating gRNA) and (ii) efficiency of Cas9 knockout by the anti-Cas9 gRNAs (**Figure 42D**)**.**

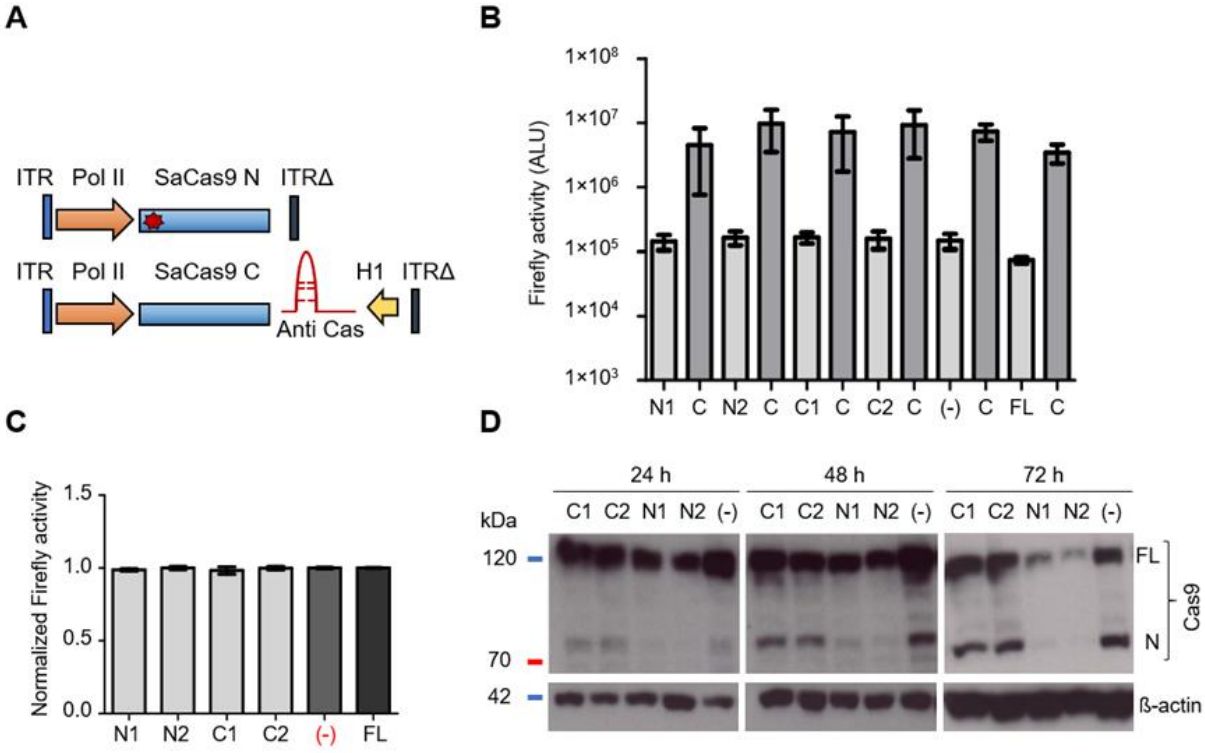


Figure 42. Characterization of SIN split Cas9 vectors.

(A) Schematic diagram of a SIN split SaCas9 system: The SaCas9 N and C parts are expressed from two separate vectors and form a full-length protein upon co-expression and intein-mediated splicing. A gRNA directed towards the Cas9 itself (Anti-Cas) was cloned under the H1 promoter to form the SIN system, while a negative control expressed no gRNA. **(B-C)** SaCas9 cleavage efficiencies with four SIN gRNAs (N1-2, C1-2) as evaluated by luciferase assay. HEK293T cells were seeded in 96-well plates at a density of 3×10^4 cells/well. In each condition, three plasmids were co-transfected: (i) the N-Cas part

(50 ng/well), (ii) the C-Cas part (50 ng/well) with an anti-Cas gRNA (N1-2, C1-2) or without gRNA as control (-), and (iii) the KS Sa luciferase reporter (2 ng/well). C = control for each luciferase knockout condition using a scrambled version of the anti-Fluc gRNA. FL = Full-length SaCas9 control (50 ng/well) expressed from the same promoter as the split SaCas9 (here, mi[nimal]CMV). Firefly luciferase activity was measured thee days post-transfection and plotted on the y-axis as ALU (mean ± range, n=2). **(C)** Alternative illustration of the data presented in (B). FLuc expression from the (-) control condition was set to 1 to calculate the fold difference in SaCas9 activity. **(D)** SaCas9 expression levels as determined by Western blotting using an antibody raised against the SaCas9 N terminus. HEK293T cells were cultured in 24-well plates and transfected with the KS Sa reporter (8 ng/well) and the two halves of the split Cas9 (400 ng each per well) with the indicated SIN gRNAs or without gRNA as control (-). Cells were harvested at 24, 48 and 72 h post-transfection for analysis (n = 2). N = N-terminal split Cas9 part.

Out of the four gRNAs tested, the two designed against the Cas9 N-terminus (N1 and N2) resulted in a clear decrease in full-length and N-terminal part Cas9 expression, as revealed by Western blot analysis at different time points (**Figure 42D**). Notably, the C-terminal part was not detected in this assay because the used anti-Cas9 antibody was raised against the N-terminus of SaCas9. Moreover, all of the tested constructs mediated up to 100-fold knockout of FLuc expression (**Figure 42B**). This is especially intriguing for the SIN split Cas9 vectors (N1 and N2) that showed in the same robust FLuc knockout when compared to the controls, despite their efficient self-inactivation of Cas9 expression (**Figure 42C-D**).

3.4.5 Towards a customizable system for concurrent target and SaCas9 inactivation using rAAV SIN split Cas9 vectors

One remaining disadvantage of the above-mentioned approach is the need to modify the rAAV target vectors to include a gRNA cassette, which complicates the application of this strategy to already available vectors. One solution would be to supply the anti-target gRNA on an additional vector, which results in a three- instead of a two-vector system. Moreover, in the setting where target and gRNA are encoded on the same vector, the anti-Cas9 gRNA can typically only be expressed from the small and weak H1 promoter because of the size limitation of scAAV vectors. Thus, in the last part of this work, we aimed to generate a tool box of customizable and thus more user-friendly ssAAV SIN CRISPR vectors. These vectors were designed to include both gRNAs (anti-target and anti-Cas), hence allowing simultaneous targeting of the transgene- and Cas9-encoding vectors in the same cell. In addition, we used different combinations of two RNA Pol III promoters to drive the expression of the gRNAs (U6 for stronger or H1 for weaker expression). The idea behind this was to become able to modify and optimize the kinetics and efficiencies of both cleavage reactions, targeting and thus inactivating the transgene and the Cas9 itself.

Because of the size of the U6 promoter as well as of the N-terminal SaCas9 part, the new constructs only fit into a ssAAV vector. Hence, their activity was compared to that of the existing cassettes that are encoded in the scAAV context (**Figure 43**).

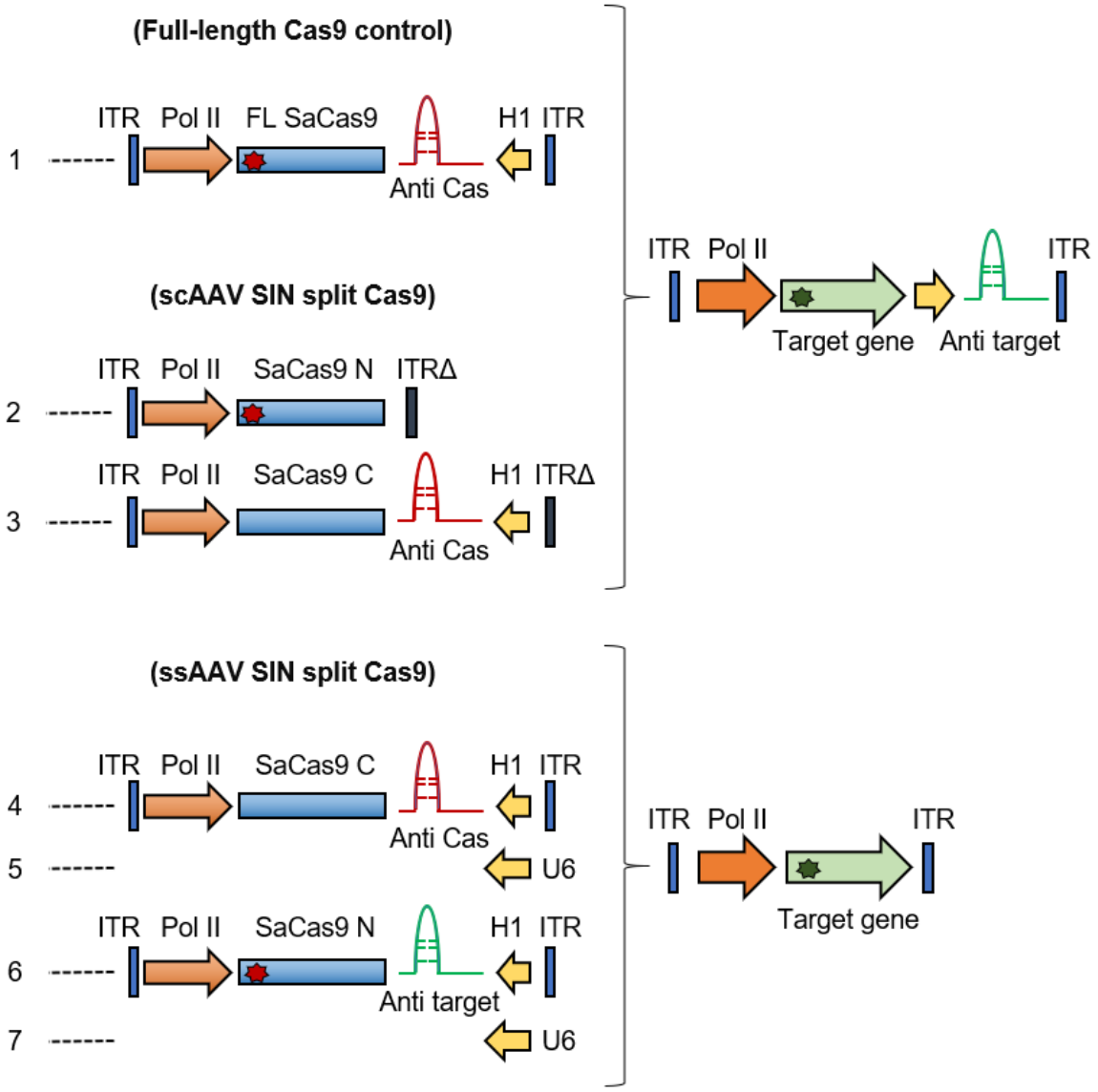


Figure 43. Schematic representation of all rAAV SIN Cas9 vectors used in this work.

Construct #1 is a control composed of a full-length (FL) SaCas9 and a gRNA cassette targeting the SaCas9 itself. Constructs #2 and #3 express the two parts of the split SaCas9 (N and C) from scAAV vectors. A gRNA cassette against the SaCas9 N-terminal part is embedded into the vector encoding the C-terminal part. Constructs 1 to 3 are always used with the KS Sa luciferase reporter. Constructs 4-7 encode the two parts of the split SaCas9 (N and C) from ssAAV vectors. gRNA cassettes against the SaCas9 itself (anti-Cas) or the target (anti-target) are located on the same vectors and are expressed from the H1 or U6 promoters. The ssAAV split SaCas9 constructs are always used with a standard luciferase reporter ssAAV-psiCheck-2. The anti-Cas9 gRNA used is the gRNA N1 (see **Figure 42**). ITRΔ = mutated AAV4 ITR that allows the packaging of rAAV vector genomes as scAAVs [241]. Pol II = RNA polymerase II promoter (here, miCMV). Stars denote the anti-Cas or anti-target gRNA binding sites within SaCas9 and the target gene, respectively.

We first validated the functionality of all the constructs shown in **Figure 43** by transfection into HEK293T cells (**Figure 44,** upper graph). Two conditions served as controls: (i) the ssAAV vector encoding the full-length SaCas9 and a gRNA cassette targeting the SaCas9 itself (construct #1 in **Figure 43**), and (ii) the SIN split SaCas9 in a scAAV context (constructs #2

and #3 in **Figure 43**). The anti-target gRNA was either embedded into the new ssAAV SIN CRISPR vectors (constructs 4-7 in **Figure 43**) or was expressed from the KS Sa reporter in the control conditions.

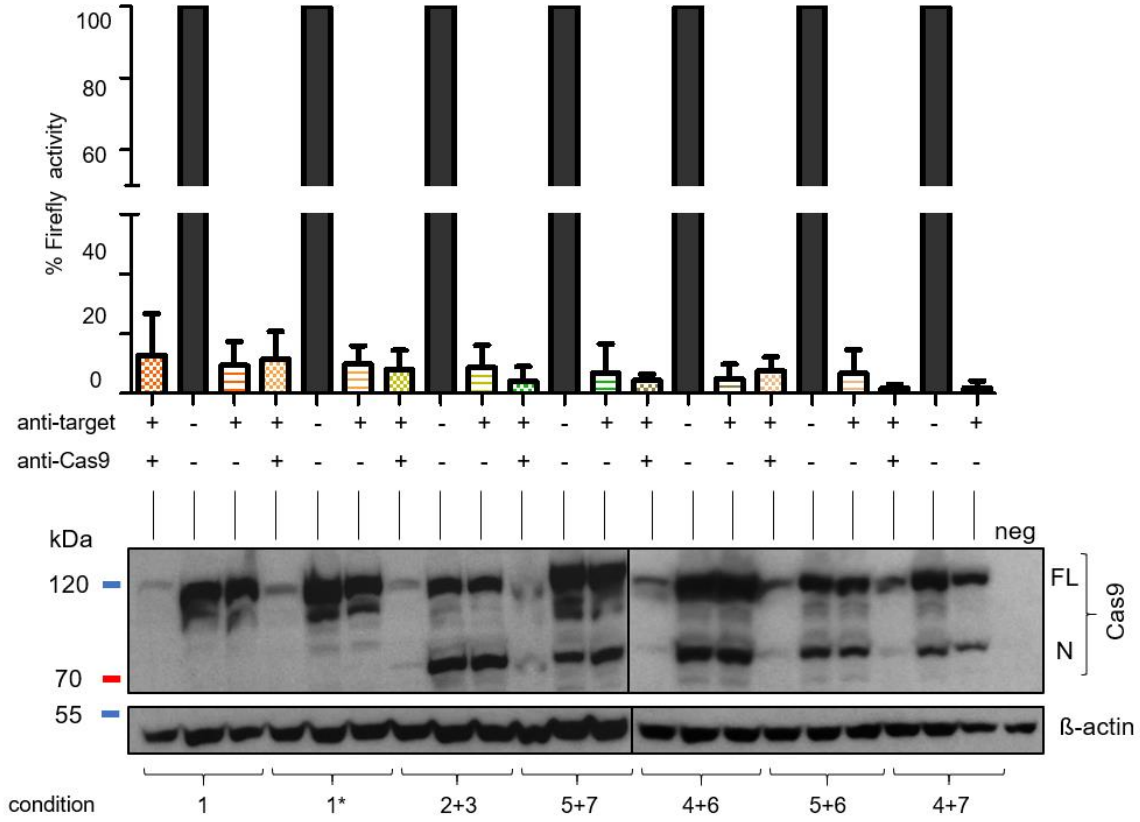


Figure 44. *In vitro* validation of ssAAV SIN CRISPR constructs.

Upper panel: HEK293T cells grown in 96-well plates (seeded at 3×10^4 cells/well) were transfected with the indicated combinations of plasmids (conditions 1 to 4+7). In all conditions, 10 ng/well reporter and 50 ng/well split or full-length (FL) SaCas9 were used. For 1, 1* and 2+3, the KS Sa reporter was co-transfected. 1* = condition 1 with twice the amount of FL SaCas9 plasmid DNA (*i.e.* 100 ng). In all other conditions, a luciferase reporter without an extra gRNA cassette was co-transfected (psiCheck-2). The constructs included either the indicated gRNAs (+) or scrambled (scr) controls (-). The remaining luciferase activity is plotted as percentage of the control (-/- condition) with two scrRNAs derived from anti-FLuc and anti-Cas9 (mean + SD, n = 3). Lower panel: Western blot analysis of the indicated conditions. HEK293T cells were cultured in 24-well plates (seeded at 1.25×10^5 cells/well). In all conditions, 40 ng/well reporter and 400 ng/well of each split SaCas9 half or FL SaCas9 were transfected. 1* = condition 1 with twice the amount of FL SaCas9 plasmid DNA (*i.e.* 800 ng). The Western blot shows the expression levels of the SaCas9 protein. neg = untransfected cells; N = N-terminal split SaCas9 part (n = 3). Luciferase assay and Western blot analysis were performed three days post-transfection.

As hoped for, all the constructs were functional and expressed sufficient amounts of SaCas9 that led to comparable knockout of FLuc activity, irrespective of the reporter used (standard or KS reporter). Also, inactivation of SaCas9 expression was detected in all cases as shown by Western blot analysis (**Figure 44**, lower panel). Importantly, as previously observed in the initial validation experiments (**Figure 42**), FLuc knockout efficiency in the

presence of an anti-Cas9 gRNA was indistinguishable from the control condition (with a non-targeting scrRNA against SaCas9).

Having proved the functionality of all our plasmid constructs, we moved to the system of interest in this work, which is the rAAV vector context. Thus, all the constructs shown in **Figure 43** (including the KS Sa reporter and the dual-luciferase cassette from the standard psiCheck-2 plasmid) were packaged into AAV2 and tested in HEK293T cells via transduction (**Figure 45**). Importantly, all the SIN split SaCas9 vectors were functional and provided robust on-target activity (FLuc knockout; **Figure 45**, upper graph). Likewise, all conditions (except for 4+7) resulted in knockout of SaCas9 expression (**Figure 45**, Western blot analysis at the bottom).

Overall, in the ssAAV SIN system, the best FLuc knockout was achieved when the U6 promoter was used to drive anti-target gRNA expression (conditions 4+7 and 5+7). In contrast, the best SaCas9 knockout was achieved in conditions 4+6 and 5+6, where the anti-target gRNA was expressed from the H1 promoter and the anti-Cas gRNA from the H1 or U6 promoter. This is followed by condition 5+7 where both gRNAs were expressed from strong U6 promoters. Interestingly, this last condition was comparable in its efficiency to the SIN split SaCas9 in the scAAV context, in which both gRNAs were driven by the H1 promoter, which might reflect higher expression levels of gRNAs from scAAV vectors

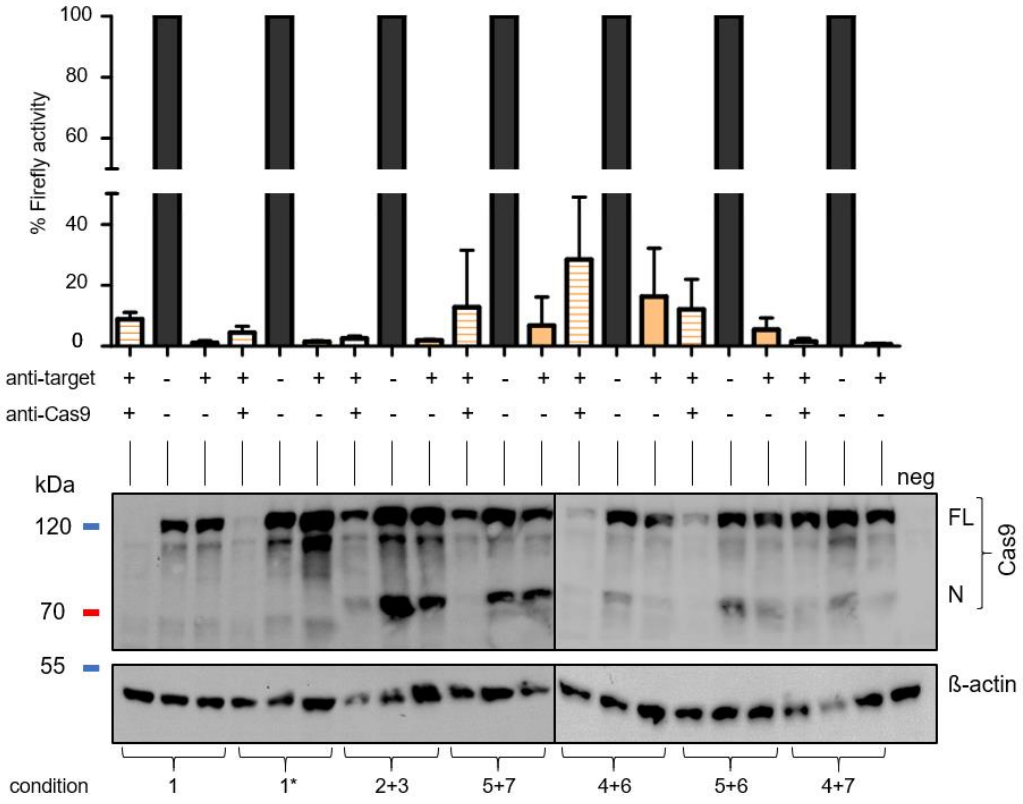


Figure 45. Functional analysis of SIN split Cas9 rAAV constructs after transduction.

