OVERLINE: GENE THERAPY

Restoring the natural tropism of AAV2 vectors for human liver

Authors:
Marti Cabanes-Creus[1], Claus V. Hallwirth[2], Adrian Westhaus[1,3], Boaz H. Ng[4], Sophia H.Y. Liao[1], Erhua Zhu[2], Renina Gale Navarro[1], Grober Baltazar[1], Matthieu Drouyer[1], Suzanne Scott[2,5], Grant J. Logan[2], Giorgia Santilli[3], Antonette Bennett[6], Samantha L. Ginn[2], Geoff McCaughan[7], Adrian J. Thrasher[3], Mavis Agbandje-McKenna[6], Ian E. Alexander[2,8] and Leszek Lisowski[1,4,9]

Affiliations:
[1]Translational Vectorology Research Unit, Children's Medical Research Institute, Faculty of Medicine and Health, The University of Sydney, Westmead, NSW 2145, Australia

[2]Gene Therapy Research Unit, Children's Medical Research Institute and The Children's Hospital at Westmead, Faculty of Medicine and Health, The University of Sydney and Sydney Children's Hospitals Network, Westmead, NSW 2145, Australia

[3]Great Ormond Institute of Child Health, University College London, WC1N 1EH London, UK

[4]Vector and Genome Engineering Facility, Children's Medical Research Institute, Faculty of Medicine and Health, The University of Sydney, Westmead, NSW 2145, Australia

[5]Commonwealth Scientific and Industrial Research Organisation (CSIRO), North Ryde, NSW 2113, Australia

[6]Department of Biochemistry and Molecular Biology, Center for Structural Biology, University of Florida, Gainesville, FL 32610, USA

[7]Liver Injury and Cancer Program, Centenary Research Institute, A.W Morrow Gastroenterology and Liver Centre, Australian National Liver Transplant Unit, Royal Prince Alfred Hospital, The University of Sydney, Sydney, NSW 2006, Australia

[8]Discipline of Child and Adolescent Health, The University of Sydney, Sydney Medical School, Faculty of Medicine and Health, Westmead, NSW 2145, Australia

[9]Military Institute of Hygiene and Epidemiology, The Biological Threats Identification and Countermeasure Centre, 24-100 Puławy, Poland

One sentence summary:
Adeno-associated viruses isolated from the human liver are tropic for human hepatocytes in vivo but this phenotype becomes attenuated after passage in culture.

Editor's Summary

Abstract

Recent clinical successes in gene therapy applications have intensified interest in using adeno-associated viruses (AAV) as vectors for therapeutic gene delivery. Although prototypical AAV2 shows robust in vitro transduction of human hepatocyte-derived cell lines, it has not translated into an effective vector for gene therapy in vivo. This is consistent with observations made in the $Fah^{-/-}/Rag2^{-/-}/Il2rg^{-/-}$ (FRG) humanized liver xenograft mouse showing that AAV2 functions poorly in this model. Here, we derived naturally hepatotropic AAV capsid sequences from primary human liver samples. We demonstrated that capsid mutations, likely acquired as an unintentional consequence of tissue culture propagation, attenuated the intrinsic human hepatic tropism of natural AAV2 and related human liver AAV isolates. These mutations resulted in amino acid changes that increased binding to heparan sulfate proteoglycan (HSPG), which has been regarded as the primary cellular receptor mediating AAV2 infection of human hepatocytes. Propagation of natural AAV variants in vitro showed tissue culture adaptation with resulting loss of tropism for human hepatocytes. In vivo re-adaptation of the prototypical AAV2 in FRG mice with a humanized liver resulted in restoration of the intrinsic hepatic tropism of AAV2 through decreased binding to HSPG. Our results challenge the notion that high affinity for HSPG is essential for AAV2 entry into human hepatocytes and suggest that natural AAV capsids of human liver origin are likely to be more effective for liver-targeted gene therapy applications than culture-adapted AAV2.