Upper panel: ss and scAAV constructs shown in **Figure 43** were packaged into AAV2 and 3×10^4 HEK293T cells (96-well format) were transduced with the indicated combinations at a MOI of 1×10^5 per SaCas9 construct and 1×10^4 for the luciferase reporters (KS Sa and standard psiCheck-2 dual-

luciferase cassette). For 1, 1* and 2+3, the KS Sa reporter was co-transduced. 1* = condition 1 with twice the MOI of the SaCas9 vector. In all other conditions, a luciferase reporter without an extra gRNA cassette was co-transduced (standard psiCheck-2). Moreover, the constructs included either the indicated gRNAs (+) or scrambled (scr) control (-). The remaining luciferase activity (mean + SD, n = 3) is plotted as percentage of the control with two scrRNAs (anti-Cas and anti-FLuc; condition -/-). Lower panel: Representative Western blot analysis showing expression levels of the SaCas9 protein for the indicated conditions (n = 3). HEK293T cells were cultured in 24-well plates (1.25×10^5 cells/well) and transduced at the same MOIs mentioned in the upper panel. FL = full-length SaCas9; neg = untransduced cells; N = N-terminal split SaCas9 part. Luciferase assay and Western blot analysis were performed five days post-transduction.

4 DISCUSSION

The field of human gene and cell therapy is currently at its all-time high, exemplified by the numerous breakthroughs that we have witnessed over the last few years, from (near) eradication of certain cancers [306] to restoration of sight [307] and cure of patients from hemophilia [142], and culminating in the first three market authorizations for gene therapy products based on recombinant AAVs (Glybera, Luxturna and, most recently, Zolgensma). As another striking evidence, the annual 2019 meeting of the American Society of Gene and Cell Therapy (ASGCT), the world's largest and most influential respective community, hit a new record with over 2000 additional attendees from academia and industry as compared to last year [308, 309]. Triggered by these successes and the rapid expansion of the field, there is now an unprecedented momentum and an urgent desire to further broaden the spectrum of diseases for which these technologies can provide therapeutic benefit. This, in turn, demands new, more potent, less immunogenic and more specific viral and non-viral vectors that concurrently provide maximum control over patient safety, which altogether fueled the concepts and studies presented here.

4.1 Development and characterization of novel primate bocaviral vectors

In the first part of this work, we expanded on a previously described pseudotyping approach [10] to study the utility of four, so-far uncharacterized primate BoV isolates as viral vectors. In essence, we successfully packaged rAAV reporter constructs into the BoV capsids to generate five pseudotyped rAAV/BoV vectors, namely, HBoV1-4 and GBoV. These rAAV/BoV vectors were then used to transduce different types of cells, from suspension cultures to monolayers and organoids, which surprisingly revealed a broad tropism *in vitro*, akin to what has been observed for rAAVs. This was not only intriguing from a biological standpoint, but also demonstrates the utility of the pseudotyped vectors for gene transfer into multiple different targets. Moreover, we present, for the first time, the successful application of DNA family shuffling (DFS) to a parvovirus other than AAV. The chimeric BoV *cap* library generated in this work was highly diverse and produced packaging-competent viral progeny. Cycling of this library in pHAE revealed interesting sequence patterns and enrichment of domains that might play a role in capsid assembly and function.

Altogether, by shedding light on the previously enigmatic tropism of primate BoVs, we not only advanced our understanding of BoV biology but also extended the range of application of BoV vectors. Moreover, we demonstrated the applicability of DFS to the BoV *cap* sequences

and envision a great value of this technique in future investigations of structure-function relationships, trafficking and distribution of antigenic domains.

4.1.1 rAAV2/HBoV1 vectors can be efficiently produced and package intact oversized rAAV genomes

rAAVs are among the most commonly used vectors for gene transfer, but one limitation is the small packaging capacity of ~4.7 kb, which restricts their application to smaller transgene cassettes [255]. Importantly, previous work has shown that rAAVs can package ssAAV vectors larger than their wt genome size [254-256]. The optimal length was, however, estimated to be between 4.1 and 4.9 kb because larger genomes were accompanied with a sharp reduction in packaging and/or transduction efficiency [254, 255]. In concordance with this, we could efficiently package genomes up to 5.1 kb into the AAV2 capsid, whereas larger genomes (5.5-6.1 kb) only resulted in a low-molecular-weight DNA smear in gels. In contrast, we show that the optimal genome length for packaging into the HBoV1 capsid is ~5.8 kb. Genomes of 6.1 kb were also encapsidated without truncations, which was evidenced by a distinct band in gels at the expected height. However, the intensity and hence packaging efficiency was lower as compared to smaller genomes. This is in line with and extends previous observations that the HBoV1 capsid can package a rAAV-*CFTR* cassette of ~5.4 kb [10].

This gain of ~800 bp in size is the first notable advantage of pseudotyped rAAV/HBoV1 vectors that we illustrated here. We harnessed this extra space to package complete CRISPR cassettes including the SpCas9 and a gRNA against the *CFTR* locus, both expressed from strong ubiquitous promoters. Smaller Cas9 proteins, such as the SaCas9 ortholog used in this work, would additionally allow the embedding of a DNA template of ~1 kb for homology-directed repair. This ability to readily package all-in-one CRISPR cassettes is remarkable and would not be possible using standard AAV capsids because of their size limitation [209].

Irrespective of size, a drawback of single-stranded AAV genomes is their need to be converted inside the nucleus into transcriptionally active double-stranded DNA. This process is rate-limiting and dependent on the cell type [310, 311]. Consequently, scAAV vectors have been developed that bypass the cell cycle-dependent conversion step and thus result in faster and more potent transgene expression [226]. However, the size of the packageable genomes is reduced by ~50% (~2.4 kb), which again limits the coding capacity. Thus, we asked whether the HBoV1 capsid can also package oversized scAAV genomes and, if so, to which extent. Our results show that up to 2.8 kb and 3.2 kb can be efficiently packaged into AAV2 and HBoV1 capsids, respectively. These extra 400 bp correspond to half the length gained for the ssAAV vector DNA, which supports and strengthens our above-mentioned observations. Curiously, Wu *et al*. reported that scAAV genomes up to 3.3 kb can be packaged into the AAV2 capsid [251]. Yet, in contrast to the smaller genomes also tested in their study, the 3.3 kb vector resulted in the least intense band, which might reflect less efficient packaging. It was also speculated

that the GC content and secondary structure of the vector genome play a role in the efficiency of the encapsidation process, which can further explain differences between laboratories [251, 254]. Another important observation in the above-mentioned study was that lowering Rep expression not only increases AAV packaging efficiency (as previously shown [134, 135]) but also favors the encapsidation of intact sc over ss molecules. This is congruent with our own observations of >95% intact scAAV genomes. Notably, here we used two different helpers with reduced Rep protein expression: pDGΔVP [137] and the conventional AAV helper (pWHC2) [136] for HBoV1 and AAV2 vector production, respectively (for more details, see section 1.2.5). Further important to note is that the sc- or ss-vector plasmids were kept constant for both, AAV and BoV vector production, ensuring that the observed increase in packaging capacity for HBoV1 was only due to the viral capsid.

In parallel to these experiments, we established a new 3-plasmid production system and compared it to the originally reported 4-plasmid transfection protocol. Both methods produced comparable rAAV2/HBoV1 viral titers of ~5×10^{10} gc/mL, congruent with previous observations of a reduced packaging efficiency in the pseudotyped vector system (5-10% of cognate AAV vectors) [10, 261]. Important to mention in this context is that Yan *et al.* independently reported an optimized 3-plasmid production system in 2018 [138]. In this study, the authors showed that rAAV2/HBoV1 vectors can be produced in the complete absence of BoV non-structural proteins to high viral titers (>4×10^{12} gc/mL). It will be interesting to test whether a combination of the two systems reported by Yan *et al*. or us, in particular a reduction of Rep expression in the NS-free system, will even further boost viral yields.

One interesting observation in our study of vector genome replication in HEK293T producer cells was that the rAAV2 control vectors produced in parallel exhibited the expected bias [312], *i.e.*, an excess of monomeric over dimeric replication forms (**Figure 9B-D,** lower gels). This was shifted in the rAAV2/HBoV1 system to a reduced accumulation of monomeric DNA replication forms for all the tested ss and scAAV constructs, which may reflect a less efficient resolution of replication intermediates (**Figure 10A-B,** lower gels). This has not been observed before using the adenoviral helper pAd4.1 [10], making it an important task for future work to clarify whether the effect results from an inherent property of pDGΔVP. Also notable in this context is that the larger scAAV genomes tested in this work (≥2.5 kb) seemed to result in more monomeric replication forms and fewer concatemers.

The less efficient resolution of concatemers has been previously reported for smaller ssAAV genomes and linked to fast preinitiation of replication without resolution of the terminal ends [312]. Thus, the effect was more prominent for small ssAAV genomes between 0.7 and 2.4 kb (corresponding to 15-50% of the wt AAV genome size) [312]. Importantly, these differences in monomeric:dimeric ratio neither affected the viral titers nor the functionality of capsids carrying

the smaller scAAV genomes, as determined by qPCR analysis and transduction of pHAE, respectively (data not shown).

For the future, it remains crucial to determine the repercussions that the oversized ss- and scAAV vector genomes have on HBoV1 infectivity. Interestingly, the above-mentioned study by Grieger *et al.* [254] revealed that DNA encapsidation not always correlates with efficient transduction. The reduced infectivity of particles encapsidating larger rAAV genomes was found to be related to preferential proteasomal targeting and degradation. This created a complication in our study that prevented us from directly addressing the effect of genome size on infectivity, since our current rAAV2/BoV generation strictly depends on proteasome inhibition, irrespective of the encapsidated genomes (reported for HBoV1[10] and also seen here for all tested BoV serotypes; data not shown). Thus, a solid conclusion on the infectivity of BoV particles carrying oversized rAAV genomes has to await the next generation of BoV vectors that is no longer or less dependent on proteasome inhibitors (see also chapter 4.1.4 below).

4.1.2 BoV production is supported by different helpers and is most efficient and unbiased using the rAAV/BoV pseudotyping approach

Over the last five decades, many BoV serotypes and variants have been discovered in non-primates and linked to different diseases and tissue distribution. Only in the last 15 years, however, primate BoVs were also detected in humans (HBoV1-4), Gorilla (GBoV) and Chimpanzee (CPBoV; GenBank: KT223502) with an unknown pathophysiological relevance for the host. This abundance of known BoV sequences fueled our second aim, which was to characterize and harness different variants for gene transfer. Here, we were inspired by the modularity of the AAV vector system in which the ITRs, *cap*, *aap* and *rep* ORFs are largely interchangeable between the serotypes (except for AAV5, see section 1.2.5.1). This high genome flexibility is intriguing and facilitated the development of multiple hybrid AAV vector systems [117-119, 313, 314]. For example, the most broadly used AAV helper plasmids nowadays co-express AAV2 *rep* together with a *cap* ORF derived from a different serotype or a synthetic *cap* variant.

Akin to the AAV Rep proteins, primate BoVs share a high sequence identity of >70% in the non-structural ORFs. Thus, we asked whether the BoV genomes offer a similar plasticity as AAV. To this end, we constructed hybrid BoV expression plasmids composed of the *ns*/*np1* ORF of HBoV1 and the *cap* ORF of a different BoV variant. Moreover, the genomes were either flanked by HBoV1 TRs to create wtBoV plasmids or devoid of TRs, yielding BoV helper plasmids. Indeed, from all the various plasmids, we were able to detect correct expression of the three viral Cap proteins (VP1-3) at the expected 1:1:10 stoichiometry [70, 302]. This reflects correct processing and splicing of the pre-mRNA and a retained regulation of expression ratios. This, in turn, indicates a similar functional organization of the BoV genomes, especially concerning the dependency of correct VP expression on upper genomic regulatory sequences

[81]. The presence of not only free VPs but also correctly assembled capsids was confirmed by electron microscopy analysis of the viral stocks. Interestingly, we detected an excess of empty over full particles in wtHBoV2, 3 and 4, in contrast to HBoV1 (not shown) and GBoV. A qPCR analysis of the viral stocks supported this observation in particular for HBoV2 and 4, where we detected ~10-fold lower titers. This implies that the NS1 proteins and/or TRs of HBoV1 might conflict with efficient packaging into HBoV2-4 capsids and suggests routes for future optimization, including *e.g.* the cloning and characterization of alternative BoV TR sequences akin to prior work in the AAV field [305]. In fact, the terminal repeat sequences play a crucial role in the packaging of viral genomes [315]. Alas, despite the wealth of information collected in the last years about BoVs, most TR sequences remain elusive. An alignment of the available partial sequences revealed a generally high sequence conservation, which is probably required for the formation of the complex secondary structures of the termini [289]. Importantly, the structure of the HBoV1 3' TR is slightly different from the ones in HBoV2-3 regarding the length of the stem loops [289], which further supports the idea to clone and study different BoV TR sequences.

Finally, the superior packaging of wt genomes into GBoV capsids is intriguing and underlines its close phylogenetic relationship to HBoV1. Interesting in this context is that the basket below the five-fold channel, which is unique to BoV and has been proposed as a portal for genome packaging [302], is highly conserved yet HBoV1 and GBoV carry a substitution of valine with isoleucine. It will be interesting to test in follow-up work whether the mutation of this amino acid can rescue the packaging defect in wtHBoV2-4.

In striking contrast to our findings with the replication-competent constructs, all BoV helper plasmids were readily compatible with cross-genera pseudotyping, akin to what was reported for HBoV1 [10]. Proof is that all scAAV reporters used in this work were packaged to comparable efficiencies into the different BoV capsids and resulted in high vector titers of $>10^{12}$ gc/mL.

Noteworthy, another chimeric BoV helper plasmid was constructed in this work by flanking the HBoV1 genome with AAV2 ITRs. This helper was tested by Stefan Holderbach, who showed that packaging of rAAV2 genomes in the presence of only BoV NS proteins is impossible. The packaging defect was, however, rescued when the adenoviral helper and the Rep proteins were supplied *in trans*, implying that the Rep but not the NS proteins are the key players in this chimeric system (S. Holderbach, BSc. thesis, 2017). This observation was supported in the above-mentioned study of Yan *et al.* in 2018 [138] and served as a basis for the construction of the chimeric BoV libraries, which are discussed below. This exclusive dependency on the AAV Rep proteins and the comparable packaging efficiencies in the rAAV2/BoV cross-genera packaging system are remarkable. They not only reflect a greater flexibility of the AAV Rep proteins than previously anticipated but also imply a different mode

of Rep-Cap interaction that probably involves structurally conserved motifs on the BoV capsids. A similar observation has been made in the chimeric B19/rAAV system, where Rep proteins were strictly required for packaging of rAAV2-based genomes into the B19 capsid [126, 127].

4.1.3 BoV vectors show a broad tropism *in vitro* and a differential reactivity to IVIg

After the successful construction of BoV helper constructs and the production of DNA-containing virions, we tested the functionality of the new rAAV/BoV vectors. Therefore, we pseudotyped two reporter constructs, namely, scAAV-YFP or scAAV-GLuc, with each BoV capsid. The first allowed for a direct visualization while the second permitted to monitor transgene expression over time. Based on our own phylogenetic analysis and the sporadic detection of HBoV2 in the lung [75], we first tested our variants in pHAE. This already confirmed the functionality of two vectors based on HBoV4 and GBoV. We moreover tested a panel of primary cells and cell lines, grown in monolayers or complex organoid structures, for their susceptibility to BoV infection. Interestingly, BoVs displayed a preferential transduction of primary cells and organoids as compared to cell lines. While the underlying reasons remain unknown, this preference has also been observed before for HBoV1 in its recombinant or wild-type form [10].

We found that many of the tested primary cells supported transduction with the pseudotyped vectors, including primary human T cells, hepatocytes and skeletal muscle cells. The equal and high transduction efficiencies in primary hepatocytes and skeletal muscle cells are notable for three reasons. First (i), they provided the first evidence for the integrity and functionality of all tested BoV variants, especially HBoV2 and HBoV3, which showed inferior transduction efficiencies in many other cell types. In addition, (ii) these data suggested that all BoV vectors can be considered in the future for gene transfer into the liver or skeletal muscle cells, and that, *vice versa,* (iii) the liver and skeletal muscle should be considered as off-targets in case of another on-target. This particularly applies to HBoV1 vectors, which are currently developed for *in vivo* gene transfer to the lung [261]. Also intriguing is the transduction profile achieved in primary human T cells. In contrast to pHAE, HBoV1 was inferior to HBoV4 and GBoV, which displayed high transduction efficiencies nearly reaching rAAV2 levels. This is interesting from the perspective of natural BoV biology because these target cells were rarely associated with HBoV infection [316, 317].

Finally, we detected different BoV cell-type specificities within a complex co-culture of pHAE, where HBoV4 was less efficient at transducing ciliated cells, as compared to HBoV1 and GBoV. This implies that HBoV4 is a more suitable serotype for gene transfer into non-ciliated cells, such as basal and club cells.

Collectively, our data demonstrate a broad *in vitro* tropism and unique expression profiles for all tested BoV serotypes. Combined with the larger packaging capacity, this opens up a wealth of intriguing future applications, such as the delivery of all-in-one CRISPR cassettes for *ex vivo* gene editing of (CAR-)T cells or *in vivo* targeting of pathogens that infect specific cells in the body, such as hepatitis viruses and liver cells, or HIV and T cells.

Further increasing the attractiveness of our new vectors is our notion of a lower reactivity of HBoV4 and GBoV with pooled human sera (IVIg), as compared to HBoV1. Our findings are compatible with the fact that GBoV is a non-human BoV variant, and they are also in line with data that HBoV4 is the least seroprevalent variant in the human population (~2%) [279]. NAbs can recognize and inactivate a viral vector before reaching its intended target organ or cells. Accordingly, variants with lower immunogenicity and lower prevalence in the human population are preferred for clinical gene transfer. It should thus be very interesting to validate and extend these findings by testing different concentrations of IVIg or individual (patient) sera. This would allow for a more comprehensive determination of virus-neutralizing antibody titers in the population, which will in turn provide important insights into the clinical applicability of bocaviral vectors [280].

4.1.4 rAAV2/BoV vectors are dependent on proteasome inhibition

Yan and colleagues reported in 2013 [10] that transduction with rAAV2/HBoV1 in pHAE can be enhanced up to 1,000-fold by the application of two proteasome inhibitors (PIs), namely, LLnL and doxorubicin. This is in line with our own observation that PIs are required for efficient transduction not only with HBoV1-based vectors but with all other rAAV/BoV vectors tested in this work (data not shown).

Importantly, the ubiquitin/proteasome pathway has previously been shown to be critical in the life cycle of different parvoviruses including MVMp [318] and AAV [257, 319]. Modulation of this pathway by mutating surface-exposed tyrosine residues [258] or by the application of pharmacological agents has informed innovative strategies for enhancing rAAV transduction. For example, Douar *et al*. [319] showed a 50-fold enhancement of rAAV2 transduction following treatment with the PI MG-132 in HeLa, HEK293 and HepG2 cells. This is consistent with our data showing a potent enhancement of rAAV and rAAV/BoV transduction in all cell types tested. In pHAE, it has been shown that the application of PIs enhances rAAV transduction from the apical, but not from the basolateral side [274]. Duan *et al.* [320] demonstrated that this synergistic effect also occurs *in vivo* in the mouse lung after co-administration of rAAV2 and the PI Z-LLL. Transduction efficiency in this setting was enhanced from non-detectable to ~10% of bronchial epithelial cells. Finally, Yan and colleagues [261] detected efficient transgene expression in the ferret lung after treatment with rAAV2/HBoV1 and the PI doxorubicin. Notably, in both *in vivo* studies, no local or systemic toxicity was detected after the application of PIs, implying clinical relevance of this strategy.

Another example of a commonly used PI is bortezomib, a US Food and Drug Administration-approved PI for treatment of multiple myeloma. This agent has been used in combination with oversized rAAVs for delivery of a 5.6 kb factor VIII cassette in hemophilia mouse and dog models [321]. The study showed a striking adjuvant effect of the combination therapy, which resulted in an enhancement of expression of up to 600% and persistent correction of the disease phenotype. Importantly, and supported by the observations in this work, only a single dose of PI was needed to augment transduction and no further application was required.

In conclusion, the dependency on PIs would not essentially prevent the use of our reported rAAV/BoV vectors *in vivo*. Nevertheless, in view of the complexity of combination therapies in terms of costs and safety considerations, one aim in this work was to generate vectors that can escape proteasomal degradation and thus transduce independently of PIs. Akin to the strategy applied in the rAAV field [258, 322], we mutated six tyrosine residues in the VP2 Cap protein of HBoV1 to phenylalanines, two of which were located on the capsid surface. While none of the HBoV1 mutants mediated an escape in the absence of PIs, two mutants (Y355 and Y466; VP2 numbering) were 10- to 6-fold less efficient at transducing primary HAE. Interestingly, residue Y355, which resulted in the strongest phenotype, lies on the capsid surface and is highly conserved among primate BoVs. Moreover, it directly flanks VR-VI and thus may contribute to loop formation and stability. By contrast, residue Y466 lies inside the capsid in close proximity to VR-VIIIB, which is a highly variable region linked to particle assembly, genome packaging and intracellular trafficking in AAVs [302, 323]. Of note, the Y466 residue is also highly conserved among HBoV1-3 and GBoV but is substituted by a phenylalanine in HBoV4, which might reflect a certain degree of tolerance that explains the milder effect on transduction. Notably, not only surface-exposed tyrosines play a role in the proteasomal processing of AAVs. Several studies have shown that also serine, lysine and threonine residues could be hotspots for ubiquitination and phosphorylation [324, 325]. The recently reported higher-resolution structure of HBoV1 [302] is likely to facilitate the future mapping and testing of such residues that are truly located on the surface.

Next to targeted capsid modification, additional options for improving rAAV/BoV vectors may arise as our understanding of wtHBoV biology increases. Recently, several studies reported how HBoV1 interacts with different cellular pathways to promote its infection, including the degradation of host proteins that would otherwise impede viral growth. For example, the NP1 protein was shown to interact with IRF-3 and interrupt its binding to the *IFNB* (interferon beta) promoter, thereby downregulating promoter activity. This, in turn, reduces IFN I signalling and promotes infection [326]. Consequently, *np1* was suggested as an early gene in the viral life cycle. Another study showed that the VP2 protein upregulates IFN production by inhibiting the proteasome-dependent degradation of RIG I [74]. It was speculated that this

benefits the establishment of latency, as previously observed with other viruses such as EBV [327]. Also, this suggests that *vp2* may act as a late gene in the viral life cycle.

Finally, it has been shown that the NS non-structural proteins also play a role in infection. NS1 and NS1-70 negatively regulate the NF-κB pathway by directly interacting with the p65 subunit of the NF-κB proteins [73]. Interesting in this context is our observation of accumulated 20S proteasome cores in the wt and recombinant BoV preparations (see *e.g.* GBoV/Gluc in **Figure 17**), which may reflect a stimulation of the immunoproteasome in response to IFN production [328]. Altogether, this implies that HBoV1 infection elicits a tightly controlled and multifactorial immune response, which the virus evades to some degree using an equally tightly regulated expression of early and late viral genes. Consequently, this raises the question whether the deletion of viral elements and ORFs in the vector system is the cause for the strict dependency of BoV transduction on the presence of PIs.

While a dissection of these possibilities was beyond the scope of this thesis, one can readily anticipate that a deeper understanding of fundamental BoV biology including endosomal escape, proteasome interactions, capsid uncoating or immune responses will inform and benefit the design of enhanced, next-generation bocaviral vectors. To this end, strategies used in the autonomous parvovirus field could be applied to entangle the effect of different viral proteins on transduction. For example, it would be interesting to test whether *cap*-replacement vectors [83-86], in which only a part of the *cap* gene is replaced by a transgene of interest, could restore transduction efficiency in the absence of PIs (see section 1.2.1 and **Figure 3B** for more details).

4.1.5 DNA family shuffling allows the creation of large, packaging-competent and chimeric BoV libraries

As discussed above, a deeper understanding of BoV biology is likely to enable rational approaches to increase viral fitness in the future. Until then, high-throughput screening methods such as random mutagenesis, domain swapping or DFS can already be applied to yield functional insights. Moreover, a combination of these methods with a specific enrichment approach that involves iterative selection rounds should enable the identification of novel variants with desired properties for therapeutic gene transfer.

In this work, we studied the applicability of DFS as a method to create a library of chimeric BoV particles. This was fueled by the tremendous success of this technology in the rAAV field, where it led to the discovery of chimeric capsid variants with distinct tropism, enhanced specificity and/or immunological profile [111, 297, 329]. Chimeric viral libraries also advanced our knowledge of AAV biology by revealing important information regarding antigenic domains and receptor binding [111, 329].

One prerequisite for the successful application of DFS is the high sequence identity between parental genes, which is ideally in a range of 50% or higher. Fortunately, with >70%

sequence identity, BoV *cap* ORFs represent suitable candidates for such an approach. Indeed, the BoV library created in this work was highly diverse with an average of nine cross-overs per clone, which is comparable to what is typically achieved using AAV *cap* sequences [300]. Importantly and congruent with previous reports [330], we detected an over-representation of parental sequences with higher identity to each other (HBoV2-4).

After the successful generation of a chimeric BoV plasmid library, the first question we asked was whether chimeric viral particles can be produced that retain their structural and functional integrity. Many previous studies have shown that structural integrity of proteins is typically not destroyed during the process of DFS, due to the recombination of functionally relevant domains [111, 331, 332]. Indeed, in the present study, BoV particles containing a viral genome were detected using DNA dot blot and qPCR analysis. Moreover, the titer of the viral library increased by ~200-fold after the first selection round. This is in line with previous studies using shuffled AAV libraries and reflects a selection for packaging-competent chimeric viruses [329]. This result was further supported by testing individual candidates from selection round three, which gave higher viral yields than rAAV2/HBoV1 vectors.

In the next step, we tested the infectivity of the candidates in pHAE. All tested chimeras were inferior to HBoV1 in their transduction ability and resulted in 10-fold lower transgene expression. This was unexpected in view of the promising data with shuffled AAV *cap* sequences, whose selection in cells frequently resulted in superior capsids [111, 297]. One reason for the observed phenotype might be the under-representation of HBoV1-derived sequences in the initial library, which could have severely compromised the functionality of the library in pHAE. Luckily, the abundance of certain sequences in the library can easily be improved by careful adjustment of input fragment concentrations. Another way to enhance library quality and thereby foster chances of success is to apply codon-optimization to the *cap* ORFs that increases the sequence similarity and thus the DNA recombination frequency [239].

Despite the sequence bias and the low functionality of the chimeras, it is intriguing that the C-terminus of VP3 was nearly entirely composed of HBoV1-derived sequences after the third selection round. Particularly interesting and thus discussed further is VR-VIIIB, which shows the highest sequence divergence in HBoV1 and GBoV. This VR corresponds to the so-called HI loop in AAV, which was shown to play a dynamic function in the virus infection pathway by controlling the exposure of the VP1 N terminus and the incorporation of VP1 subunits at the 5-fold axis. Notable in this context is also the above-mentioned change in HBoV1 of a conserved tyrosine residue flanking the HI loop (Y466) into a phenylalanine, which resulted in a 6-fold loss in transduction (see sections 3.1.3 and 4.1.4) and thus reflects the importance of this loop in HBoV1 infection. Importantly, it was shown that polarized cells can shuttle viruses into different endocytic compartments, some of which are “dead ends” and do not contribute to transduction [333]. Thus, it will now be interesting to investigate whether this

region plays a role in proper post-entry processing of HBoV1, which may in turn underlie its higher transduction in pHAE as compared to the other BoV isolates tested here.

Moreover, the HI loop was linked to capsid assembly and genome packaging. For example, substitution of the HI loop in AAV2 with that from AAV1 or 8 (both 83% identical to AAV2 on the capsid amino acid sequence) resulted in 2-fold lower viral titers [323]. By contrast, swapping the HI loop from a less homologous AAV serotype (AAV4 and AAV5, having 61% and 59% amino acid identity to AAV2, respectively) resulted in a 1-log reduction in virus titer (AAV4 HI) or complete absence of capsid assembly (AAV5 HI). Importantly, the infectivity of the HI loop swap mutants was not compromised (except for AAV2 carrying the HI loop from AAV5) which initially suggested that this loop might be less important for infection. Mutations of highly conserved residues within the loop, however, led to improper trafficking, which resulted in a substantial decrease in infectivity [323].

This is also particularly interesting in view of a mutation that we recently reported in patient samples at amino acid position 590 (VP1 numbering), which lies in the VR-VIIIB region and affected viral serum load [281]. In the course of this other study, we experimentally verified that the presence of a serine versus a threonine at position 590 affects viral titers but not infectivity (Kai-Philipp Linse, MD thesis), which is strikingly congruent with what was reported for AAV2 [323] and supports the hypothesis of an evolutionarily conserved function of this region. Collectively, our data suggest a critical role of the HI loop and flanking amino acids in the infectious pathway of HBoV1. It should now be intriguing to perform site-directed mutagenesis or to swap the HI loop in the HBoV1 capsid with orthologs, to study whether it has implications for wtHBoV1 infectivity and assembly as well.

The VR-IX region also showed a preference for HBoV1-derived sequences. This might be a result of the lower recombination rates between VR-IIIB and VR-IX in the unselected library. Interestingly, despite the high heterogeneity of this region, the structural arrangement of the loops is conserved among HBoVs [302]. Therefore, this region and VR-VIII were suggested as determinants of the host tropism [302]. This, however, does not rule out an as-of-yet undefined role in the HBoV1 infection pathway, akin to what was reported for MVMp [334].

After one selection round in pHAE, another striking pattern was the site-specific, nearly exclusive accumulation of HBoV4-derived sequences in VR-III to VR-IV, which extended to involve VR-V after the third selection round. Notably, this pattern was equally observed in the presence and absence of PIs. VR-III is similar among HBoV2-4, whereas the highest divergence is observed in VR-IV and VR-V [302]. Thus, in view of the higher transduction ability of HBoV4 as compared to HBoV2 and HBoV3, VR-IV and VR-V might be important for HBoV4 cell binding and/or internalization.

Important to mention is that VR-III is extended in HBoV1 and GboV, and has been suggested as a determinant of the respiratory tropism of HBoV1 [302]. In our study, we could not

find evidence for this hypothesis. In fact, the HBoV1/GBoV VR-III was completely depleted after three selection rounds. However, this does not rule out an important role in transduction, in view of the 10-fold lower transduction ability of HBoV4 as compared to HBoV1 and GBoV. One explanation for the nonetheless striking preferential accumulation of HBoV4 VR-IV and VR-V is that these domains might trigger a superior binding to the cell surface, which, however, leads to a less efficient entry pathway. The under-representation of HBoV1 and GBoV sequences in the initial library and the low selection pressure for a productive transduction might have further supported the loss of these sequences.

Interesting was also the enrichment of HBoV2 in VR-VIII to VR-IX in the presence of PIs but not in their absence, and the gradual disappearance of this pattern over the selection rounds. One possible explanation could be an initial accumulation of variants that produced better but had a lower infectivity and were thus lost with an increased selection pressure for infectivity. The presence of PIs could have protected these viral capsids from degradation and thereby favored their accumulation in the library after the first selection round. Supporting this hypothesis is the reduction in library titers from 9×10^{11} (after selection round 1) to 2×10^{11} gc/mL (after selection round 2), which is congruent with previous observations with AAV libraries [329].

While initially surprising, the reduced infectivity of the resulting chimeric capsids is an important result as it highlights crucial considerations about library construction and study design, and thus provides a roadmap for future work. For instance, the loss of HBoV1 sequences during the initial assembly may have negatively influenced library performance and selection in pHAE. It would thus be interesting and crucial to now test this BoV library in other cell types to gather information about its general viability and applicability, and to assess whether the result obtained here is truly specific for pHAE. Concurrently, to increase chances of isolating a good transducer, a selection step on the RNA level can be included, as this will inherently enrich particles that mediate complete transduction, rather than capsids that merely deliver their DNA cargo to cells but may not express the encoded transgene. Luckily, the implementation of such a RNA- instead of DNA-based selection scheme should be possible in view of the ubiquitous and high activity of the bocaviral p5 promoter, which will drive robust *cap* mRNA expression in library-infected cells and thus facilitate RNA rescue and reverse transcription for subcloning into a secondary library [335].

4.2 KS and SIN are promising designs for inactivation of transgene expression from rAAV vectors

Among all viral vectors that are currently used in the gene therapy field, rAAVs have proved an outstanding potential and have thus become lead candidates for *in vivo* applications. This success is particularly due to their high safety profile, which is characterized by a low

genotoxicity and a mild host immune response [336]. Nevertheless, it would be important for the next vector generation to incorporate a system that allows to efficiently shut off transgene expression from rAAV vectors in the event of adverse side effects or toxicity. The latter has indeed been observed *e.g.* after rAAV-mediated shRNA delivery into the mouse liver, which ultimately led to hepatotoxicity and even animal fatalities because of saturation of the endogenous miRNA pathway [241]. Moreover, it was shown that AAV9-based vectors could also elicit severe hepatotoxicity in non-human primates when administered at a high dose and that this toxicity seems to result from oxidative stress [337]. Importantly, the latter study continues to be discussed very controversially in the field because of the low number of animals and the lack of toxicity in other studies that involved the same dosages in humans. Regardless of the exact reasons for these discrepancies (that certainly require clarification and independent validation), the observed toxicities were unexpected and, in some of the animals, even life-threatening.

These two examples and other similar findings clearly highlight the urgent need for the implementation and preclinical evaluation of new safety measures in rAAV vectors, including switches that allow for a rapid vector incapacitation in the event of adverse patient reactions. Thus, in this part of the work, we focused on establishing and validating different novel means for deliberate inactivation of gene expression from rAAV vectors subsequent to transduction. To this end, we constructed both, “kill switch” (KS) and “self-inactivating” (SIN) rAAV vectors as two innovative approaches to control the duration of transgene expression. Two critical components in both systems are the endonuclease Cas9 and customized gRNAs that are directed against the vector-encoded transgene. Our successful demonstration of the great potential of these systems for *in vitro* and *in vivo* inactivation of rAAV transgene expression allows us to envision numerous future applications in biological research and gene therapy, as further discussed below.

4.2.1 Kill-switch rAAV vectors can be efficiently inactivated *in vitro* and *in vivo*

To build a simple yet effective kill switch that permits to inactivate transgene expression at any given time, two plasmids were constructed (see **Figure 39A** and **Figure 46A**): (i) a KS reporter that constitutively expresses a transgene (FLuc) together with a gRNA directed against this transgene, and (ii) the Cas9 endonuclease. In combination, these form an off-switch because transgene expression is switched off only when the Cas9 is expressed or delivered *in trans*. As two alternative strategies, the gRNA cassette can also be expressed (i) from the same vector as the Cas9 or (ii) from a third, separate plasmid. We did not further pursue the first of these two alternative approaches because it was hardly compatible with the size limit of rAAV vectors (~4.7 kb), especially when the SpCas9 (4.1 kb) is used (as in the first validation experiments of the KS reporter, sections 3.4.2 and 3.4.3). We compared, however,

the second approach to our original KS system (see above; transgene and gRNA are on the same vector) and showed a superior inactivation of the FLuc transgene in the KS reporter (~10-fold). This is conceivable in view of the smaller number of plasmids that have to enter the same cell (two in the case of the genuine KS *versus* three in the alternative approach).

Interestingly, we detected a reduced expression of the RLuc reporter, which is co-expressed from the same KS plasmid as FLuc but was not targeted by the Cas9. Also, T7 assays performed on the gRNA target region revealed a very low editing efficiency that did not correlate with the robust FLuc functional knockout (data not shown). Since the T7 assay reflects the degree of NHEJ (the most common DNA repair pathway in the cell) and since NHEJ of FLuc should not have affected the RLuc reporter, we concluded that the plasmid DNA is preferably degraded after Cas9 cleavage rather than being repaired by NHEJ.

Next, we transferred our system into a rAAV context and asked whether AAV vector genomes can be targeted *in vivo* in a similar way. This was important to test because rAAV vector genomes can assume various molecular forms in which they persist and express [305], and because it was unclear whether the predominant forms - episomal, circular monomers or concatemers - can be recognized and cleaved by CRISPR/Cas9 akin to plasmid DNA. To this end, we packaged the KS reporter into AAV8 and injected it into transgenic mice expressing Cas9 in a Cre-dependent fashion. Hence, to activate the endogenous Cas9 expression, we co-delivered an AAV8 vector expressing Cre under a liver-specific promoter. While following the kinetics of FLuc transgene expression over time, we detected a “pulse-like” profile characterized by an initial accumulation of the reporter and a subsequent degradation phase because of Cas9 activation and cleavage of the reporter.

Interestingly, the FLuc expression kinetics in our *in vivo* study and the degradation of plasmid DNA in transfection experiments were congruent with a previous *in vitro* study by Moore *et al.*, who used a “self-destructive” one-plasmid system composed of: (i) a reporter gene, (ii) the Cas9 endonuclease and (iii) a gRNA cassette directed against the reporter. In this study, the authors noted a bell-like expression of the Cas9-targeted reporter gene and a concomitant knockout of another reporter, which was co-expressed from the same plasmid. The latter strengthens our own data and hypothesis of plasmid DNA degradation after Cas9-mediated cleavage. Another interesting observation in this reported study is that the amplitude and duration of reporter expression can be controlled by altering the Cas9-gRNA affinity to their target sequence. We envision that a similar adjustment of transgene expression from our kill-switch circuit might be of particular value in the field of regenerative medicine. There, it was shown that cellular reprogramming factors can be delivered by rAAV [100] or Sendai virus vectors [338] to various organs *in vivo* to induce the formation of pluripotent stem cells. However, it was also noted that short-term cyclic induction of their expression is required to promote safe tissue regeneration and to reduce the incidence of adverse teratoma formation [339]. Thus, it will be

very interesting to restrict and control the expression of these potent factors by including self-targeting gRNAs. In this context, it might not only be important to restrict the transgene expression, but also to control the duration and amplitude of the pulse. Here, it would be promising to test gRNAs with different affinities to their targets as this would permit vector fine-tuning *in vivo*.

Taken together, our data clearly motivate the future expansion of these experiments and the continued development of the overall concept towards application in humans. In this respect, we acknowledge that the clinical relevance of the Cas9-transgenic mice that we used in our pilot work is obviously limited, yet this experiment provided us with the first seminal proof-of-principle that rAAV vectors can indeed be efficiently targeted *in vivo* using the SpCas9 endonuclease. Important to mention in this context is that after we had obtained these results, a study by Li *et al.* was published in 2018 [340] that independently supports our data and conclusion, by also showing efficient cleavage of rAAV vectors in the mouse liver. Further noteworthy is that in our own study, we had used the relatively weak H1 RNA Pol III promoter to drive gRNA expression. Thus, it should be possible to even further improve the observed ~50% reduction in transgene expression by using stronger RNA Pol III promoters, such as U6.

Another intriguing option for future improvement towards use in humans is to deliver Cas9 mRNA *in vivo* using LNPs (lipid nanoparticles) instead of encoding it in viral vectors [341-343]. The advantages of such an alternative Cas9 delivery system are manifold and include *e.g.*, the high tolerability of LNPs and the transient expression of the Cas9, which was shown to decrease the risk of off-targeting [303]. Even more encouraging is that a new generation of LNPs has now be formulated that allows for concomitant delivery of Cas9 mRNA and chemically synthesized gRNAs. The promise of this approach has recently been exemplified by Finn *et al.* [341], who targeted the *Ttr* gene in the mouse liver. TTR protein levels started to go down significantly 20 h after intravenous LNP administration and reached their lowest levels after 180 h. This fast onset, good safety profile and small window of Cas9 expression make this new system very attractive for future applications in humans. However, it remains to be tested whether such a system can target rAAV vectors that persist in multiple copies in a cell and, if so, with the same efficiency as an endogenous target. Until then, there is consensus that rAAV vectors remain the most attractive tools for Cas9 delivery to different organs *in vivo* [12, 14, 209].

Collectively, the rAAV KS vector constructed in this work could be potently targeted and inactivated by Cas9 *in vitro* and *in vivo*, illustrating its great potential for a large variety of future applications including but not limited to human gene therapy. In fact, we have recently also reported that the rAAV KS reporter is a sensitive sensor with a large dynamic range and that it could be used as a tool to answer fundamental biological questions beyond its use in therapeutic applications [186].

4.2.2 KS rAAV vectors can be combined with scAAV SIN split Cas9 vectors to mediate potent reduction in transgene and Cas9 expression

Despite all the encouraging prospects of using Cas9 mRNA or RNP complexes, the delivery of these molecules to target cells is still inefficient, especially in *in vivo* applications. While the latter can be solved by expressing Cas9 from rAAV vectors, one limitation and readily conceivable disadvantage in their application is the long-lasting expression, which, as noted above, increases the risk of Cas9 off-targeting and immune responses. Here, we therefore devised and validated a new generation of self-inactivating (SIN) AAV/CRISPR vectors that combine efficient *in vivo* delivery with restricted and thus safer Cas9 expression.

To this end, it was important to realize that two parameters predominantly govern the efficiency of *in vivo* delivery and expression of rAAV vectors: (i) the AAV capsid and (ii) the rAAV genome. For example, AAV8 and AAV6 capsids efficiently transduce hepatocytes and blood cells, respectively [344, 345], making them preferred candidates for gene transfer into these types of cells. This is also reflected by the use of AAV8 in various liver-directed gene therapy trials in humans [345], and by the frequent pre-clinical exploitation of AAV6 for *ex vivo* delivery of CRISPR repair templates in *e.g.* hematopoietic stem cells [346]. Similarly, the currently available, AAV-based gene therapy products harness the efficiency of AAV1 in the muscle (Glybera) or AAV2 in the eye (Luxturna), or the ability of AAV9 to potently cross the blood-brain barrier (Zolgensma) [140, 144]. Complementing these efforts to fine-tune AAV vector efficiency and activity on the capsid level, others including us have shown that similar improvements can be obtained on the rAAV genome level. In particular, it was shown that scAAV vectors result in a faster and often more potent transgene expression as compared to ssAAVs [226].

Based on these two criteria, we aimed to optimize the Cas9 expression levels by first harnessing a novel split SaCas9 variant from our group that can be packaged into two scAAV vectors and therefore potentially results in faster onset and higher Cas9 expression than conventional ssAAV-Cas9 vectors (Carolin Schmelas; MSc. Thesis). Using this split Cas9 as a template, we added an anti-Cas9 gRNA cassette on top that is directed towards the N-terminal part of the Cas9 itself, to create a two vector-based scAAV SIN CRISPR system (see **Figure 46B**). Next, we packaged our constructs into the AAV2 capsid, which potently transduces HEK293T cells (up to 100%; data not shown). When these cells were co-transduced with the scAAV SIN CRISPR vectors and our KS reporter, a remarkably high concomitant anti-target and anti-Cas9 editing was achieved (section 3.4.5).

Importantly, we have also already transferred our system into an *in vivo* context and performed a pilot experiment in mice to test the performance of our scAAV SIN split Cas9 system in comparison to conventional, full-length Cas9-expressing ssAAV vectors (ssAAV FLCas9). In this experiment, we aimed to predominantly target the mouse liver and thus packaged our constructs into AAV8, which is, as mentioned above, a potent AAV serotype in

the liver. Strikingly, our preliminary data imply superior Cas9 expression levels not only in the liver but also in nearly all other organs that we analyzed in parallel (work performed together with Carolin Schmelas; data not shown). The varying degrees of increases in Cas9 expression that we measured are congruent with previous reports showing that the improvement in expression observed with scAAV vectors is cell- [226] and tissue-type-dependent [347]. In the liver, the knockout efficiency of the KS-luciferase reporter was ~2-fold higher when the scAAV split Cas9 system was used as compared to ssAAV FLCas9. This fold increase in efficiency is in concordance with previous observations [348], showing a moderate increase in transgene expression (1.5- to 2.6-fold) from scAAV vectors in the mouse liver as compared to ssAAVs. The higher expression from scAAVs was linked to a larger number of transduced cells, improved vector stability and the preferable formation of circular monomeric forms [345, 347, 349] , which are the transcriptionally active forms of rAAV vectors in cells.