Introduction

Adeno-associated viruses (AAVs) were discovered in the 1960s as contaminants of adenoviral isolates(*1*) and subsequently shown to be endemic in both human(*2*) and non-human primate populations(*3*). The vectorization of AAV2, a human isolate, in 1984(*4*) set in motion the development of an increasingly powerful and modular gene transfer and editing system that is now delivering therapeutic success in the treatment of genetic diseases such as spinal muscular atrophy(*5*), RPE65 deficiency(*6*) and hemophilia B (*7*). These successes have hinged on continuing development of the AAV vector system, driven primarily by advances in AAV capsid technology, the principal determinant of vector tropism. The liver, in particular, is a key target for the development of more efficient AAV vector delivery given its direct involvement in an extensive set of genetic and acquired diseases. Successes in liver-targeted gene therapy have thus far primarily been for conditions in which the secretion of functional proteins is impaired and where gene transfer to a relatively low proportion of hepatocytes confers therapeutic benefit, such as hemophilia A and B(*7-9*). Substantially higher targeting efficiencies will be required to bring the majority of liver disorders, such as the cell-autonomous ornithine transcarbamylase deficiency (OTCD)(*10*) within technological reach of AAV-mediated gene transfer. This necessitates a deeper understanding of the determinants of human liver tropism and the development of AAV vectors with increased liver transduction efficiency.

We have recently shown that the prototypical human isolate, AAV2(*11*), contains a liver-specific enhancer-promoter element in the 3' untranslated region (UTR) between the *cap* stop codon and the right-hand inverted terminal repeat (ITR) (*12*). This 124-nucleotide sequence contains binding sites for human master hepatic transcription factors including HNF1α, HNF4α, HNF6 and GATA6, providing direct evidence that AAV2 is evolutionarily associated

with infection of the human liver (*12*). Although little is known about the life-cycle of AAV2, membrane-associated heparan sulfate proteoglycan (HSPG) is widely accepted as a critical determinant of the ability of AAV2 to attach to and transduce cells (*13*). In support of AAV2's dependence on HSPG for cell entry, soluble heparin was shown to compete for AAV2 binding to permissive cells, leading to decreased transduction (*13*). The determination of the AAV2 crystal structure(*14*) together with extensive mutagenesis analysis(*15*), have localized the HSPG binding domain in AAV2 to a patch of basic amino acids residues on the capsid surface(*16*), with five basic residues (R484, R487, K532, R585 and R588) being essential for the interaction of AAV2 with HSPG(*17*). These residues facilitate high efficiency transduction of human hepatocyte-derived cell lines(*18, 19*).

Despite the evolutionary relationship with human liver, clinical use of the AAV2 capsid resulted in unexpectedly low efficacy in a liver-directed gene therapy trial (*20*). While this early study demonstrated that AAV2-based vectors could transduce human hepatocytes in vivo, peak therapeutic values of 3-11% of factor IX were observed only in subjects infused with the highest dose tested of 2×10^{12} vg/kg via the hepatic artery. These results underpinned the clinical evaluation of alternative AAV serotypes, such as the non-human primate-derived AAV8(*21*), in liver-targeted clinical trials. However, despite the fact that AAV8 mediated over 10-fold higher gene transfer into the liver in animal models, including non-human primates(*8, 9, 22-24*), in human studies the peak concentrations of vector-encoded human factor IX were similar for individuals treated with AAV2 and AAV8 (*9, 20*). These results are consistent with subsequent observations made by ourselves and others in a xenograft mouse model of human liver, the $Fah^{-/-}/Rag2^{-/-}/Il2rg^{-/-}$ (FRG)(*25*), where AAV2 and AAV8 each only transduced around 3% of human hepatocytes (*19*).

More recently, directed evolution-based selection of AAV capsid libraries on this model has yielded AAV vectors with improved human hepatotropism, such as AAV-LK03(*19*) and AAV-NP59 (*26*). By studying the correlation between capsid sequence/structure of AAV-NP59, which differs from prototypical AAV2 in only eleven amino acid positions, we have recently reported that attenuation of the heparan sulfate proteoglycan binding affinity enhances transduction performance of AAV2 in vivo in the hFRG model, most likely due to reduced vector sequestration on HSPG-rich matrixes, leading to overall improved biodistribution and human hepatocyte transduction (*27*).

A key question arising from these observations relates to the mechanistic understanding behind AAV2's high affinity for HSPG, given the observed detrimental effect on in vivo vector function. In this study we postulated that high affinity to HSPG could be an end result of tissue culture replicative adaptation, rather than a natural property of AAV2. We hypothesized that tissue culture replication of naturally occurring AAV variants would ultimately lead to selection of clones that acquired mutations leading to stronger HSPG binding. Finally, we hypothesized that the opposite phenomenon would occur upon in vivo replication of AAV2 in primary human hepatocytes, where the viral population would shift to HSPG de-targeted variants.