A particularly intruiguing and thus noteworthy observation in our most recent pilot study was the up to 300-fold increased expression of split Cas9 from SIN scAAV vectors in the mouse skeletal muscle, as compared to ssAAV FLCas9. This is well in line with a study by Wang *et al.* [347] that revealed a 15-fold higher expression from scAAV as compared to ssAAV vectors in the mouse muscle. Thus, we are now planning a more comprehensive and further optimized study to test whether the superior performance of our system in this tissue will also translate into better *in vivo* knockout efficiencies. If observed, it will then be very tempting to test our vectors in animal models of human muscle disorders that are amenable to treatment with CRISPR, such as *mdx* or Ex50 mice which recapitulate frequent mutations that cause Duchenne muscular dystrophy in humans and that can, in principle, be repaired by AAV/CRISPR-induced exon skipping [350, 351].

Yet another interesting target for scAAV split Cas9 application is the eye, based on data by *e.g.*, Yoki *et al.* who showed that scAAV vectors outperformed ssAAVs in photoreceptor cells within the mouse eye [352]. By contrast, the improvement in transduction efficiency in retinal pigment epithelial cells was only moderate, which again supports the existence of a cell-type dependent effect.

Notably, in parallel to our work and that by Li *et al*. in AAV vectors [340], several other research groups have also begun to restrict Cas9 expression in a lentiviral vector context by applying, *e.g*, the “lentiSLiCES” [303] or the “KamiCas9” strategy [353]. In both approaches, the Cas9 is inactivated by supplying an anti-Cas9 gRNA but they differ in the positioning of the gRNA cassette. In the lentiSLiCES system, a lentiviral vector was engineered that simultaneously expresses the Cas9, the target gRNA and the anti-Cas9 gRNA. To prevent Cas9 cleavage during vector production, the expression of Cas9 is regulated using tetracycline-responsive elements and producer cell lines stably expressing the Tet repressor. By contrast, in the KamiCas9 system, the Cas9 nuclease and the gRNAs were expressed from

separate transfer vectors. Importantly, regardless of gRNA positioning and the vector system used, all the above-mentioned studies and our own data consistently show high and comparable on-target activities *in vitro* and *in vivo*, with and without the SIN genotype.

Finally, it is pivotal to mention that Cas9 activity could also be inhibited on the protein level using phage-derived anti-CRISPR inhibitors (Acr proteins) that have recently been discovered and adapted for use in mammalian cells [186, 354, 355] (see section 1.6). One interesting application of these proteins in the future is to *e.g.*, prevent Cas9 activity during vector production, in case this process would be rate-limited by premature self-cleavage of the Cas9-encoding vector. Acr proteins can also be utilized for spatio-temporal control of Cas9 expression, as discussed in section 4.2.3 and as jointly reported by colleagues and us in two recently published studies [186, 356].

4.2.3 Towards a costumizable all-in-one rAAV SIN split Cas9 system for temporal and/or spatial control of Cas9 expression

The perhaps most critical improvement of the current AAV/CRISPR vector generation including our new SIN split Cas9 variants towards clinical use will be to implement features that allow to spatially and temporally control Cas9 expression in treated patients using endogenous or exogenous triggers. As discussed in section 4.2.2, the split Cas9 scAAV vectors that were implemented here can outperform ssAAV counterparts in terms of Cas9 expression; however, they are concurrently more limited with respect to the incorporation of regulatory elements owing to the limited size of scAAV vectors. Therefore, we aimed to further expand the power and potential of the split Cas9 systems by transferring the split halves into ssAAV vectors, along with the anti-target gRNA and the anti-Cas9 gRNA. As compared to the scAAV context, this setting provides multiple additional and beneficial features, such as: (i) a higher flexibility in the choice of promoters and regulatory elements, and (ii) an enhanced spatial and/or temporal control that can be acquired by coupling functional modules to the split halves of Cas9, or by expressing other components *in trans*, which do not fit into scAAVs. Importantly, we showed that the ssAAV SIN split Cas9 vectors constructed in this work can mediate potent knockout of both, target gene and Cas9. In addition, we demonstrated that these processes can be optimized by regulating the kinetics of anti-Cas and anti-target gRNA expression.

For future work, we envision several combinatorial approaches that will further restrict and control Cas9 activity, and hence increase the safety index of CRISPR applications (see **Figure 46C**). For example, as discussed in section 1.7.2, there are several options to control the process of reconstitution of the split Cas9 into a full-length protein, including induction by using different triggers and by fusing Cas9 with cognate elements, such as those responding to ligands (*e.g.* rapamycin [216]) or to the exposure to light [357]. While the first approach requires the application (typically injection) of chemicals, which is an invasive procedure, light-based

strategies are usually safer and are currently only limited by the low tissue penetration, especially of the most-commonly applied green and blue light.

Indeed, photoactivatable domains have already been shown to be readily combinable with the split Cas9 itself or the Acr proteins (**Figure 46C**; inducible systems). In the first approach, Nihongaki *et al.* [357] constructed a photoactivatable split Cas9 (paCas9) by fusing light-sensitive proteins (called magnets) to each split Cas9 half. Only upon exposure to blue light, these domains can dimerize and hence reconstitute a functional full-length Cas9. In the second strategy, Bubeck [186] in collaboration with our lab followed a different indirect approach to restrict Cas9 activity, which is to engineer a photoactivatable domain (LOV2) into the SpCas9 inhbitor AcrIIA4. Co-expression of this hybrid inhibitor together with the SpCas9 allowed for spatio-temporal and light-dependent control of Cas9 protein activity. One advantage of this latter approach is that it does not require any further modifications of the Cas9 or gRNA (as compared to *e.g.*, paCas9), rendering it compatible with all commonly used full-length and split Cas9 versions available to date.

In conclusion, the optogenetic regulation of AAV/CRISPR activity is highly promising and would confer an additional level of regulation to our SIN split Cas9 system. However, further optimization is still required to reach tissues that are deeper in the body without the use of invasive strategies. Until then, we and others are engineering alternative Cas9 systems that are restricted in their expression by endogenous triggers or that can autonomously regulate themselves. For example, our ssAAV SIN CRISPR vectors would be readily combinable with strategies using miRNA-dependent regulation, to specifically turn off (mir-OFF) [358] or activate (mir-ON) [356] Cas9 expression (**Figure 46C**; post-transcriptional or translational control). Particularly exciting in this respect is the on-switch approach that has resulted from collaborative work of our group with the Niopek lab and has recently been jointly reported by Hoffman *et al.* [356]. In this approach, miRNAs were used to regulate the expression of Acr proteins, by fusing the *acr* cDNA with cognate miRNA binding sites. Consequently, Cas9 expression can be turned on in the presence of certain cell-specific miRNAs that downregulate the expression of the Acr protein. The *acr* ORF could be easily cloned into our ssAAV SIN split Cas9 vectors, which would permit tissue-specific (as defined by the selected miRNA) but transient Cas9 expression and thus combine the best of both worlds.

The encouraging *in vitro* data reported for regulated Acr protein expression [186, 356] furthermore support their future application for Cas9 inhibition *in vivo*, as a novel and perhaps clinically relevant option to spatio-temporally control Cas9 activity and, thus, therapeutic gene editing or chromatin modification. However, it should be considered that, similar to the bacterial Cas9 [359], (i) Acr proteins can elicit an immune response because of their phage origin, and/or that (ii) the pre-existing immunity in humans may hamper the *in vivo* application of these proteins. Moreover, in contrast to the SIN CRISPR approach, the inhibition of Cas9 expression

can occur on multiple levels (dependent on the Acr protein [182, 183]) but the Cas9 vector DNA and its expression remain unaltered. Therefore, it is most likely that a combination of different techniques and strategies will ultimately be needed to maximize control over AAV/CRISPR vectors *in vivo* and to hence maximize patient safety.

Finally, Shen *et al.* have recently described an autonomous Cas9 system that combines a SIN approach with translational control [360]. Therefore, akin to miRNA-dependent regulation, K-turn motifs were placed at the 5′ UTR of the Cas9 mRNA, to which the L7Ae protein can bind in order to prevent Cas9 translation (**Figure 46C**; translational control). In contrast to miRNAs, the L7Ae protein is not endogenously expressed and has to be encoded on the same Cas9 vector *in cis.* Strikingly, the combination of a SIN full-length Cas9 with K-turn/L7Ae translational control resulted in a synergistic effect (>95% reduction of Cas9 protein). It should now be interesting to investigate whether such an extra inhibition on the level of Cas9 translation is also beneficial when juxtaposed with our AAV SIN split Cas9 approach.

In the context of therapeutic gene editing, all the above-mentioned approaches to spatially and temporally control Cas9 expression serve a common ultimate purpose, which is patient safety. One important parameter that has briefly been mentioned before (section 3.4) is the immune response against the Cas9, which can lead to serious adverse side effects [304, 359]. While many of the discussed strategies offer a confinement of Cas9 activity, Cas9 expression (mRNA/protein) itself is often not controlled. This highlights an additional advantage of the self-targeted SIN CRISPR systems including the one reported in this work, which target the Cas9 on the DNA level and thus further impede mRNA and protein expression. Important to mention along these lines is that the immunogenicity of the split parts and/or intein residues used in this work is currently unknown and requires further investigation. Our analysis of Cas9 protein expression / degradation kinetics revealed a high stability of the split Cas9 N part in HEK293T cells (the split C part is not recognized by the antibody, see sections 3.4.4 and 3.4.5). Consequently, another layer of safety can be introduced by reducing the stability of the free split parts. This can be achieved, for instance, by fusing the inteins to protein-destabilizing domains, such as SopE, ER50 or ddFKBP [361]. Then, intein-mediated protein splicing excises out the inteins and the destabilizing domains, thereby rescuing the reconstituted protein from degradation. Importantly, this approach can only be applied to split Cas9 in the ssAAV context because the smallest of these domains (SopE) is around ~300 bp and thus cannot be used together with scAAV split Cas9 vectors.

As a whole, we have successfully constructed SIN split Cas9 vectors and confirmed efficient on-target activity and self-inactivation of these vectors. Moreover, we demonstrated the applicability of both, ss- and scAAV vectors for the delivery of the SIN CRISPR components. Importantly, the choice of one of the discussed systems over the other is highly

dependent on the intended application, route of delivery and site of activity (see **Figure 46** for an overview of designs and applications). For example, scAAV split Cas9 vectors would be a good choice for target tissues that are hardly transduced with conventional ssAAV vectors. Optogenetic control might be an option for applications in the eye or other sites within the body that are easy to access by non-invasive means, such as skin or the gums. Finally, translational regulation using miRNAs might be especially promising in tissues with specific and high miRNA expression levels, *e.g.*, miR-122 in the liver [362].

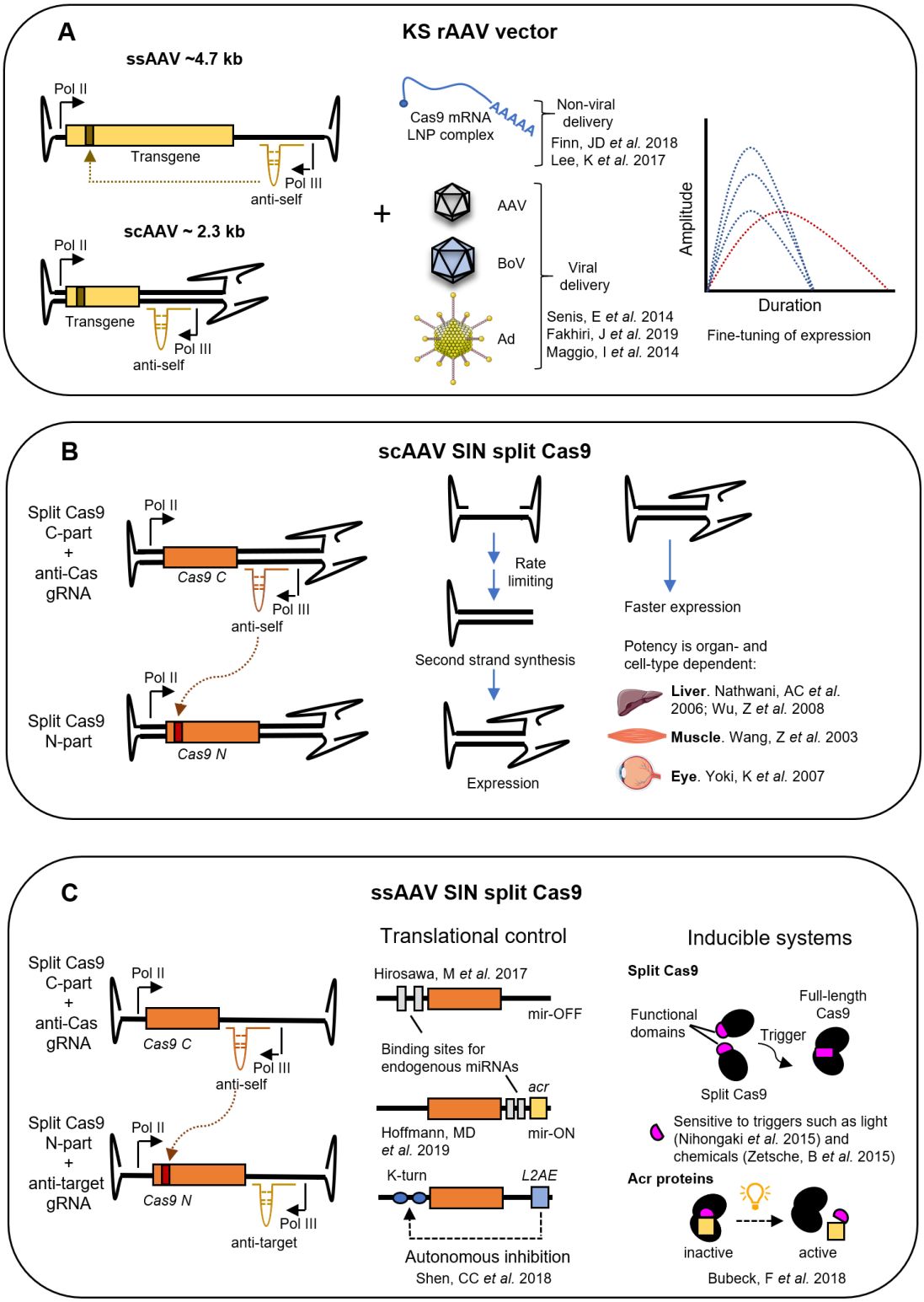


Figure 46. Deliberate inactivation of rAAV vectors using CRISPR/Cas9 and means for optimized spatio-temporal control of Cas9 expression.

(A) Kill-switch rAAV vectors as described in this work. An anti-self gRNA is constitutively expressed and promotes the cleavage of the vectors upon expression of Cas9 *in trans* (viral-/non-viral-mediated delivery). The dose of the rAAV-transgene vector defines the “amplitude” of expression. In addition, the affinity of the Cas9/gRNA complex to the target site affects both, the kinetics and amplitude. **(B)** SIN split Cas9 vectors expressed from a scAAV backbone. The gRNA expression cassette against the Cas9 N-terminus is positioned on the C-terminal split Cas9 vector. Expression from scAAV is usually faster

and more potent than from ssAAV because these vectors bypass the second-strand DNA synthesis step. The degree of fold increase is cell- and organ-dependent. **(C)** SIN split Cas9 vectors expressed from a ssAAV backbone. The anti-self and anti-target gRNAs are positioned on the split Cas C and split Cas N halves, respectively. Because ssAAV genomes are twice the size of scAAVs, additional modules can be incorporated to allow for a control on the level of protein translation and/or formation of functionally active Cas9 protein. This figure contains free clipart from https://smart.servier.com/ and https://www.vecteezy.com/.

5 REFERENCES

1. Barber, MA (1911). A technic for the inoculation of bacteria and other substances into living cells. *J Infect Dis* **8**: 348–360.
2. Jackson, DA, Symons, RH, and Berg, P (1972). Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. *Proc Natl Acad Sci USA* **69**: 2904-2909.
3. Cohen, SN, and Chang, AC (1973). Recircularization and autonomous replication of a sheared R-factor DNA segment in Escherichia coli transformants. *Proc Natl Acad Sci USA* **70**: 1293-1297.
4. Jaenisch, R, and Mintz, B (1974). Simian virus 40 DNA sequences in DNA of healthy adult mice derived from preimplantation blastocysts injected with viral DNA. *Proc Natl Acad Sci USA* **71**: 1250-1254.
5. Culver, KW, Anderson, WF, and Blaese, RM (1991). Lymphocyte gene therapy. *Hum Gene Ther* **2**: 107-109.
6. Blaese, RM, Culver, KW, Miller, AD, Carter, CS, Fleisher, T, Clerici, M*, et al.* (1995). T lymphocyte-directed gene therapy for ADA-SCID: initial trial results after 4 years. *Science* **270**: 475-480.
7. Marchini, A, Bonifati, S, Scott, EM, Angelova, AL, and Rommelaere, J (2015). Oncolytic parvoviruses: from basic virology to clinical applications. *Virol J* **12**: 6.
8. Berns, KI, and Muzyczka, N (2017). AAV: An overview of unanswered questions. *Hum Gene Ther* **28**: 308-313.
9. Cotmore, S, and Tattersall, P (2006). Structure and organization of the viral genome. *Parvoviruses*. Edward Arnold: Great Britain, 74-90 pp.
10. Yan, Z, Keiser, W, Song, Y, Deng, X, and Cheng, F (2013). A novel chimeric adenoassociated virus 2/human bocavirus 1 parvovirus vector efficiently transduces human airway epithelia. *Mol Ther* **21**: 2181-2194.
11. Swiech, L, Heidenreich, M, Banerjee, A, Habib, N, Li, Y, Trombetta, J*, et al.* (2015). In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9. *Nat Biotechnol* **33**: 102-106.
12. Ran, FA, Cong, L, Yan, WX, Scott, DA, Gootenberg, JS, Kriz, AJ*, et al.* (2015). In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**: 186-191.
13. Ruan, GX, Barry, E, Yu, D, Lukason, M, Cheng, SH, and Scaria, A (2017). CRISPR/Cas9-mediated genome editing as a therapeutic approach for leber congenital amaurosis 10. *Mol Ther* **25**: 331-341.
14. Nelson, CE, Hakim, CH, Ousterout, DG, Thakore, PI, Moreb, EA, Castellanos Rivera, RM*, et al.* (2016). In vivo genome editing improves muscle function in a mouse model of Duchenne muscular dystrophy. *Science* **351**: 403-407.
15. Atchison, RW, Casto, BC, and Hammon, WM (1965). Adenovirus-associated defective virus particles. *Science* **149**: 754-756.
16. Buller, RM, Janik, JE, Sebring, ED, and Rose, JA (1981). Herpes simplex virus types 1 and 2 completely help adenovirus-associated virus replication. *J Virol* **40**: 241-247.
17. Walz, C, Deprez, A, Dupressoir, T, Durst, M, Rabreau, M, and Schlehofer, JR (1997). Interaction of human papillomavirus type 16 and adeno-associated virus type 2 co-infecting human cervical epithelium. *J Gen Virol* **78**: 1441-1452.
18. Schlehofer, JR, Ehrbar, M, and zur Hausen, H (1986). Vaccinia virus, herpes simplex virus, and carcinogens induce DNA amplification in a human cell line and support replication of a helpervirus dependent parvovirus. *Virology* **152**: 110-117.
19. Wang, Z, Deng, X, Zou, W, Engelhardt, JF, Yan, Z, and Qiu, J (2017). Human bocavirus 1 is a novel helper for adeno-associated virus replication. *J Virol* **91**: e00710-17.

20. Hoggan, MD, Blacklow, NR, and Rowe, WP (1966). Studies of small DNA viruses found in various adenovirus preparations: physical, biological, and immunological characteristics. *Proc Natl Acad Sci USA* **55**: 1467-1474.
21. Glauser, DL, Strasser, R, Laimbacher, AS, Saydam, O, Clement, N, Linden, RM*, et al.* (2007). Live covisualization of competing adeno-associated virus and herpes simplex virus type 1 DNA replication: molecular mechanisms of interaction. *J Virol* **81**: 4732-4743.
22. Linden, RM, Ward, P, Giraud, C, Winocour, E, and Berns, KI (1996). Site-specific integration by adeno-associated virus. *Proc Natl Acad Sci USA* **93**: 11288-11294.
23. Kotin, RM, Siniscalco, M, Samulski, RJ, Zhu, XD, Hunter, L, Laughlin, CA*, et al.* (1990). Site-specific integration by adeno-associated virus. *Proc Natl Acad Sci USA* **87**: 2211-2215.
24. Samulski, RJ, Zhu, X, Xiao, X, Brook, JD, Housman, DE, Epstein, N*, et al.* (1991). Targeted integration of adeno-associated virus (AAV) into human chromosome 19. *EMBO J* **10**: 3941-3950.
25. Nault, JC, Datta, S, Imbeaud, S, Franconi, A, Mallet, M, Couchy, G*, et al.* (2015). Recurrent AAV2-related insertional mutagenesis in human hepatocellular carcinomas. *Nat Genet* **47**: 1187-1193.
26. Logan, GJ, Dane, AP, Hallwirth, CV, Smyth, CM, Wilkie, EE, Amaya, AK*, et al.* (2017). Identification of liver-specific enhancer-promoter activity in the 3' untranslated region of the wild-type AAV2 genome. *Nat Genet* **49**: 1267-1273.
27. Smith, J, Herrero, R, Erles, K, Grimm, D, Munoz, N, Bosch, FX*, et al.* (2001). Adeno-associated virus seropositivity and HPV-induced cervical cancer in Spain and Colombia. *Int J Cancer* **94**: 520-526.
28. Mayor, HD, Drake, S, Stahmann, J, and Mumford, DM (1976). Antibodies to adeno-associated satellite virus and herpes simplex in sera from cancer patients and normal adults. *Am J Obstet Gynecol* **126**: 100-104.
29. Mayor, HD, Houlditch, GS, and Mumford, DM (1973). Influence of adeno-associated satellite virus on adenovirus-induced tumours in hamsters. *Nat New Biol* **241**: 44-46.
30. Berns, KI, Byrne, BJ, Flotte, TR, Gao, G, Hauswirth, WW, Herzog, RW*, et al.* (2015). Adeno-associated virus type 2 and hepatocellular carcinoma? *Hum Gene Ther* **26**: 779-781.
31. Hermonat, PL, 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 USA* **81**: 6466-6470.
32. Lebkowski, JS, McNally, MM, Okarma, TB, and Lerch, LB (1988). Adeno-associated virus: a vector system for efficient introduction and integration of DNA into a variety of mammalian cell types. *Mol Cell Biol* **8**: 3988-3996.
33. Sonntag, F, Kother, K, Schmidt, K, Weghofer, M, Raupp, C, Nieto, K*, et al.* (2011). The assembly-activating protein promotes capsid assembly of different adeno-associated virus serotypes. *J Virol* **85**: 12686-12697.
34. Gao, G, Vandenberghe, LH, and Wilson, JM (2005). New recombinant serotypes of AAV vectors. *Curr Gene Ther* **5**: 285-297.
35. Bello, A, Chand, A, Aviles, J, Soule, G, Auricchio, A, and Kobinger, GP (2014). Novel adeno-associated viruses derived from pig tissues transduce most major organs in mice. *Sci Rep* **4**: 6644.
36. Penzes, JJ, Pham, HT, Benko, M, and Tijssen, P (2015). Novel parvoviruses in reptiles and genome sequence of a lizard parvovirus shed light on Dependoparvovirus genus evolution. *J Gen Virol* **96**: 2769-2779.
37. Summerford, C, and Samulski, RJ (1998). Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions. *J Virol* **72**: 1438-1445.
38. Murlidharan, G, Samulski, RJ, and Asokan, A (2014). Biology of adeno-associated viral vectors in the central nervous system. *Front Mol Neurosci* **7**: 76.

39. Pillay, S, Zou, W, Cheng, F, Puschnik, AS, Meyer, NL, Ganaie, SS*, et al.* (2017). AAV serotypes have distinctive interactions with domains of the cellular receptor AAVR. *Virol* **91**: e00391-17.
40. Pillay, S, Meyer, NL, Puschnik, AS, Davulcu, O, Diep, J, Ishikawa, Y*, et al.* (2016). An essential receptor for adeno-associated virus infection. *Nature* **530**: 108-112.
41. Bowles, DE, Rabinowitz, JE, and Samulski, RJ (2006). The genus Depenovirus. *Parvoviruses*. Edward Arnold: Great Britain, 15-22 pp.
42. Kisary, J (1979). Interaction in replication between goose parvovirus strain B and duck plague herpesvirus. *Arch Virol* **59**: 81-88.
43. Zadori, Z, Stefancsik, R, Rauch, T, and Kisary, J (1995). Analysis of the complete nucleotide sequences of goose and muscovy duck parvoviruses indicates common ancestral origin with adeno-associated virus 2. *Virology* **212**: 562-573.
44. Qiu, J, Cheng, F, Yoto, Y, Zadori, Z, and Pintel, D (2005). The expression strategy of goose parvovirus exhibits features of both the Dependovirus and Parvovirus genera. *J Virol* **79**: 11035-11044.
45. Toolan, HW (1967). Lack of oncogenic effect of the H-viruses for hamsters. *Nature* **214**: 1036.
46. Toolan, HW, Rhode, SL, 3rd, and Gierthy, JF (1982). Inhibition of 7,12-dimethylbenz(a)anthracene-induced tumors in Syrian hamsters by prior infection with H-1 parvovirus. *Cancer Res* **42**: 2552-2555.
47. Guetta, E, Graziani, Y, and Tal, J (1986). Suppression of Ehrlich ascites tumors in mice by minute virus of mice. *J Natl Cancer Inst* **76**: 1177-1180.
48. Dupressoir, T, Vanacker, JM, Cornelis, JJ, Duponchel, N, and Rommelaere, J (1989). Inhibition by parvovirus H-1 of the formation of tumors in nude mice and colonies in vitro by transformed human mammary epithelial cells. *Cancer Res* **49**: 3203-3208.
49. Heegaard, ED, and Brown, KE (2002). Human parvovirus B19. *Clin Microbiol Rev* **15**: 485-505.
50. Rayet, B, Lopez-Guerrero, JA, Rommelaere, J, and Dinsart, C (1998). Induction of programmed cell death by parvovirus H-1 in U937 cells: connection with the tumor necrosis factor alpha signalling pathway. *J Virol* **72**: 8893-8903.
51. Malerba, M, Daeffler, L, Rommelaere, J, and Iggo, RD (2003). Replicating parvoviruses that target colon cancer cells. *J Virol* **77**: 6683-6691.
52. Salome, N, van Hille, B, Duponchel, N, Meneguzzi, G, Cuzin, F, Rommelaere, J*, et al.* (1990). Sensitization of transformed rat cells to parvovirus MVMp is restricted to specific oncogenes. *Oncogene* **5**: 123-130.
53. Legrand, C, Mousset, S, Salome, N, and Rommelaere, J (1992). Cooperation of oncogenes in cell transformation and sensitization to killing by the parvovirus minute virus of mice. *J Gen Virol* **73**: 2003-2009.
54. Telerman, A, Tuynder, M, Dupressoir, T, Robaye, B, Sigaux, F, Shaulian, E*, et al.* (1993). A model for tumor suppression using H-1 parvovirus. *Proc Natl Acad Sci USA* **90**: 8702-8706.
55. Angelova, AL, Geletneky, K, Nuesch, JP, and Rommelaere, J (2015). Tumor selectivity of oncolytic parvoviruses: from in vitro and animal models to cancer patients. *Front Bioeng Biotechnol* **3**: 55.
56. Cornelis, JJ, Deleu, L, Koch, U, and Rommelaere, J (2006). Parvovirus oncosuppression. *Parvoviruses.* Edward Arnold: Great Britain, 365-375 pp.
57. Bashir, T, Horlein, R, Rommelaere, J, and Willwand, K (2000). Cyclin A activates the DNA polymerase delta-dependent elongation machinery in vitro: a parvovirus DNA replication model. *Proc Natl Acad Sci USA* **97**: 5522-5527.
58. Daeffler, L, Horlein, R, Rommelaere, J, and Nuesch, JP (2003). Modulation of minute virus of mice cytotoxic activities through site-directed mutagenesis within the NS coding region. *J Virol* **77**: 12466-12478.
59. Abinanti, FR, and Warfield, MS (1961). Recovery of a hemadsorbing virus (HADEN) from the gastrointestinal tract of calves. *Virology* **14**: 288-289.
60. Binn, LN, Lazar, EC, Eddy, GA, and Kajima, M (1970). Recovery and characterization of a minute virus of canines. *Infect Immun* **1**: 503-508.

61. Harrison, LR, Styer, EL, Pursell, AR, Carmichael, LE, and Nietfeld, JC (1992). Fatal disease in nursing puppies associated with minute virus of canines. *J Vet Diagn Invest* **4**: 19-22.
62. Jarplid, B, Johansson, H, and Carmichael, LE (1996). A fatal case of pup infection with minute virus of canines (MVC). *J Vet Diagn Invest* **8**: 484-487.
63. Barnes, MA, Wright, RE, Bodine, AB, and Alberty, CF (1982). Frequency of bluetongue and bovine parvovirus infection in cattle in South Carolina dairy herds. *Am J Vet Res* **43**: 1078-1080.
64. Allander, T, Tammi, MT, Eriksson, M, Bjerkner, A, Tiveljung-Lindell, A, and Andersson, B (2005). Cloning of a human parvovirus by molecular screening of respiratory tract samples. *Proc Natl Acad Sci USA* **102**: 12891-12896.
65. Arthur, JL, Higgins, GD, Davidson, GP, Givney, RC, and Ratcliff, RM (2009). A novel bocavirus associated with acute gastroenteritis in Australian children. *PLoS Pathog* **5**: e1000391.
66. Kapoor, A, Mehta, N, Esper, F, Poljsak-Prijatelj, M, Quan, PL, Qaisar, N*, et al.* (2010). Identification and characterization of a new bocavirus species in gorillas. *PLoS One* **5**: e11948.
67. Manteufel, J, and Truyen, U (2008). Animal bocaviruses: a brief review. *Intervirology* **51**: 328-334.
68. Tu, M, Liu, F, Chen, S, Wang, M, and Cheng, A (2015). Role of capsid proteins in parvoviruses infection. *Virol J* **12**: 114.
69. Schwartz, D, Green, B, Carmichael, LE, and Parrish, CR (2002). The canine minute virus (minute virus of canines) is a distinct parvovirus that is most similar to bovine parvovirus. *Virology* **302**: 219-223.
70. Zou, W, Cheng, F, Shen, W, Engelhardt, JF, Yan, Z, and Qiu, J (2016). Nonstructural protein NP1 of human bocavirus 1 plays a critical role in the expression of viral capsid proteins. *J Virol* **90**: 4658-4669.
71. Weissbrich, B, Neske, F, Schubert, J, Tollmann, F, Blath, K, Blessing, K*, et al.* (2006). Frequent detection of bocavirus DNA in German children with respiratory tract infections. *BMC Infect Dis* **6**: 109.
72. Huang, Q, Deng, X, Yan, Z, Cheng, F, Luo, Y, Shen, W*, et al.* (2012). Establishment of a reverse genetics system for studying human bocavirus in human airway epithelia. *PLoS Pathog* **8**: e1002899.
73. Liu, Q, Zhang, Z, Zheng, Z, Zheng, C, Liu, Y, Hu, Q*, et al.* (2016). Human bocavirus NS1 and NS1-70 proteins inhibit TNF-α-Mediated activation of NF-κB by targeting p65. *Sci Rep* **6**: 28481.
74. Luo, H, Zhang, Z, Zheng, Z, Ke, X, Zhang, X, Li, Q*, et al.* (2013). Human bocavirus VP2 upregulates IFN-beta pathway by inhibiting ring finger protein 125-mediated ubiquitination of retinoic acid-inducible gene-I. *J Immunol* **191**: 660-669.
75. Song, JR, Jin, Y, Xie, ZP, Gao, HC, Xiao, NG, Chen, WX*, et al.* (2010). Novel human bocavirus in children with acute respiratory tract infection. *Emerg Infect Dis* **16**: 324-327.
76. Qiu, J, Cheng, F, Johnson, FB, and Pintel, D (2007). The transcription profile of the bocavirus bovine parvovirus is unlike those of previously characterized parvoviruses. *J Virol* **81**: 12080-12085.
77. Sukhu, L, Fasina, O, Burger, L, Rai, A, Qiu, J, and Pintel, DJ (2013). Characterization of the nonstructural proteins of the bocavirus minute virus of canines. *J Virol* **87**: 1098-1104.
78. Shen, W, Deng, X, Zou, W, Cheng, F, Engelhardt, JF, Yan, Z*, et al.* (2015). Identification and functional analysis of novel nonstructural proteins of human bocavirus 1. *J Virol* **89**: 10097-10109.
79. Fasina, OO, Dong, Y, and Pintel, DJ (2016). NP1 protein of the bocaparvovirus minute virus of canines controls access to the viral capsid genes via its role in RNA processing. *J Virol* **90**: 1718-1728.

80. Fasina, OO, Stupps, S, Figueroa-Cuilan, W, and Pintel, DJ (2017). Minute virus of canines NP1 protein governs the expression of a subset of essential nonstructural proteins via its role in RNA processing. *J Virol* **91**: e00260-17.
81. Liu, X, Hao, S, Chen, Z, Xu, H, Wang, H, Huang, M*, et al.* (2018). The 5' untranslated region of human bocavirus capsid transcripts regulates viral mRNA biogenesis and alternative translation. *J Virol* **92**: e00443-18.
82. Wang, Z, Shen, W, Cheng, F, Deng, X, Engelhardt, JF, Yan, Z*, et al.* (2017). Parvovirus expresses a small noncoding RNA that plays an essential role in virus replication. *J Virol* **91**: e02375-16.
83. Kestler, J, Neeb, B, Struyf, S, Van Damme, J, Cotmore, SF, D'Abramo, A*, et al.* (1999). Cis requirements for the efficient production of recombinant DNA vectors based on autonomous parvoviruses. *Hum Gene Ther* **10**: 1619-1632.
84. Palmer, GA, Brogdon, JL, Constant, SL, and Tattersall, P (2004). A nonproliferating parvovirus vaccine vector elicits sustained, protective humoral immunity following a single intravenous or intranasal inoculation. *J Virol* **78**: 1101-1108.
85. Haag, A, Menten, P, Van Damme, J, Dinsart, C, Rommelaere, J, and Cornelis, JJ (2000). Highly efficient transduction and expression of cytokine genes in human tumor cells by means of autonomous parvovirus vectors; generation of antitumor responses in recipient mice. *Hum Gene Ther* **11**: 597-609.
86. Giese, NA, Raykov, Z, DeMartino, L, Vecchi, A, Sozzani, S, Dinsart, C*, et al.* (2002). Suppression of metastatic hemangiosarcoma by a parvovirus MVMp vector transducing the IP-10 chemokine into immunocompetent mice. *Cancer Gene Ther* **9**: 432-442.
87. Maxwell, IH, Maxwell, F, Rhode, SL, 3rd, Corsini, J, and Carlson, JO (1993). Recombinant LuIII autonomous parvovirus as a transient transducing vector for human cells. *Hum Gene Ther* **4**: 441-450.
88. Spitzer, AL, Maxwell, F, Corsini, J, and Maxwell, IH (1996). Species specificity for transduction of cultured cells by a recombinant LuIII rodent parvovirus genome encapsidated by canine parvovirus or feline panleukopenia virus. *J Gen Virol* **77**: 1787-1792.
89. Spitzer, AL, Parrish, CR, and Maxwell, IH (1997). Tropic determinant for canine parvovirus and feline panleukopenia virus functions through the capsid protein VP2. *J Gen Virol* **78**: 925-928.
90. Maxwell, IH, Spitzer, AL, Long, CJ, and Maxwell, F (1996). Autonomous parvovirus transduction of a gene under control of tissue-specific or inducible promoters. *Gene Ther* **3**: 28-36.
91. David, RM, and Doherty, AT (2017). Viral vectors: the road to reducing genotoxicity. *Toxicol Sci* **155**: 315-325.
92. Bertran, J, Yang, Y, Hargrove, P, Vanin, EF, and Nienhuis, AW (1998). Targeted integration of a recombinant globin gene adeno-associated viral vector into human chromosome 19. *Ann N Y Acad Sci* **850**: 163-177.
93. Surosky, RT, Urabe, M, Godwin, SG, McQuiston, SA, Kurtzman, GJ, Ozawa, K*, et al.* (1997). Adeno-associated virus Rep proteins target DNA sequences to a unique locus in the human genome. *J Virol* **71**: 7951-7959.
94. Philpott, NJ, Giraud-Wali, C, Dupuis, C, Gomos, J, Hamilton, H, Berns, KI*, et al.* (2002). Efficient integration of recombinant adeno-associated virus DNA vectors requires a p5-rep sequence in cis. *J Virol* **76**: 5411-5421.
95. Henckaerts, E, Dutheil, N, Zeltner, N, Kattman, S, Kohlbrenner, E, Ward, P*, et al.* (2009). Site-specific integration of adeno-associated virus involves partial duplication of the target locus. *Proc Natl Acad Sci USA* **106**: 7571-7576.
96. Yang, Q, Chen, F, and Trempe, JP (1994). Characterization of cell lines that inducibly express the adeno-associated virus Rep proteins. *J Virol* **68**: 4847-4856.
97. Satoh, W, Hirai, Y, Tamayose, K, and Shimada, T (2000). Site-specific integration of an adeno-associated virus vector plasmid mediated by regulated expression of Rep based on Cre-loxP recombination. *J Virol* **74**: 10631-10638.

98. Ponnazhagan, S, Erikson, D, Kearns, WG, Zhou, SZ, Nahreini, P, Wang, XS*, et al.* (1997). Lack of site-specific integration of the recombinant adeno-associated virus 2 genomes in human cells. *Hum Gene Ther* **8**: 275-284.
99. Nakai, H, Wu, X, Fuess, S, Storm, TA, Munroe, D, Montini, E*, et al.* (2005). Large-scale molecular characterization of adeno-associated virus vector integration in mouse liver. *J Virol* **79**: 3606-3614.
100. Senis, E, Mosteiro, L, Wilkening, S, Wiedtke, E, Nowrouzi, A, Afzal, S*, et al.* (2018). AAV vector-mediated in vivo reprogramming into pluripotency. *Nat Commun* **9**: 2651.
101. Chandler, RJ, LaFave, MC, Varshney, GK, Trivedi, NS, Carrillo-Carrasco, N, Senac, JS*, et al.* (2015). Vector design influences hepatic genotoxicity after adeno-associated virus gene therapy. *J Clin Invest* **125**: 870-880.
102. Donsante, A, Miller, DG, Li, Y, Vogler, C, Brunt, EM, Russell, DW*, et al.* (2007). AAV vector integration sites in mouse hepatocellular carcinoma. *Science* **317**: 477.
103. Chandler, RJ, LaFave, MC, Varshney, GK, Burgess, SM, and Venditti, CP (2016). Genotoxicity in mice following AAV gene delivery: a safety concern for human gene therapy? *Mol Ther* **24**: 198-201.
104. Gil-Farina, I, Fronza, R, Kaeppel, C, Lopez-Franco, E, Ferreira, V, D'Avola, D*, et al.* (2016). Recombinant AAV integration is not associated with hepatic genotoxicity in nonhuman primates and patients. *Mol Ther* **24**: 1100-1105.
105. Kaeppel, C, Beattie, SG, Fronza, R, van Logtenstein, R, Salmon, F, Schmidt, S*, et al.* (2013). A largely random AAV integration profile after LPLD gene therapy. *Nat Med* **19**: 889-891.
106. Nathwani, AC, Rosales, C, McIntosh, J, Rastegarlari, G, Nathwani, D, Raj, D*, 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.
107. Penaud-Budloo, M, Le Guiner, C, Nowrouzi, A, Toromanoff, A, Cherel, Y, Chenuaud, P*, et al.* (2008). Adeno-associated virus vector genomes persist as episomal chromatin in primate muscle. *J Virol* **82**: 7875-7885.
108. Ward, P, Elias, P, and Linden, RM (2003). Rescue of the adeno-associated virus genome from a plasmid vector: evidence for rescue by replication. *J Virol* **77**: 11480-11490.
109. Chiorini, JA, Afione, S, and Kotin, RM (1999). Adeno-associated virus (AAV) type 5 Rep protein cleaves a unique terminal resolution site compared with other AAV serotypes. *J Virol* **73**: 4293-4298.
110. Boutin, S, Monteilhet, V, Veron, P, Leborgne, C, Benveniste, O, Montus, MF*, et al.* (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.
111. Grimm, D, Lee, JS, Wang, L, Desai, T, Akache, B, Storm, TA*, et al.* (2008). In vitro and in vivo gene therapy vector evolution via multispecies interbreeding and retargeting of adeno-associated viruses. *J Virol* **82**: 5887-5911.
112. Excoffon, KJ, Koerber, JT, Dickey, DD, Murtha, M, Keshavjee, S, Kaspar, BK*, et al.* (2009). Directed evolution of adeno-associated virus to an infectious respiratory virus. *Proc Natl Acad Sci USA* **106**: 3865-3870.
113. Adachi, K, and Nakai, H (2010). A new recombinant adeno-associated virus (Aav)-based random peptide display library system: infection-defective Aav1.9-3 as a novel detargeted platform for vector evolution. *Gene Ther Regul* **5**: 31-55.
114. Perabo, L, Buning, H, Kofler, DM, Ried, MU, Girod, A, Wendtner, CM*, et al.* (2003). In vitro selection of viral vectors with modified tropism: the adeno-associated virus display. *Mol Ther* **8**: 151-157.
115. Korbelin, J, Sieber, T, Michelfelder, S, Lunding, L, Spies, E, Hunger, A*, 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.