Results

The human liver is a rich source of AAV variants

In order to identify AAV capsid residues that were evolutionarily selected to support efficient targeting of primary human liver we performed an in silico analysis of capsid sequences of natural hepatotropic AAV variants. To do so, in addition to analyzing published sequences(*28, 29*) we undertook a screen of primary human liver samples to identify naturally hepatotropic wildtype AAV variants. A survey of 72 liver samples for the presence of proviral AAV sequences using multiplex PCR identified 30 samples (42 %) that were AAV-positive. Of those, seven (designated AAV-hu.Lvr01-07) yielded amplicons encompassing the entire *cap* open reading frame when amplified with a *rep2*-specific primer and a reverse primer placed at the AAV2 3'UTR region. From among the seven novel isolates, the *rep* coding sequence was determined for four. A phylogeny of deduced capsid amino acid sequences revealed that two of the identified sequences clustered with AAV2 (clade B), while the remaining five were intermediate between AAV2 and AAV13 (**Fig. 1A**), and likely corresponded to the previously described hybrid sequences comprising AAV clade C (*2, 30*). Consistent with the previously postulated origin of AAVs in clade C (*2*), the *rep* sequences of hu.Lvr01, hu.Lvr02, hu.Lvr05 and hu.Lvr07 clustered phylogenetically with AAV2 (**fig. S1**).

AAVs isolated from human liver transduced primary human hepatocytes in vivo with high efficiency

We next investigated the ability of the new human liver AAV isolates to transduce primary human hepatocytes in a xenograft mouse model of human liver, the humanized $Fah^{-/-}/Rag2^{-/-}/Il2rg^{-/-}$ (hFRG) mouse (*25*). To do so, six barcoded (BC) reporter constructs expressing eGFP under the control of a liver-specific promoter (LSP) (ssAAV-LSP1-eGFP-BC-WPRE-BGHpA)(*31*) were independently packaged into each of the AAV-hu.Lvr01-07 variants, reference serotypes AAV2 and AAV13, as well as AAV8 and AAV-LK03 as positive controls for functional transduction of murine and human hepatocytes, respectively(*19*). Two hFRG

mice were injected intravenously with 5.5×10^{11} vector genomes (vg) containing an equimolar mix of the individual variants (5×10^{10} vg of each capsid). Next-generation sequence (NGS) analysis of the barcoded transgene was performed to measure the ability of each variant to transduce murine liver cells and human hepatocytes. Specifically, NGS at the DNA level was used to quantify the cell entry (physical transduction), while NGS at the RNA/cDNA was used to quantify transgene expression (functional transduction). Interestingly, all the human liver-derived variants were more efficient at transduction of primary human hepatocytes than the respective closest reference serotype (AAV2 or AAV13), at both DNA and RNA levels (**Fig. 1B-C**). As expected, based on the known ability of AAV8 to functionally transduce murine hepatocytes with high efficiency(*19*), the majority of cDNA barcodes recovered from the murine liver mapped to this capsid (**fig. S2).**

Immunohistochemical analysis of all AAV-hu.Lvr variants and controls in individual hFRG mice showed that six out of the seven liver-isolated variants were more efficient at functional transduction of primary human hepatocytes than AAV-LK03, the positive control (**Fig. 1D-E**).

Natural AAV isolates were deficient in HSPG binding

Comparison of the two AAV2-like sequences, AAV-hu.Lvr02 and AAV-hu.Lvr06, revealed that they lacked two critical arginine residues associated with AAV2 HSPG binding (R585 and R588, **fig. S3A**), which is in keeping with most previously published clade B variants detected in human tissues (*2, 28-30, 32*). Similarly, all AAV13-like variants identified by us and others(*29*) lacked the structural equivalent of the key lysine 528 residue required for HSPG binding (**fig. S3B**).

Given that all isolated variants lacked key residues of the HSPG binding motif, the heparin binding capacity (a surrogate for HSPG binding(*17*)) of the isolated AAV capsids hu.Lvr01 to 07 was analyzed. In contrast to the prototypical AAV2 and AAV13, which bound to and eluted

from a HiTrap Heparin Column with an elution peak maximum at ~454 mM and ~332 mM NaCl, respectively, all seven liver-derived isolates were unable to efficiently bind to the column (**Table 1 and figs. S4-S11**). In vitro competition assays, using soluble heparin as an analogue of HSPG, provided further evidence that the human liver isolates did not interact with heparin at the tested concentration (**Fig. 1F**). Notably, prototypical AAV2 transduced HuH-7, a human hepatoma-derived cell line, at a considerably higher efficiency than the primary human liver variants, and the transduction was significantly ($p = 0.0079$) inhibited by heparin.