116. Deverman, BE, Pravdo, PL, Simpson, BP, Kumar, SR, Chan, KY, Banerjee, A*, et al.* (2016). Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. *Nat Biotechnol* **34**: 204-209.
117. Beck, SE, Jones, LA, Chesnut, K, Walsh, SM, Reynolds, TC, Carter, BJ*, et al.* (1999). Repeated delivery of adeno-associated virus vectors to the rabbit airway. *J Virol* **73**: 9446-9455.
118. Rabinowitz, JE, Rolling, F, Li, C, Conrath, H, Xiao, W, Xiao, X*, et al.* (2002). Cross-packaging of a single adeno-associated virus (AAV) type 2 vector genome into multiple AAV serotypes enables transduction with broad specificity. *J Virol* **76**: 791-801.
119. Grimm, D, Kay, MA, and Kleinschmidt, JA (2003). Helper virus-free, optically controllable, and two-plasmid-based production of adeno-associated virus vectors of serotypes 1 to 6. *Mol Ther* **7**: 839-850.
120. Rutledge, EA, Halbert, CL, and Russell, DW (1998). Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2. *J Virol* **72**: 309-319.
121. Chiorini, JA, Yang, L, Liu, Y, Safer, B, and Kotin, RM (1997). Cloning of adeno-associated virus type 4 (AAV4) and generation of recombinant AAV4 particles. *J Virol* **71**: 6823-6833.
122. Chiorini, JA, Kim, F, Yang, L, and Kotin, RM (1999). Cloning and characterization of adeno-associated virus type 5. *J Virol* **73**: 1309-1319.
123. Maxwell, IH, Long, CJ, Carlson, JO, Rhode, SL, 3rd, and Maxwell, F (1993). Encapsidation of a recombinant LuIII parvovirus genome by H1 virus and the fibrotropic or lymphotropic strains of minute virus of mice. *J Gen Virol* **74**: 1175-1179.
124. Wrzesinski, C, Tesfay, L, Salome, N, Jauniaux, JC, Rommelaere, J, Cornelis, J*, et al.* (2003). Chimeric and pseudotyped parvoviruses minimize the contamination of recombinant stocks with replication-competent viruses and identify a DNA sequence that restricts parvovirus H-1 in mouse cells. *J Virol* **77**: 3851-3858.
125. Corsini, J, Carlson, JO, Maxwell, F, and Maxwell, IH (1995). Symmetric-strand packaging of recombinant parvovirus LuIII genomes that retain only the terminal regions. *J Virol* **69**: 2692-2696.
126. Srivastava, CH, Samulski, RJ, Lu, L, Larsen, SH, and Srivastava, A (1989). Construction of a recombinant human parvovirus B19: adeno-associated virus 2 (AAV) DNA inverted terminal repeats are functional in an AAV-B19 hybrid virus. *Proc Natl Acad Sci USA* **86**: 8078-8082.
127. Ponnazhagan, S, Weigel, KA, Raikwar, SP, Mukherjee, P, Yoder, MC, and Srivastava, A (1998). Recombinant human parvovirus B19 vectors: erythroid cell-specific delivery and expression of transduced genes. *J Virol* **72**: 5224-5230.
128. Weigel-Kelley, KA, Yoder, MC, and Srivastava, A (2001). Recombinant human parvovirus B19 vectors: erythrocyte P antigen is necessary but not sufficient for successful transduction of human hematopoietic cells. *J Virol* **75**: 4110-4116.
129. Gallinella, G (2013). Parvovirus B19 Achievements and Challenges. *ISRN Virology* **2013**: 898730.
130. Wang, XS, Yoder, MC, Zhou, SZ, and Srivastava, A (1995). Parvovirus B19 promoter at map unit 6 confers autonomous replication competence and erythroid specificity to adeno-associated virus 2 in primary human hematopoietic progenitor cells. *Proc Natl Acad Sci USA* **92**: 12416-12420.
131. Kruger, L, Eskerski, H, Dinsart, C, Cornelis, J, Rommelaere, J, Haberkorn, U*, et al.* (2008). Augmented transgene expression in transformed cells using a parvoviral hybrid vector. *Cancer Gene Ther* **15**: 252-267.
132. Moolten, FL (1986). Tumor chemosensitivity conferred by inserted herpes thymidine kinase genes: paradigm for a prospective cancer control strategy. *Cancer Res* **46**: 5276-5281.
133. Fuks, F, Deleu, L, Dinsart, C, Rommelaere, J, and Faisst, S (1996). Ras oncogene-dependent activation of the P4 promoter of minute virus of mice through a proximal P4 element interacting with the Ets family of transcription factors. *J Virol* **70**: 1331-1339.

134. Li, J, Samulski, RJ, and Xiao, X (1997). Role for highly regulated rep gene expression in adeno-associated virus vector production. *J Virol* **71**: 5236-5243.
135. Grimm, D, Kern, A, Rittner, K, and Kleinschmidt, JA (1998). Novel tools for production and purification of recombinant adenoassociated virus vectors. *Hum Gene Ther* **9**: 2745-2760.
136. Gao, GP, Alvira, MR, Wang, L, Calcedo, R, Johnston, J, and Wilson, JM (2002). Novel adeno-associated viruses from rhesus monkeys as vectors for human gene therapy. *Proc Natl Acad Sci USA* **99**: 11854-11859.
137. Dubielzig, R, King, JA, Weger, S, Kern, A, and Kleinschmidt, JA (1999). Adeno-associated virus type 2 protein interactions: formation of pre-encapsidation complexes. *J Virol* **73**: 8989-8998.
138. Yan, Z, Zou, W, Feng, Z, Shen, W, Park, SY, Deng, X*, et al.* (2018). Establishment of a high yield rAAV/HBoV vector production system independent of Bocavirus non-structural proteins. *Hum Gene Ther*. doi: 10.1089/hum.2018.173.
139. Bennett, J (2017). Taking stock of retinal gene therapy: looking back and moving forward. *Mol Ther* **25**: 1076-1094.
140. Mendell, JR, Al-Zaidy, S, Shell, R, Arnold, WD, Rodino-Klapac, LR, Prior, TW*, et al.* (2017). Single-dose gene-replacement therapy for spinal muscular atrophy. *N Engl J Med* **377**: 1713-1722.
141. Mueller, C, Gernoux, G, Gruntman, AM, Borel, F, Reeves, EP, Calcedo, R*, et al.* (2017). 5 year expression and neutrophil defect repair after gene therapy in alpha-1 antitrypsin deficiency. *Mol Ther* **25**: 1387-1394.
142. George, LA, Sullivan, SK, Giermasz, A, Rasko, JEJ, Samelson-Jones, BJ, Ducore, J*, et al.* (2017). Hemophilia B gene therapy with a high-specific-activity factor IX variant. *N Engl J Med* **377**: 2215-2227.
143. US Food & Drug Administration (2017, December 19). FDA approves novel gene therapy to treat patients with a rare form of inherited vision loss. Retrieved from https://www.fda.gov/NewsEvents/Newsroom/PressAnnouncements/ucm589467.htm.
144. Al-Zaidy, S, Pickard, AS, Kotha, K, Alfano, LN, Lowes, L, Paul, G*, et al.* (2019). Health outcomes in spinal muscular atrophy type 1 following AVXS-101 gene replacement therapy. *Pediatr Pulmonol* **54**: 179-185.
145. Angelova, AL, Aprahamian, M, Balboni, G, Delecluse, HJ, Feederle, R, Kiprianova, I*, et al.* (2009). Oncolytic rat parvovirus H-1PV, a candidate for the treatment of human lymphoma: In vitro and in vivo studies. *Mol Ther* **17**: 1164-1172.
146. Geletneky, K, Hajda, J, Angelova, AL, Leuchs, B, Capper, D, Bartsch, AJ*, et al.* (2017). Oncolytic H-1 parvovirus shows safety and signs of immunogenic activity in a first phase I/IIa glioblastoma trial. *Mol Ther* **25**: 2620-2634.
147. Geletneky, K, Bartsch, A, Weiss, C, Bernhard, H, Marchini, A, Rommelaere, J*, et al.* (2018). ATIM-40. High rate of objective anti-tumor response in 9 patients with glioblastoma after viro-immunotherapy with oncolytic parvovirus H-1 in combination with bevacicumab and PD-1 checkpoint blockade. *Neuro-Oncology* **20**: vi10.
148. Stern, MJ, Ames, GF, Smith, NH, Robinson, EC, and Higgins, CF (1984). Repetitive extragenic palindromic sequences: a major component of the bacterial genome. *Cell* **37**: 1015-1026.
149. Ishino, Y, Shinagawa, H, Makino, K, Amemura, M, and Nakata, A (1987). Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol* **169**: 5429-5433.
150. Nakata, A, Amemura, M, and Makino, K (1989). Unusual nucleotide arrangement with repeated sequences in the Escherichia coli K-12 chromosome. *J Bacteriol* **171**: 3553-3556.
151. Hermans, PW, van Soolingen, D, Bik, EM, de Haas, PE, Dale, JW, and van Embden, JD (1991). Insertion element IS987 from Mycobacterium bovis BCG is located in a hot-spot integration region for insertion elements in Mycobacterium tuberculosis complex strains. *Infect Immun* **59**: 2695-2705.
152. Mojica, FJ, Ferrer, C, Juez, G, and Rodriguez-Valera, F (1995). Long stretches of short tandem repeats are present in the largest replicons of the Archaea Haloferax

mediterranei and Haloferax volcanii and could be involved in replicon partitioning. *Mol Microbiol* **17**: 85-93.
153. Jansen, R, van Embden, JD, Gaastra, W, and Schouls, LM (2002). Identification of a novel family of sequence repeats among prokaryotes. *OMICS* **6**: 23-33.
154. Jansen, R, Embden, JD, Gaastra, W, and Schouls, LM (2002). Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol* **43**: 1565-1575.
155. Mojica, FJ, Diez-Villasenor, C, Garcia-Martinez, J, and Soria, E (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* **60**: 174-182.
156. Bolotin, A, Quinquis, B, Sorokin, A, and Ehrlich, SD (2005). Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* **151**: 2551-2561.
157. Mojica, FJ, Diez-Villasenor, C, Garcia-Martinez, J, and Almendros, C (2009). Short motif sequences determine the targets of the prokaryotic CRISPR defence system. *Microbiology* **155**: 733-740.
158. Yosef, I, Goren, MG, and Qimron, U (2012). Proteins and DNA elements essential for the CRISPR adaptation process in Escherichia coli. *Nucleic Acids Res* **40**: 5569-5576.
159. Marraffini, LA, and Sontheimer, EJ (2010). Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature* **463**: 568-571.
160. Barrangou, R, Fremaux, C, Deveau, H, Richards, M, Boyaval, P, Moineau, S*, et al.* (2007). CRISPR provides acquired resistance against viruses in prokaryotes. *Science* **315**: 1709-1712.
161. Mohanraju, P, Makarova, KS, Zetsche, B, Zhang, F, Koonin, EV, and van der Oost, J (2016). Diverse evolutionary roots and mechanistic variations of the CRISPR-Cas systems. *Science* **353**: aad5147.
162. Wei, Y, Chesne, MT, Terns, RM, and Terns, MP (2015). Sequences spanning the leader-repeat junction mediate CRISPR adaptation to phage in Streptococcus thermophilus. *Nucleic Acids Res* **43**: 1749-1758.
163. Koonin, EV, and Krupovic, M (2015). Evolution of adaptive immunity from transposable elements combined with innate immune systems. *Nat Rev Genet* **16**: 184-192.
164. Fineran, PC, and Charpentier, E (2012). Memory of viral infections by CRISPR-Cas adaptive immune systems: acquisition of new information. *Virology* **434**: 202-209.
165. Datsenko, KA, Pougach, K, Tikhonov, A, Wanner, BL, Severinov, K, and Semenova, E (2012). Molecular memory of prior infections activates the CRISPR/Cas adaptive bacterial immunity system. *Nat Commun* **3**: 945.
166. Pougach, K, Semenova, E, Bogdanova, E, Datsenko, KA, Djordjevic, M, Wanner, BL*, et al.* (2010). Transcription, processing and function of CRISPR cassettes in Escherichia coli. *Mol Microbiol* **77**: 1367-1379.
167. Brouns, SJ, Jore, MM, Lundgren, M, Westra, ER, Slijkhuis, RJ, Snijders, AP*, et al.* (2008). Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science* **321**: 960-964.
168. Deltcheva, E, Chylinski, K, Sharma, CM, Gonzales, K, Chao, Y, Pirzada, ZA*, et al.* (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* **471**: 602-607.
169. Wiedenheft, B, Lander, GC, Zhou, K, Jore, MM, Brouns, SJJ, van der Oost, J*, et al.* (2011). Structures of the RNA-guided surveillance complex from a bacterial immune system. *Nature* **477**: 486-489.
170. Westra, ER, van Erp, PB, Kunne, T, Wong, SP, Staals, RH, Seegers, CL*, et al.* (2012). CRISPR immunity relies on the consecutive binding and degradation of negatively supercoiled invader DNA by Cascade and Cas3. *Mol Cell* **46**: 595-605.
171. Peng, W, Feng, M, Feng, X, Liang, YX, and She, Q (2015). An archaeal CRISPR type III-B system exhibiting distinctive RNA targeting features and mediating dual RNA and DNA interference. *Nucleic Acids Res* **43**: 406-417.
172. Louwen, R, Horst-Kreft, D, de Boer, AG, van der Graaf, L, de Knegt, G, Hamersma, M*, et al.* (2013). A novel link between Campylobacter jejuni bacteriophage defence, virulence and Guillain-Barre syndrome. *Eur J Clin Microbiol Infect Dis* **32**: 207-226.

173. Westra, ER, Buckling, A, and Fineran, PC (2014). CRISPR-Cas systems: beyond adaptive immunity. *Nat Rev Microbiol* **12**: 317-326.
174. Jurczak-Kurek, A, Gasior, T, Nejman-Falenczyk, B, Bloch, S, Dydecka, A, Topka, G*, et al.* (2016). Biodiversity of bacteriophages: morphological and biological properties of a large group of phages isolated from urban sewage. *Sci Rep* **6**: 34338.
175. Suttle, CA (2005). Viruses in the sea. *Nature* **437**: 356-361.
176. Seed, KD (2015). Battling phages: how bacteria defend against viral attack. *PLoS Pathog* **11**: e1004847.
177. Cumby, N, Edwards, AM, Davidson, AR, and Maxwell, KL (2012). The bacteriophage HK97 gp15 moron element encodes a novel superinfection exclusion protein. *J Bacteriol* **194**: 5012-5019.
178. Makarova, KS, Wolf, YI, Alkhnbashi, OS, Costa, F, Shah, SA, Saunders, SJ*, et al.* (2015). An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol* **13**: 722-736.
179. Bivalkar-Mehla, S, Vakharia, J, Mehla, R, Abreha, M, Kanwar, JR, Tikoo, A*, et al.* (2011). Viral RNA silencing suppressors (RSS): novel strategy of viruses to ablate the host RNA interference (RNAi) defense system. *Virus Res* **155**: 1-9.
180. Bondy-Denomy, J, Pawluk, A, Maxwell, KL, and Davidson, AR (2013). Bacteriophage genes that inactivate the CRISPR/Cas bacterial immune system. *Nature* **493**: 429-432.
181. Pawluk, A, Staals, RH, Taylor, C, Watson, BN, Saha, S, Fineran, PC*, et al.* (2016). Inactivation of CRISPR-Cas systems by anti-CRISPR proteins in diverse bacterial species. *Nat Microbiol* **1**: 16085.
182. Borges, AL, Davidson, AR, and Bondy-Denomy, J (2017). The discovery, mechanisms, and evolutionary impact of anti-CRISPRs. *Annu Rev Virol* **4**: 37-59.
183. Bondy-Denomy, J, Garcia, B, Strum, S, Du, M, Rollins, MF, Hidalgo-Reyes, Y*, et al.* (2015). Multiple mechanisms for CRISPR-Cas inhibition by anti-CRISPR proteins. *Nature* **526**: 136-139.
184. Vorontsova, D, Datsenko, KA, Medvedeva, S, Bondy-Denomy, J, Savitskaya, EE, Pougach, K*, et al.* (2015). Foreign DNA acquisition by the I-F CRISPR-Cas system requires all components of the interference machinery. *Nucleic Acids Res* **43**: 10848-10860.
185. Pawluk, A, Amrani, N, Zhang, Y, Garcia, B, Hidalgo-Reyes, Y, Lee, J*, et al.* (2016). Naturally occurring off-switches for CRISPR-Cas9. *Cell* **167**: 1829-1838 e1829.
186. Bubeck, F, Hoffmann, MD, Harteveld, Z, Aschenbrenner, S, Bietz, A, Waldhauer, MC*, et al.* (2018). Engineered anti-CRISPR proteins for optogenetic control of CRISPR-Cas9. *Nat Methods* **15**: 924-927.
187. Sapranauskas, R, Gasiunas, G, Fremaux, C, Barrangou, R, Horvath, P, and Siksnys, V (2011). The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Res* **39**: 9275-9282.
188. Cong, L, Ran, FA, Cox, D, Lin, S, Barretto, R, Habib, N*, et al.* (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**: 819-823.
189. Mali, P, Yang, L, Esvelt, KM, Aach, J, Guell, M, DiCarlo, JE*, et al.* (2013). RNA-guided human genome engineering via Cas9. *Science* **339**: 823-826.
190. Jinek, M, Chylinski, K, Fonfara, I, Hauer, M, Doudna, JA, and Charpentier, E (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**: 816-821.
191. Garneau, JE, Dupuis, ME, Villion, M, Romero, DA, Barrangou, R, Boyaval, P*, et al.* (2010). The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**: 67-71.
192. Gasiunas, G, Barrangou, R, Horvath, P, and Siksnys, V (2012). Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci USA* **109**: E2579-2586.
193. Maruyama, T, Dougan, SK, Truttmann, MC, Bilate, AM, Ingram, JR, and Ploegh, HL (2015). Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat Biotechnol* **33**: 538-542.

194. Li, W, Xu, H, Xiao, T, Cong, L, Love, MI, Zhang, F*, et al.* (2014). MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol* **15**: 554.
195. Chen, S, Sanjana, NE, Zheng, K, Shalem, O, Lee, K, Shi, X*, et al.* (2015). Genome-wide CRISPR screen in a mouse model of tumor growth and metastasis. *Cell* **160**: 1246-1260.
196. Song, J, Yang, D, Xu, J, Zhu, T, Chen, YE, and Zhang, J (2016). RS-1 enhances CRISPR/Cas9- and TALEN-mediated knock-in efficiency. *Nat Commun* **7**: 10548.
197. Savic, N, Ringnalda, FC, Lindsay, H, Berk, C, Bargsten, K, Li, Y*, et al.* (2018). Covalent linkage of the DNA repair template to the CRISPR-Cas9 nuclease enhances homology-directed repair. *Elife* **7**: e33761.
198. Mali, P, Aach, J, Stranges, PB, Esvelt, KM, Moosburner, M, Kosuri, S*, et al.* (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. *Nat Biotechnol* **31**: 833-838.
199. Bikard, D, Jiang, W, Samai, P, Hochschild, A, Zhang, F, and Marraffini, LA (2013). Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Res* **41**: 7429-7437.
200. Yin, H, Xue, W, Chen, S, Bogorad, RL, Benedetti, E, Grompe, M*, et al.* (2014). Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. *Nat Biotechnol* **32**: 551-553.
201. Guan, Y, Ma, Y, Li, Q, Sun, Z, Ma, L, Wu, L*, et al.* (2016). CRISPR/Cas9-mediated somatic correction of a novel coagulator factor IX gene mutation ameliorates hemophilia in mouse. *EMBO Mol Med* **8**: 477-488.
202. Bonamassa, B, Hai, L, and Liu, D (2011). Hydrodynamic gene delivery and its applications in pharmaceutical research. *Pharm Res* **28**: 694-701.
203. Suda, T, and Liu, D (2007). Hydrodynamic gene delivery: its principles and applications. *Mol Ther* **15**: 2063-2069.
204. Zuris, JA, Thompson, DB, Shu, Y, Guilinger, JP, Bessen, JL, Hu, JH*, et al.* (2015). Cationic lipid-mediated delivery of proteins enables efficient protein-based genome editing in vitro and in vivo. *Nat Biotechnol* **33**: 73-80.
205. Lee, K, Conboy, M, Park, HM, Jiang, F, Kim, HJ, Dewitt, MA*, et al.* (2017). Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair. *Nat Biomed Eng* **1**: 889-901.
206. Kang, YK, Kwon, K, Ryu, JS, Lee, HN, Park, C, and Chung, HJ (2017). Nonviral genome editing based on a polymer-derivatized CRISPR nanocomplex for targeting bacterial pathogens and antibiotic resistance. *Bioconjug Chem* **28**: 957-967.
207. Cheng, R, Peng, J, Yan, Y, Cao, P, Wang, J, Qiu, C*, et al.* (2014). Efficient gene editing in adult mouse livers via adenoviral delivery of CRISPR/Cas9. *FEBS Lett* **588**: 3954-3958.
208. Holmgaard, A, Askou, AL, Benckendorff, JNE, Thomsen, EA, Cai, Y, Bek, T*, et al.* (2017). In vivo knockout of the Vegfa gene by lentiviral delivery of CRISPR/Cas9 in mouse retinal pigment epithelium cells. *Mol Ther Nucleic Acids* **9**: 89-99.
209. Senis, E, Fatouros, C, Grosse, S, Wiedtke, E, Niopek, D, Mueller, AK*, et al.* (2014). CRISPR/Cas9-mediated genome engineering: an adeno-associated viral (AAV) vector toolbox. *Biotechnol J* **9**: 1402-1412.
210. Kim, E, Koo, T, Park, SW, Kim, D, Kim, K, Cho, HY*, et al.* (2017). In vivo genome editing with a small Cas9 orthologue derived from Campylobacter jejuni. *Nat Commun* **8**: 14500.
211. Esvelt, KM, Mali, P, Braff, JL, Moosburner, M, Yaung, SJ, and Church, GM (2013). Orthogonal Cas9 proteins for RNA-guided gene regulation and editing. *Nat Methods* **10**: 1116-1121.
212. Tabebordbar, M, Zhu, K, Cheng, JKW, Chew, WL, Widrick, JJ, Yan, WX*, et al.* (2016). In vivo gene editing in dystrophic mouse muscle and muscle stem cells. *Science* **351**: 407-411.