AAVs rapidly acquired mutations leading to amino acid changes, enhancing HSPG binding during in vitro replication

The substantial difference in the in vivo and in vitro performance of the human liver isolates and prototypical AAV2 (**Fig. 1 B-C, Fig. 1F**), combined with the differential interactions with heparin, led us to hypothesize that strong HSPG binding might not be a common property among natural AAVs, but rather a characteristic acquired during in vitro propagation as a result of viral adaptation to a new environment.

Given the similarly poor in vitro transduction performance of AAV-hu.Lvr03, hu.Lvr04 and hu.Lvr07 on HuH-7 cells (**Fig. 1F**), AAV-hu.Lvr07 was selected to test the hypothesis of tissue culture adaptation. To do so, a replication-competent (RC) version (ITR2-rep2-cap.hu.Lvr07-ITR2) of AAV-hu.Lvr07, referred to as RC-AAV-hu.Lvr07, was subjected to an in vitro adaptation protocol on HuH-7 cells in the presence of human adenovirus 5 (hAd5) (**Fig. 2A)**. No AAV was detected in HuH-7 cells infected only with hAd5 (**fig. S12**). After four rounds of tissue culture replication, 98.57 % of the isolated capsids harbored a glutamate to lysine change at position 530 (E530K) and 1.14 % of clones had acquired an arginine at position 593 (G593R). Both these amino acids have positively charged side-chains and have been shown to facilitate heparin binding at corresponding structural regions in AAV13 and AAV3B,

respectively (**Table 2**)(*16, 33*). In-depth analysis of the capsid gene region containing both modified residues using high-throughput NGS demonstrated that the original AAV-hu.Lvr07 was displaced rapidly after the initial appearance of E530K and G593R in round 1 and 2, respectively (**Fig. 2B-C**). The same phenomenon was observed upon hAd5-supported iterative replication of a RC version of AAV-hu.Lvr02 (ITR2-rep2-cap.hu.Lvr02-ITR2) in HuH-7 cells. The initial single-variant population was rapidly dominated by a variant that acquired two point mutations that resulted in substitution of acidic amino acids at positions 469 (D469N) and 555 (E555K) (**fig. S13**).

In vitro adaptation improved cell culture vector function and attenuated in vivo performance

To test whether the observed changes in the AAV-hu.Lvr07 capsid sequence affected transduction characteristics, the two new AAV-hu.Lvr07 variants, harboring solely the individual amino acid substitutions (E530K or G593R), were used to package the AAV-LSP1-eGFP-WPRE-BGHpA vector and were evaluated in vitro on HuH-7 cells. FACS analysis revealed that culture-adapted (ca)AAV-hu.Lvr07^{E530K} and caAAV-hu.Lvr07^{G593R} functionally transduced HuH-7 cells more efficiently than AAV-hu.Lvr07 (**Fig. 2D**). Furthermore, co-incubation with soluble heparin reduced the observed improvement in transduction to the extent observed for natural AAV-hu.Lvr07 (**Fig. 2D**). In contrast to the natural AAV-hu.Lvr07, the tissue culture-adapted variants (E530K and G593R) were shown to efficiently bind to the HiTrap Heparin Column (**Table 1**, **figs. S14-S15)**. This was anticipated, as structural analysis revealed that the E530K change was adjacent to residues R487, K527, and K532, which are part of the AAV13 HSPG binding region(*16*), and created a basic patch on the capsid surface in close proximity to R484 at base of the 3-fold protrusions (**Fig. 2E-F**, **Table 2**). An identical phenotype was observed for caAAV-hu.Lvr02$^{D469N+E555K}$, which in contrast to the natural

serotype was found to efficiently bind to the HiTrap Heparin column (**fig. S16**) and to transduce HuH-7 with significantly ($p = 0.0286$) higher efficiency (**fig. S17**).

We next wanted to investigate whether change in HSPG-binding properties would have a detrimental effect in vivo. For that, six barcoded reporter constructs were packaged into reference serotypes AAV2, AAV8 and AAV13, as well as AAV-hu.Lvr07 and its tissue culture adapted version (caAAV-hu.Lvr07^{E530K}). Two hFRG mice were injected intravenously with a total of 2.5×10^{11} vg containing an equimolar mix of the individual variants (5×10^{10} vg of each capsid). Interestingly, the single amino acid substitution at the corresponding structural region in AAV13 (E530K) sufficed to significantly ($p = 0.0022$) attenuate the in vivo performance of the otherwise highly human hepatotropic AAV-hu.Lvr07 to an AAV13-like phenotype (**Fig. 2G-H**).