213. Truong, DJ, Kuhner, K, Kuhn, R, Werfel, S, Engelhardt, S, Wurst, W*, et al.* (2015). Development of an intein-mediated split-Cas9 system for gene therapy. *Nucleic Acids Res* **43**: 6450-6458.
214. Duan, D, Yue, Y, and Engelhardt, JF (2001). Expanding AAV packaging capacity with trans-splicing or overlapping vectors: a quantitative comparison. *Mol Ther* **4**: 383-391.
215. Wright, AV, Sternberg, SH, Taylor, DW, Staahl, BT, Bardales, JA, Kornfeld, JE*, et al.* (2015). Rational design of a split-Cas9 enzyme complex. *Proc Natl Acad Sci USA* **112**: 2984-2989.
216. Zetsche, B, Volz, SE, and Zhang, F (2015). A split-Cas9 architecture for inducible genome editing and transcription modulation. *Nat Biotechnol* **33**: 139-142.
217. Nishimasu, H, Cong, L, Yan, WX, Ran, FA, Zetsche, B, Li, Y*, et al.* (2015). Crystal structure of Staphylococcus aureus Cas9. *Cell* **162**: 1113-1126.
218. Maude, SL, Frey, N, Shaw, PA, Aplenc, R, Barrett, DM, Bunin, NJ*, et al.* (2014). Chimeric antigen receptor T cells for sustained remissions in leukemia. *N Engl J Med* **371**: 1507-1517.
219. Brentjens, RJ, Davila, ML, Riviere, I, Park, J, Wang, X, Cowell, LG*, et al.* (2013). CD19-targeted T cells rapidly induce molecular remissions in adults with chemotherapy-refractory acute lymphoblastic leukemia. *Sci Transl Med* **5**: 177ra138.
220. Zhang, C, Liu, J, Zhong, JF, and Zhang, X (2017). Engineering CAR-T cells. *Biomark Res* **5**: 22.
221. US Food & Drug Administration (2018, May 03). FDA approves tisagenlecleucel for adults with relapsed or refractory large B-cell lymphoma. Retrieved from https://www.fda.gov/Drugs/InformationOnDrugs/ApprovedDrugs/ucm606540.htm.
222. MacLeod, DT, Antony, J, Martin, AJ, Moser, RJ, Hekele, A, Wetzel, KJ*, et al.* (2017). Integration of a CD19 CAR into the TCR alpha chain locus streamlines production of allogeneic gene-edited CAR T Cells. *Mol Ther* **25**: 949-961.
223. Zakrzewski, JL, Suh, D, Markley, JC, Smith, OM, King, C, Goldberg, GL*, et al.* (2008). Tumor immunotherapy across MHC barriers using allogeneic T-cell precursors. *Nat Biotechnol* **26**: 453-461.
224. Yoon, DH, Osborn, MJ, Tolar, J, and Kim, CJ (2018). Incorporation of immune checkpoint blockade into chimeric antigen receptor T Cells (CAR-Ts): combination or built-in CAR-T. *Int J Mol Sci* **19**: E340.
225. Sheridan, C (2018, December 14). Go-ahead for first in-body CRISPR medicine testing - A CRISPR-Cas9-based therapeutic from Editas is poised to enter clinical trials for treating blindness. *Nat Biotechnol*. doi: 10.1038/d41587-018-00003-2.
226. McCarty, DM, Monahan, PE, and Samulski, RJ (2001). Self-complementary recombinant adeno-associated virus (scAAV) vectors promote efficient transduction independently of DNA synthesis. *Gene Ther* **8**: 1248-1254.
227. Pear, WS, Nolan, GP, Scott, ML, and Baltimore, D (1993). Production of high-titer helper-free retroviruses by transient transfection. *Proc Natl Acad Sci USA* **90**: 8392-8396.
228. Lieber, M, Mazzetta, J, Nelson-Rees, W, Kaplan, M, and Todaro, G (1975). Establishment of a continuous tumor-cell line (panc-1) from a human carcinoma of the exocrine pancreas. *Int J Cancer* **15**: 741-747.
229. Gey, G, Coffman, W, and Kubicek, M (1952). Tissue culture studies of the proliferative capacity of cervical carcinoma and normal epithelium. *Cancer research* **12**: 264–265.
230. Nakabayashi, H, Taketa, K, Miyano, K, Yamane, T, and Sato, J (1982). Growth of human hepatoma cells lines with differentiated functions in chemically defined medium. *Cancer research* **42**: 3858-3863.
231. Raschke, WC, Baird, S, Ralph, P, and Nakoinz, I (1978). Functional macrophage cell lines transformed by Abelson leukemia virus. *Cell* **15**: 261-267.
232. Xu, L, Hui, AY, Albanis, E, Arthur, MJ, O'Byrne, SM, Blaner, WS*, et al.* (2005). Human hepatic stellate cell lines, LX-1 and LX-2: new tools for analysis of hepatic fibrosis. *Gut* **54**: 142-151.

233. Soule, HD, Vazguez, J, Long, A, Albert, S, and Brennan, M (1973). A human cell line from a pleural effusion derived from a breast carcinoma. *J Natl Cancer Inst* **51**: 1409-1416.
234. Dharmsathaphorn, K, McRoberts, JA, Mandel, KG, Tisdale, LD, and Masui, H (1984). A human colonic tumor cell line that maintains vectorial electrolyte transport. *Am J Physiol* **246**: G204-208.
235. Zabner, J, Karp, P, Seiler, M, Phillips, SL, Mitchell, CJ, Saavedra, M*, et al.* (2003). Development of cystic fibrosis and noncystic fibrosis airway cell lines. *Am J Physiol Lung Cell Mol Physiol* **284**: L844-854.
236. Chojnacki, S, Cowley, A, Lee, J, Foix, A, and Lopez, R (2017). Programmatic access to bioinformatics tools from EMBL-EBI update: 2017. *Nucleic Acids Res* **45**: W550-W553.
237. Schneider, CA, Rasband, WS, and Eliceiri, KW (2012). NIH Image to ImageJ: 25 years of image analysis. *Nature methods* **9**: 671-675.
238. Kumar, S, Stecher, G, Li, M, Knyaz, C, and Tamura, K (2018). MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol* **35**: 1547-1549.
239. Schurmann, N, Trabuco, LG, Bender, C, Russell, RB, and Grimm, D (2013). Molecular dissection of human Argonaute proteins by DNA shuffling. *Nat Struct Mol Biol* **20**: 818-826.
240. Samulski, RJ, Berns, KI, 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 USA* **79**: 2077-2081.
241. Grimm, D, Streetz, KL, Jopling, CL, Storm, TA, Pandey, K, Davis, CR*, et al.* (2006). Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature* **441**: 537-541.
242. Chen, B, Gilbert, LA, Cimini, BA, Schnitzbauer, J, Zhang, W, Li, GW*, et al.* (2013). Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* **155**: 1479-1491.
243. Shalem, O, Sanjana, NE, Hartenian, E, Shi, X, Scott, DA, Mikkelson, T*, et al.* (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**: 84-87.
244. Adikusuma, F, Pfitzner, C, and Thomas, PQ (2017). Versatile single-step-assembly CRISPR/Cas9 vectors for dual gRNA expression. *PLoS One* **12**: e0187236.
245. Xie, J, Mao, Q, Tai, PWL, He, R, Ai, J, Su, Q*, et al.* (2017). Short DNA hairpins compromise recombinant adeno-associated virus genome homogeneity. *Mol Ther* **25**: 1363-1374.
246. Thermo Fisher scientific, Tm calculator for Phusion or Phire DNA polymerase: https://www.thermofisher.com/de/de/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html.
247. Allawi, HT, and SantaLucia, J, Jr. (1997). Thermodynamics and NMR of internal G.T mismatches in DNA. *Biochemistry* **36**: 10581-10594.
248. Samulski, RJ, Srivastava, A, Berns, KI, and Muzyczka, N (1983). Rescue of adeno-associated virus from recombinant plasmids: gene correction within the terminal repeats of AAV. *Cell* **33**: 135-143.
249. Bardelli, M, Zarate-Perez, F, Agundez, L, Jolinon, N, Michael Linden, R, Escalante, CR*, et al.* (2017). Analysis of replicative intermediates of adeno-associated virus through Hirt extraction and Southern blotting. *Bio Protoc* **7**: e2271.
250. Sachs, N, Papaspyropoulos, A, Zomer-van Ommen, DD, Heo, I, Bottinger, L, Klay, D*, et al.* (2019). Long-term expanding human airway organoids for disease modeling. *EMBO J* **38**: e100300.
251. Wu, J, Zhao, W, Zhong, L, Han, Z, Li, B, Ma, W*, et al.* (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.

252. Fakhiri, J, Schneider, MA, Puschhof, J, Stanifer, M, Schildgen, V, Holderbach, S*, et al.* (2019). Novel chimeric gene therapy vectors based on adeno-associated virus and four different mammalian bocaviruses. *Mol Ther Methods Clin Dev* **12**: 202-222.
253. Brandenburger, A, Coessens, E, El Bakkouri, K, and Velu, T (1999). Influence of sequence and size of DNA on packaging efficiency of parvovirus MVM-based vectors. *Hum Gene Ther* **10**: 1229-1238.
254. Grieger, JC, and Samulski, RJ (2005). Packaging capacity of adeno-associated virus serotypes: impact of larger genomes on infectivity and postentry steps. *J Virol* **79**: 9933-9944.
255. Dong, JY, Fan, PD, and Frizzell, RA (1996). Quantitative analysis of the packaging capacity of recombinant adeno-associated virus. *Hum Gene Ther* **7**: 2101-2112.
256. Wu, Z, Yang, H, and Colosi, P (2010). Effect of genome size on AAV vector packaging. *Mol Ther* **18**: 80-86.
257. Yan, Z, Zak, R, Luxton, GW, Ritchie, TC, Bantel-Schaal, U, and Engelhardt, JF (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.
258. Zhong, L, Li, B, Mah, CS, Govindasamy, L, Agbandje-McKenna, M, Cooper, M*, et al.* (2008). Next generation of adeno-associated virus 2 vectors: point mutations in tyrosines lead to high-efficiency transduction at lower doses. *Proc Natl Acad Sci USA* **105**: 7827-7832.
259. Salganik, M, Aydemir, F, Nam, HJ, McKenna, R, Agbandje-McKenna, M, and Muzyczka, N (2014). Adeno-associated virus capsid proteins may play a role in transcription and second-strand synthesis of recombinant genomes. *J Virol* **88**: 1071-1079.
260. Gurda, BL, Parent, KN, Bladek, H, Sinkovits, RS, DiMattia, MA, Rence, C*, et al.* (2010). Human bocavirus capsid structure: insights into the structural repertoire of the parvoviridae. *J Virol* **84**: 5880-5889.
261. Yan, Z, Feng, Z, Sun, X, Zhang, Y, Zou, W, Wang, Z*, et al.* (2017). Human bocavirus type-1 capsid facilitates the transduction of ferret airways by adeno-associated virus genomes. *Hum Gene Ther* **28**: 612-625.
262. Kapoor, A, Simmonds, P, Slikas, E, Li, L, Bodhidatta, L, Sethabutr, O*, et al.* (2010). Human bocaviruses are highly diverse, dispersed, recombination prone, and prevalent in enteric infections. *J Infect Dis* **201**: 1633-1643.
263. Tamura, K, and Nei, M (1993). Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol Biol Evol* **10**: 512-526.
264. Li, L, Pesavento, PA, Leutenegger, CM, Estrada, M, Coffey, LL, Naccache, SN*, et al.* (2013). A novel bocavirus in canine liver. *Virol J* **10**: 54.
265. Engelhardt, J, Yan, Z, and Qiu, J (2017). Cis and trans requirements for terminal resolution of Human Bocavirus 1, WO patent application *WO2017205739A1*, published 30 November 2017.
266. Ayuso, E, Mingozzi, F, Montane, J, Leon, X, Anguela, XM, Haurigot, V*, et al.* (2010). High AAV vector purity results in serotype- and tissue-independent enhancement of transduction efficiency. *Gene Ther* **17**: 503-510.
267. Grieger, JC, Choi, VW, and Samulski, RJ (2006). Production and characterization of adeno-associated viral vectors. *Nat Protoc* **1**: 1412-1428.
268. Strobel, B, Miller, FD, Rist, W, and Lamla, T (2015). Comparative analysis of cesium chloride- and iodixanol-based purification of recombinant adeno-associated viral vectors for preclinical applications. *Hum Gene Ther Methods* **26**: 147-157.
269. Gao, K, Li, M, Zhong, L, Su, Q, Li, J, Li, S*, et al.* (2014). Empty virions in AAV8 vector preparations reduce transduction efficiency and may cause total viral particle dose-limiting side-effects. *Mol Ther Methods Clin Dev* **1**: 20139.
270. Fountaine, H, Harnish, P, Andrew, E, and Grynne, B (1996). Safety, tolerance, and pharmacokinetics of iodixanol injection, a nonionic, isosmolar, hexa-iodinated contrast agent. *Acad Radiol* **3**: S475-484.

271. Khan, IF, Hirata, RK, and Russell, DW (2011). AAV-mediated gene targeting methods for human cells. *Nat Protoc* **6**: 482-501.
272. Vironova. Characterization of AAV samples using TEM. *Mini TEM automated imaging and image analysis*. Retrieved from https://www.vironova.com/characterization-of-aav-samples-using-tem/.
273. Ikai, A, Nishigai, M, Tanaka, K, and Ichihara, A (1991). Electron microscopy of 26 S complex containing 20 S proteasome. *FEBS Lett* **292**: 21-24.
274. Yan, Z, Zak, R, Zhang, Y, Ding, W, Godwin, S, Munson, K*, et al.* (2004). Distinct classes of proteasome-modulating agents cooperatively augment recombinant adeno-associated virus type 2 and type 5-mediated transduction from the apical surfaces of human airway epithelia. *J Virol* **78**: 2863-2874.
275. Barkauskas, CE, Chung, MI, Fioret, B, Gao, X, Katsura, H, and Hogan, BL (2017). Lung organoids: current uses and future promise. *Development* **144**: 986-997.
276. Drost, J, and Clevers, H (2018). Organoids in cancer research. *Nat Rev Cancer* **18**: 407-418.
277. Rock, JR, Randell, SH, and Hogan, BL (2010). Airway basal stem cells: a perspective on their roles in epithelial homeostasis and remodeling. *Dis Model Mech* **3**: 545-556.
278. Plopper, CG, Nishio, SJ, Alley, JL, Kass, P, and Hyde, DM (1992). The role of the nonciliated bronchiolar epithelial (Clara) cell as the progenitor cell during bronchiolar epithelial differentiation in the perinatal rabbit lung. *Am J Respir Cell Mol Biol* **7**: 606-613.
279. Kantola, K, Hedman, L, Arthur, J, Alibeto, A, Delwart, E, Jartti, T*, et al.* (2011). Seroepidemiology of human bocaviruses 1-4. *J Infect Dis* **204**: 1403-1412.
280. Meliani, A, Leborgne, C, Triffault, S, Jeanson-Leh, L, Veron, P, and Mingozzi, F (2015). Determination of anti-adeno-associated virus vector neutralizing antibody titer with an in vitro reporter system. *Hum Gene Ther Methods* **26**: 45-53.
281. Tabatabai, J, Fakhiri, J, Meyburg, J, Linse, KP, Xu, M, Soderlund-Venermo, M*, et al.* (2019). Severe human bocavirus 1 respiratory tract infection in an immunodeficient child with fatal outcome. *Pediatr Infect Dis J*. doi: 10.1097/INF.0000000000002354.
282. Dupuis-Maguiraga, L, Noret, M, Brun, S, Le Grand, R, Gras, G, and Roques, P (2012). Chikungunya disease: infection-associated markers from the acute to the chronic phase of arbovirus-induced arthralgia. *PLoS Negl Trop Dis* **6**: e1446.
283. Zhang, J, and Crumpacker, C (2015). Hematopoietic stem and immune cells in chronic HIV infection. *Stem Cells Int* **2015**: 148064.
284. Bonvicini, F, Manaresi, E, Gentilomi, GA, Di Furio, F, Zerbini, M, Musiani, M*, et al.* (2011). Evidence of human bocavirus viremia in healthy blood donors. *Diagn Microbiol Infect Dis* **71**: 460-462.
285. Karbalaie Niya, MH, Ajdarkosh, H, Safarnezhad Tameshkel, F, Panahi, M, Tabasi, M, Bouzari, B*, et al.* (2018). The molecular detection of human bocavirus (HBoV) in colorectal tissue with malignant and non-malignant lesions. *Asian Pac J Cancer Prev* **19**: 3295-3299.
286. Abdel-Moneim, AS, El-Fol, HA, Kamel, MM, Soliman, AS, Mahdi, EA, El-Gammal, AS*, et al.* (2016). Screening of human bocavirus in surgically excised cancer specimens. *Arch Virol* **161**: 2095-2102.
287. Schildgen, V, Longo, Y, Pieper, M, and Schildgen, O (2018). T84 air-liquid interface cultures enable isolation of human bocavirus. *Influenza Other Respir Viruses*. doi: 10.1111/irv.12567.
288. Ghietto, LM, Toigo D'Angelo, AP, Viale, FA, and Adamo, MP (2017). Human bocavirus 1 infection of CACO-2 cell line cultures. *Virology* **510**: 273-280.
289. Kapoor, A, Hornig, M, Asokan, A, Williams, B, Henriquez, JA, and Lipkin, WI (2011). Bocavirus episome in infected human tissue contains non-identical termini. *PLoS One* **6**: e21362.
290. Yamaoka, S, Okada, K, Ito, N, Okadera, K, Mitake, H, Nakagawa, K*, et al.* (2017). Defect of rabies virus phosphoprotein in its interferon-antagonist activity negatively affects viral replication in muscle cells. *J Vet Med Sci* **79**: 1394-1397.

291. Desdouits, M, Munier, S, Prevost, MC, Jeannin, P, Butler-Browne, G, Ozden, S*, et al.* (2013). Productive infection of human skeletal muscle cells by pandemic and seasonal influenza A (H1N1) viruses. *PLoS One* **8**: e79628.
292. Brebion, A, Vanlieferinghen, P, Déchelotte, P, Boutry, M, Peigue-Lafeuille, H, and Henquell, C (2014). Fatal subacute myocarditis associated with human bocavirus 2 in a 13-month-old child. *J Clin Microbiol* **52**: 1006-1008.
293. Kuethe, F, Lindner, J, Matschke, K, Wenzel, JJ, Norja, P, Ploetze, K*, et al.* (2009). Prevalence of parvovirus B19 and human bocavirus DNA in the heart of patients with no evidence of dilated cardiomyopathy or myocarditis. *Clin Infect Dis* **49**: 1660-1666.
294. White, ES (2015). Lung extracellular matrix and fibroblast function. *Ann Am Thorac Soc* **12**: S30-33.
295. Friedman, HM, Macarak, EJ, MacGregor, RR, Wolfe, J, and Kefalides, NA (1981). Virus infection of endothelial cells. *J Infect Dis* **143**: 266-273.
296. Hansen, M, Brockmann, M, Schildgen, V, and Schildgen, O (2019). Human bocavirus is detected in human placenta and aborted tissues. *Influenza Other Respir Viruses* **13**: 106-109.
297. Koerber, JT, Jang, JH, and Schaffer, DV (2008). DNA shuffling of adeno-associated virus yields functionally diverse viral progeny. *Mol Ther* **16**: 1703-1709.
298. Perabo, L, Endell, J, King, S, Lux, K, Goldnau, D, Hallek, M*, et al.* (2006). Combinatorial engineering of a gene therapy vector: directed evolution of adeno-associated virus. *J Gene Med* **8**: 155-162.
299. 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.
300. Herrmann, AK, Bender, C, Kienle, E, Grosse, S, El Andari, J, Botta, J*, et al.* (2018). A robust and all-inclusive pipeline for shuffling of Adeno-associated viruses (AAV). *ACS Synth Biol* **8**: 194-206.
301. Qu, XW, Liu, WP, Qi, ZY, Duan, ZJ, Zheng, LS, Kuang, ZZ*, et al.* (2008). Phospholipase A2-like activity of human bocavirus VP1 unique region. *Biochem Biophys Res Commun* **365**: 158-163.
302. Mietzsch, M, Kailasan, S, Garrison, J, Ilyas, M, Chipman, P, Kantola, K*, et al.* (2017). Structural insights into human bocaparvoviruses. *J Virol* **91**: e00261-17.
303. Petris, G, Casini, A, Montagna, C, Lorenzin, F, Prandi, D, Romanel, A*, et al.* (2017). Hit and go CAS9 delivered through a lentiviral based self-limiting circuit. *Nat Commun* **8**: 15334.
304. Chew, WL, Tabebordbar, M, Cheng, JK, Mali, P, Wu, EY, Ng, AH*, et al.* (2016). A multifunctional AAV-CRISPR-Cas9 and its host response. *Nat Methods* **13**: 868-874.
305. Grimm, D, Pandey, K, Nakai, H, Storm, TA, and Kay, MA (2006). Liver transduction with recombinant adeno-associated virus is primarily restricted by capsid serotype not vector genotype. *J Virol* **80**: 426-439.
306. Schuster, SJ, Svoboda, J, Chong, EA, Nasta, SD, Mato, AR, Anak, O*, et al.* (2017). Chimeric antigen receptor T Cells in refractory B-Cell lymphomas. *N Engl J Med* **377**: 2545-2554.
307. Bennett, J, Wellman, J, Marshall, KA, McCague, S, Ashtari, M, DiStefano-Pappas, J*, et al.* (2016). Safety and durability of effect of contralateral-eye administration of AAV2 gene therapy in patients with childhood-onset blindness caused by RPE65 mutations: a follow-on phase 1 trial. *Lancet* **388**: 661-672.
308. Gao, G (2019, May 09). The case for growth. Retrieved from https://www.asgct.org/research/news/may-2019/the-case-for-growth.
309. Cross, R (2019, May 02). Crowds swarm gene-therapy conference as excitement hits an all-time high. c&en **97**. Retrieved from https://cen.acs.org/business/Crowds-swarm-gene-therapy-conference/97/i18.
310. Fisher, KJ, Gao, GP, Weitzman, MD, DeMatteo, R, Burda, JF, and Wilson, JM (1996). Transduction with recombinant adeno-associated virus for gene therapy is limited by leading-strand synthesis. *J Virol* **70**: 520-532.

311. Ferrari, FK, Samulski, T, Shenk, T, and Samulski, RJ (1996). Second-strand synthesis is a rate-limiting step for efficient transduction by recombinant adeno-associated virus vectors. *J Virol* **70**: 3227-3234.
312. Tullis, GE, and Shenk, T (2000). Efficient replication of adeno-associated virus type 2 vectors: a cis-acting element outside of the terminal repeats and a minimal size. *J Virol* **74**: 11511-11521.
313. Herrmann, AK, Grosse, S, Borner, K, Kramer, C, Wiedtke, E, Gunkel, M*, et al.* (2019). Impact of the assembly-activating protein on molecular evolution of synthetic adeno-associated virus capsids. *Hum Gene Ther* **30**: 21-35.
314. Grosse, S, Penaud-Budloo, M, Herrmann, AK, Borner, K, Fakhiri, J, Laketa, V*, 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**: e01198-17.
315. Ling, C, Wang, Y, Lu, Y, Wang, L, Jayandharan, GR, Aslanidi, GV*, et al.* (2015). The adeno-associated virus genome packaging puzzle. *J Mol Genet Med* **9**: 175.
316. Jartti, T, Hedman, K, Jartti, L, Ruuskanen, O, Allander, T, and Soderlund-Venermo, M (2012). Human bocavirus-the first 5 years. *Rev Med Virol* **22**: 46-64.
317. Guido, M, Tumolo, MR, Verri, T, Romano, A, Serio, F, De Giorgi, M*, et al.* (2016). Human bocavirus: current knowledge and future challenges. *World J Gastroenterol* **22**: 8684-8697.
318. Ros, C, Burckhardt, CJ, and Kempf, C (2002). Cytoplasmic trafficking of minute virus of mice: low-pH requirement, routing to late endosomes, and proteasome interaction. *J Virol* **76**: 12634-12645.
319. Douar, AM, Poulard, K, Stockholm, D, and Danos, O (2001). Intracellular trafficking of adeno-associated virus vectors: routing to the late endosomal compartment and proteasome degradation. *J Virol* **75**: 1824-1833.
320. Duan, D, Yue, Y, Yan, Z, Yang, J, and Engelhardt, JF (2000). Endosomal processing limits gene transfer to polarized airway epithelia by adeno-associated virus. *J Clin Invest* **105**: 1573-1587.
321. Monahan, PE, Lothrop, CD, Sun, J, Hirsch, ML, Kafri, T, Kantor, B*, et al.* (2010). Proteasome inhibitors enhance gene delivery by AAV virus vectors expressing large genomes in hemophilia mouse and dog models: a strategy for broad clinical application. *Mol Ther* **18**: 1907-1916.
322. Zhong, L, Li, B, Jayandharan, G, Mah, CS, Govindasamy, L, Agbandje-McKenna, M*, et al.* (2008). Tyrosine-phosphorylation of AAV2 vectors and its consequences on viral intracellular trafficking and transgene expression. *Virology* **381**: 194-202.
323. DiPrimio, N, Asokan, A, Govindasamy, L, Agbandje-McKenna, M, and Samulski, RJ (2008). Surface loop dynamics in adeno-associated virus capsid assembly. *J Virol* **82**: 5178-5189.
324. Mao, Y, Wang, X, Yan, R, Hu, W, Li, A, Wang, S*, et al.* (2016). Single point mutation in adeno-associated viral vectors-DJ capsid leads to improvement for gene delivery in vivo. *BMC Biotechnol* **16**: 1.
325. Gabriel, N, Hareendran, S, Sen, D, Gadkari, RA, Sudha, G, Selot, R*, et al.* (2013). Bioengineering of AAV2 capsid at specific serine, threonine, or lysine residues improves its transduction efficiency in vitro and in vivo. *Hum Gene Ther Methods* **24**: 80-93.
326. Zhang, Z, Zheng, Z, Luo, H, Meng, J, Li, H, Li, Q*, et al.* (2012). Human bocavirus NP1 inhibits IFN-β production by blocking association of IFN regulatory factor 3 with IFNB promoter. *J Immunol* **189**: 1144-1153.
327. Xu, D, Brumm, K, and Zhang, L (2006). The latent membrane protein 1 of Epstein-Barr virus (EBV) primes EBV latency cells for type I interferon production. *J Biol Chem* **281**: 9163-9169.
328. McCarthy, MK, and Weinberg, JB (2015). The immunoproteasome and viral infection: a complex regulator of inflammation. *Front Microbiol* **6**: 21.

329. Li, W, Asokan, A, Wu, Z, Van Dyke, T, DiPrimio, N, Johnson, JS*, et al.* (2008). Engineering and selection of shuffled AAV genomes: a new strategy for producing targeted biological nanoparticles. *Mol Ther* **16**: 1252-1260.
330. Joern, JM, Meinhold, P, and Arnold, FH (2002). Analysis of shuffled gene libraries. *J Mol Biol* **316**: 643-656.
331. Stemmer, WP (1994). DNA shuffling by random fragmentation and reassembly: in vitro recombination for molecular evolution. *Proc Natl Acad Sci USA* **91**: 10747-10751.
332. Ness, JE, Welch, M, Giver, L, Bueno, M, Cherry, JR, Borchert, TV*, et al.* (1999). DNA shuffling of subgenomic sequences of subtilisin. *Nat Biotechnol* **17**: 893-896.
333. Nonnenmacher, M, and Weber, T (2012). Intracellular transport of recombinant adeno-associated virus vectors. *Gene Ther* **19**: 649-658.
334. Kailasan, S, Halder, S, Gurda, B, Bladek, H, Chipman, PR, McKenna, R*, et al.* (2015). Structure of an enteric pathogen, bovine parvovirus. *J Virol* **89**: 2603-2614.
335. Li, J, Yang, Y, Dong, Y, Li, Y, Huang, Y, Yi, Q*, et al.* (2013). Key elements of the human bocavirus type 1 (HBoV1) promoter and its trans-activation by NS1 protein. *Virol J* **10**: 315.
336. Mingozzi, F, and High, KA (2013). Immune responses to AAV vectors: overcoming barriers to successful gene therapy. *Blood* **122**: 23-36.
337. Hinderer, C, Katz, N, Buza, EL, Dyer, C, Goode, T, Bell, P*, et al.* (2018). Severe toxicity in nonhuman primates and piglets following high-dose intravenous administration of an adeno-associated virus vector expressing human SMN. *Hum Gene Ther* **29**: 285-298.
338. Miyamoto, K, Akiyama, M, Tamura, F, Isomi, M, Yamakawa, H, Sadahiro, T*, et al.* (2018). Direct in vivo reprogramming with Sendai virus vectors improves cardiac function after myocardial infarction. *Cell Stem Cell* **22**: 91-103 e105.
339. Ocampo, A, Reddy, P, Martinez-Redondo, P, Platero-Luengo, A, Hatanaka, F, Hishida, T*, et al.* (2016). In vivo amelioration of age-associated hallmarks by partial reprogramming. *Cell* **167**: 1719-1733 e1712.
340. Li, A, Lee, CM, Hurley, AE, Jarrett, KE, De Giorgi, M, Lu, W*, et al.* (2019). A self-deleting AAV-CRISPR System for in vivo genome editing. *Mol Ther Methods Clin Dev* **12**: 111-122.
341. Finn, JD, Smith, AR, Patel, MC, Shaw, L, Youniss, MR, van Heteren, J*, et al.* (2018). A single administration of CRISPR/Cas9 lipid nanoparticles achieves robust and persistent in vivo genome editing. *Cell Rep* **22**: 2227-2235.
342. Miller, JB, Zhang, S, Kos, P, Xiong, H, Zhou, K, Perelman, SS*, et al.* (2017). Non-viral CRISPR/Cas gene editing in vitro and in vivo enabled by synthetic nanoparticle co-delivery of Cas9 mRNA and sgRNA. *Angew Chem Int Ed Engl* **56**: 1059-1063.
343. Yin, H, Song, CQ, Suresh, S, Wu, Q, Walsh, S, Rhym, LH*, et al.* (2017). Structure-guided chemical modification of guide RNA enables potent non-viral in vivo genome editing. *Nat Biotechnol* **35**: 1179-1187.
344. Ling, C, Bhukhai, K, Yin, Z, Tan, M, Yoder, MC, Leboulch, P*, et al.* (2016). High-efficiency transduction of primary human hematopoietic stem/progenitor cells by AAV6 vectors: strategies for overcoming donor-variation and implications in genome editing. *Sci Rep* **6**: 35495.
345. Nathwani, AC, Gray, JT, Ng, CY, Zhou, J, Spence, Y, Waddington, SN*, 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.
346. Bak, RO, Dever, DP, and Porteus, MH (2018). CRISPR/Cas9 genome editing in human hematopoietic stem cells. *Nat Protoc* **13**: 358-376.
347. Wang, Z, Ma, HI, 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.
348. Wu, Z, Sun, J, Zhang, T, Yin, C, Yin, F, Van Dyke, T*, et al.* (2008). Optimization of self-complementary AAV vectors for liver-directed expression results in sustained correction of hemophilia B at low vector dose. *Mol Ther* **16**: 280-289.

349. Choi, VW, Samulski, RJ, and McCarty, DM (2005). Effects of adeno-associated virus DNA hairpin structure on recombination. *J Virol* **79**: 6801-6807.
350. Amoasii, L, Long, C, Li, H, Mireault, AA, Shelton, JM, Sanchez-Ortiz, E*, et al.* (2017). Single-cut genome editing restores dystrophin expression in a new mouse model of muscular dystrophy. *Sci Transl Med* **9**: eaan8081.
351. Nelson, CE, Wu, Y, Gemberling, MP, Oliver, ML, Waller, MA, Bohning, JD*, et al.* (2019). Long-term evaluation of AAV-CRISPR genome editing for Duchenne muscular dystrophy. *Nat Med* **25**: 427-432.
352. Yokoi, K, Kachi, S, Zhang, HS, Gregory, PD, Spratt, SK, Samulski, RJ*, et al.* (2007). Ocular gene transfer with self-complementary AAV vectors. *Invest Ophthalmol Vis Sci* **48**: 3324-3328.
353. Merienne, N, Vachey, G, de Longprez, L, Meunier, C, Zimmer, V, Perriard, G*, et al.* (2017). The self-inactivating KamiCas9 system for the editing of CNS disease genes. *Cell Rep* **20**: 2980-2991.
354. Lee, J, Mir, A, Edraki, A, Garcia, B, Amrani, N, Lou, HE*, et al.* (2018). Potent Cas9 inhibition in bacterial and human cells by AcrIIC4 and AcrIIC5 anti-CRISPR proteins. *MBio* **9**: e02321-18.
355. Shin, J, Jiang, F, Liu, JJ, Bray, NL, Rauch, BJ, Baik, SH*, et al.* (2017). Disabling Cas9 by an anti-CRISPR DNA mimic. *Sci Adv* **3**: e1701620.
356. Hoffmann, MD, Aschenbrenner, S, Grosse, S, Rapti, K, Domenger, C, Fakhiri, J*, et al.* (2019). Cell-specific CRISPR-Cas9 activation by microRNA-dependent expression of anti-CRISPR proteins. *Nucleic Acids Res*. doi: 10.1093/nar/gkz271.
357. Nihongaki, Y, Kawano, F, Nakajima, T, and Sato, M (2015). Photoactivatable CRISPR-Cas9 for optogenetic genome editing. *Nat Biotechnol* **33**: 755-760.
358. Hirosawa, M, Fujita, Y, Parr, CJC, Hayashi, K, Kashida, S, Hotta, A*, et al.* (2017). Cell-type-specific genome editing with a microRNA-responsive CRISPR-Cas9 switch. *Nucleic Acids Res* **45**: e118.
359. Charlesworth, CT, Deshpande, PS, Dever, DP, Camarena, J, Lemgart, VT, Cromer, MK*, et al.* (2019). Identification of preexisting adaptive immunity to Cas9 proteins in humans. *Nat Med* **25**: 249-254.
360. Shen, CC, Hsu, MN, Chang, CW, Lin, MW, Hwu, JR, Tu, Y*, et al.* (2018). Synthetic switch to minimize CRISPR off-target effects by self-restricting Cas9 transcription and translation. *Nucleic Acids Res*. doi: 10.1093/nar/gky1165.
361. Pan, D, Xuan, B, Sun, Y, Huang, S, Xie, M, Bai, Y*, et al.* (2016). An intein-mediated modulation of protein stability system and its application to study human cytomegalovirus essential gene function. *Sci Rep* **6**: 26167.
362. Lagos-Quintana, M, Rauhut, R, Yalcin, A, Meyer, J, Lendeckel, W, and Tuschl, T (2002). Identification of tissue-specific microRNAs from mouse. *Curr Biol* **12**: 735-739.

6 PUBLICATIONS, PRESENTATIONS AND PATENT APPLICATIONS

6.1 Publications

- **Fakhiri**, **J**, Schneider, MA, Puschhof, J, Stanifer, M, Schildgen, V, Holderbach, S, Voss, Y, El Andari, J, Schildgen, O, Boulant, S, Meister, M, Clevers, H, Yan, Z, Qiu, J and Grimm, D (2019). Novel chimeric gene therapy vectors based on adeno-associated virus and four different mammalian bocaviruses. *Mol Ther Methods Clin Dev* **12**: 202-222.
- Tabatabai, J, **Fakhiri**, **J**, Meyburg, J, Linse, KP, Xu, M, Soderlund-Venermo, M, Grimm, D and Schnitzler, P (2019). Severe human bocavirus 1 respiratory tract infection in an immunodeficient child with fatal outcome. *Pediatr Infect Dis J*. doi: 10.1097/INF.0000000000002354.
- **Fakhiri**, **J**, Nickl, M and Grimm, D (2019). Rapid and simple screening of CRISPR g(uide)RNAs in cultured cells using Adeno-associated viral (AAV) vectors. *Methods Mol Biol* **1961**:111-126.
- Hoffmann, MD, Aschenbrenner, S, Grosse, S, Rapti, K, Domenger, C, **Fakhiri, J**, Mastel, M, Börner, K, Eils, R, Grimm, D and Niopek, D (2019). Cell-specific CRISPR-Cas9 activation by microRNA-dependent expression of anti-CRISPR proteins. *Nucleic Acids Res*. https://doi.org/10.1093/nar/gkz271.
- Bubeck, F, Hoffmann, MD, Harteveld, Z, Aschenbrenner, S, Bietz, A, Waldhauer, MC, Börner, K, **Fakhiri**, **J**, Schmelas, C, Dietz, L, Grimm, D, Correia, B, Eils, R, and Niopek, D (2018). Engineered anti-CRISPR proteins for optogenetic control of CRISPR-Cas9. *Nat Methods* **15**: 924-927.
- Grosse, S, Penauld-Budloo, M, Herrmann, AK, Börner, K, **Fakhiri**, **J**, Laketa, V, Krämer, C, Wiedtke, E, Gunkel, M, Menard, L, Ayuso, E and Grimm, D (2017). Relevance of assembly-activating protein for adeno-associated virus vector production and capsid protein stability in mammalian and insect cells. *J Virol* **91**: e01198-17.

6.2 Presentations

- Annual meeting of the American Society of Gene & Cell Therapy (ASGCT). Mai 2016, Washington, USA. Talk: “New chimeric gene therapy vectors based on four different mammalian bocaviruses“. (Travel award from ASGCT society)
- Annual meeting of the German Society of Gene and Cell Therapy (DG-GT). September 2016, Heidelberg, Germany. Talk: “A CRISPR/Cas9 inducible kill-switch for efficient shutdown of transgene expression from rAAV vectors“.
- XVI International parvovirus workshop (Biannual meeting). June 2016, Ajaccio, France. Talk: “New chimeric gene therapy vectors based on four different mammalian bocaviruses“. (Travel award from the HBIGS graduate school)
- Annual meeting of the European Society of Gene and Cell Therapy (ESGCT). October 2017, Berlin, Germany. Talk: “New chimeric gene therapy vectors based on four different mammalian bocaviruses“.
- Annual meeting of the American Society of Gene & Cell Therapy (ASGCT). Mai 2018, Chicago, USA. Talk: “Molecular evolution of the next generation of bocaviral vectors”. (Travel award from ASGCT and University of Heidelberg Graduate Academy)

6.3 Patent applications

- "Means and methods for inactivating therapeutic DNA in a cell" (2016). International Application Number: *PCT/EP2017/059187*
- "Modulated Cas-Inhibitors" (2017). International Application Number: *PCT/EP2018/077164*
- "Use of anti-CRISPR polypeptides for specific activation of Cas nucleases" (2017). International Application Number: *PCT/EP2018/072350*

7 ACKNOWLEDGEMENTS

An erster Stelle möchte ich mich bei dem Betreuer meiner Arbeit Prof. Dr. Dirk Grimm bedanken, der mich in seiner Arbeitsgruppe aufgenommen und mir die Möglichkeit gegeben hat, an vielen spannenden Projekten zu arbeiten. Ohne seinen wertvollen, wissenschaftlichen Rat und den vielen inspirierenden Anstößen, wäre die Arbeit in dieser Form nicht entstanden. Ihm ist es zu verdanken, dass ich mich während meiner Promotionszeit in verschiedenen Bereichen der Gentherapie einbringen konnte und mit mehreren Arbeitsgruppen erfolgreich kooperiert habe. Die Präsentation und Publikation der daraus entstandenen Forschungsergebnisse, hat er immer mit sehr viel Begeisterung und Einsatz unterstützt. Ich möchte ihm auch herzlich danken für die sehr hilfreichen Vorschläge zur Verbesserung meiner Arbeiten, Anträge und Manuskripte. All diese Erfahrungen haben sowohl meine wissenschaftliche als auch persönliche Entwicklung sehr geprägt und dafür allein, bin ich zu großem Dank verpflichtet.

Mein Dank gilt auch Prof. Dr. Christof von Kalle für die Bereitschaft, das Erstgutachten dieser Arbeit zu übernehmen. Außerdem möchte ich mich bei den weiteren Mitgliedern meiner Prüfungskommission Prof. Dr. Marc Freichel und Dr. Alessia Ruggieri bedanken.

Ich danke auch Prof. Dr. Christof von Kalle, Prof. Dr. Marc Freichel und Dr. Steeve Boulant für den ermutigenden und konstruktiven, wissenschaftlichen Austausch während der TAC-meetings, der meine Projekte deutlich vorangetrieben hat.

Dankbar bin ich auch für die durchgehende Finanzierung meiner Forschungsarbeit. Für die ersten drei Jahre bin ich besonders der Heidelberg Biosciences International Graduate School (HBIGS), für die Vergabe eines Promotionsstipendiums verbunden. Auch in dieser Hinsicht bin ich Prof. Dr. Dirk Grimm sehr dankbar für das erfolgreiche Einwerben von Forschungsgeldern (Cystic Fibrosis Foundation) und Innovationsfonds (FRONTIER), die den Abschluss meines Promotionsvorhabens entscheidend unterstützt haben.

Ich möchte auch all meinen ehemaligen und jetzigen Kollegen in der AG Grimm für die nette, offene und freundschaftliche Arbeitsatmosphäre danken. Es war immer schön mit euch zusammenzuarbeiten! Eure wertvollen Anregungen, Diskussionsbeiträge und konstruktive Kritik haben meine Arbeit enorm bereichert. Besonders danken möchte ich an dieser Stelle Carolin Schmelas, die an der Fortführung eines bedeutenden Teils dieser Arbeit beteiligt ist. Es war immer schön mit dir zu planen und zu arbeiten, ob an der Bench oder im „Maus Hotel“. Sehr geholfen hat mir auch Ellen Wiedtke bei den ersten Verpackungsversuchen von (meinen) Bocaviren. Sehr dankbar bin ich Dr. Janina Haar und Chiara Krämer für die Bereitstellung

verschiedener primärer Zellen, die eine große Wende in meiner Dissertation brachten. Unterstützung bekam ich auch von Cinja Sackmann, besonders wenn es mal durch härtere experimentelle Zeiten ging. Außerdem danke ich Stefan Holderbach für die gute Zusammenarbeit bei der Optimierung und Generierung von der chimären, bocaviralen Bibliothek, die eine interessante Basis für Zukunftsexperimente setzt. Ein besonders großer Dank gebührt Dr. Jihad El Andari (Jad), der mir mit viel Rat, Unterstützung und einem großartigen Humor zur Seite stand.

Für vielfältige Unterstützung bin ich weiteren Kollegen am Department für Infektiologie, Virologie, Universitätsklinikum Heidelberg, verbunden: Robin Burk für die ersten Elektronenmikroskop-Aufnahmen von Bocaviren, David Bejarano und Dr. Kathleen Börner für die Bereitstellung von primären PBMCs und Macrophagen, Manuela Nickl für die Isolierung und das Teilen von primären T-Zellen und Dr. Andrea Imle für die hilfreichen, praktischen Ratschläge.

Ein ganz besonderer Dank gilt Dr. Michael Meister und Dr. Marc Schneider, für ihren kontinuierlichen Supply mit primären Lungenepithelzellen, die uns erste unentbehrliche Einblicke in die Biologie der Bocaviren gewährt haben. Prof. Dr. Oliver Schildgen und Dr. Verena Schildgen danke ich für die darauffolgenden Validierungsexperimente in Cufi-8 Zellen.

Auch danken möchte ich Prof. Dr. Hans Clevers, Dr. Steeve Boulant, Dr. Megan Stanifer und Jens Puschhof für die Bereitstellung von primären Organoiden (Mini-Organen), die diese Arbeit sehr bereichert haben.

Ich danke Iris Augustin für ihre Hilfe bei den ersten Experimenten zur in vivo Validierung des KS Reporters. Diese Ergebnisse formten eine vielversprechende Basis für weitere Ansätze in dieser Arbeit.

Ein ganz großer Dank gilt den Freunden, die ich in dieser Zeit gewonnen habe: Manuela (Nickl), Vilmante (Zitkute) und Jad (El Andari). Ihr habt mir dabei geholfen die Motivation immer aufrechtzuerhalten!

Eine wichtige und herausragende Stellung nimmt meine Familie ein, der ich diese Arbeit widme. Ohne ihre liebevolle Fürsorge und besondere Unterstützung in allen Phasen meines Lebens, wäre ich nicht da, wo ich heute bin. Ihnen gilt mein ganz besonderer Dank!