In vivo propagation of the prototypical AAV2 led to rapid acquisition of HSPG de-targeting substitutions

In view of the rapid kinetics of mutagenesis observed in vitro and the resultant change in vector performance following tissue culture adaptation, we investigated whether it was possible to "re-adapt" the prototypical AAV2 on primary human hepatocytes in order to improve in vivo function. To this end, a replication-competent version of AAV2 (RC-AAV2, ITR2-*rep2*-*cap2*-ITR2) was used to perform an iterative in vivo adaptation experiment on primary human hepatocytes in hFRG mice in the presence of hAd5 (**Fig. 3A)**. Naïve FRG mice and an hFRG mouse injected with hAd5 only were used as negative controls (**Methods**, **fig. S26**). Sequence analysis of random full-length *cap* genes (n = 24) recovered after the first round of the in vivo adaptation experiment revealed substantial enrichment (16.6 % of the population) of an AAV2 clone harboring a point mutation leading to a R588I substitution, affecting one of the key arginine residues involved in HSPG binding. Interestingly, 12.0 % of clones analyzed after

round 2 harbored an R487Q amino acid substitution. While this clone was selected against, or outcompeted by R588I, in the subsequent rounds of selection, R487Q affected another key arginine residue involved in heparin binding and was therefore included as a variant in subsequent analysis (**Table 1**). Importantly, clones harboring other random DNA point-mutations, including silent ones, were observed in the three rounds of selection, as detailed in **tables S7-S9**.

To gain a better understanding of the kinetics of the population genetic shift, Illumina high-throughput NGS of the 150-bp region surrounding the affected positions G1460A (amino acid 487) (**Fig. 3B**) and G1763T (amino acid 588) (**Fig. 3C**) was performed in the initial virus stock and after each round of in vivo selection. As shown in **Fig. 3C**, the R588 present in AAV2 was rapidly replaced by R588I within two rounds of selection. NGS analysis of the pre-adapted RC-AAV2 preparation suggested that the identified variants resulted from random mutagenesis events that gave rise to AAV2 variants with improved biodistribution and overall in vivo tropism for primary human hepatocytes. Supporting this hypothesis, a third variant, R588T, was detected after the second round of selection and accounted for 1.37 % of the total reads. However, AAV2-R588T was no longer detectable after the subsequent round of selection.

To further validate the hypothesis that the observed variants resulted from random mutagenesis events occurring during viral replication, we repeated the in vivo adaptation study using a new preparation of viral RC-AAV2. During the second independent in vivo adaptation experiment, AAV2-R588T rapidly expanded over other capsid variants and accounted for 88 % of the population following the second round of selection (**Fig. 3D**). As mentioned earlier, the interaction between the 3-fold protrusions of AAV2 and HSPG has been reported to be mediated mainly by interactions of the sugar sulfates with positively charged basic residues located on the face of the 3-fold protrusions(*34*). The three amino acid substitutions, R487Q and R588I/T, disrupt and reduce the positive charge of AAV2 (**Fig. 3E-F**), and would therefore

reduce the strength of the AAV2-HSPG interaction, consistent with our observations using the HiTrap column (**Table 1**, **figs. S20-S22)**.

Single amino acid changes acquired during in vivo readaptation restored the intrinsic human hepatotropism of AAV2

To evaluate the functionality of the new AAV variants, the three clones AAV2-R588I, AAV2-R588T, and AAV2-R487Q were vectorized and used to package the AAV-LSP1-eGFP-BC-WPRE-BGHpA cassette. To more comprehensively understand the effect of the amino acid substitutions on vector function, the three variants were tested in vitro on HuH-7 cells (**Fig. 3G**) and in vivo on primary human hepatocytes in hFRG mice (**Fig. 4A-B**). As expected, all variants showed reduced transduction of HuH-7 at the MOIs tested when compared to the prototypical AAV2 (**Fig. 3G**) and consistent with natural AAV liver isolates (**Fig. 1F**), performance was not affected by soluble heparin at the concentration tested (**Fig. 3G**). In striking contrast to the in vitro results, all three variants were found to outperform AAV2 and the human hepatotropic bioengineered capsid AAV-LK03 in both entry (DNA) and expression (cDNA) analyses in vivo using primary human hepatocytes (**Fig. 4A-B**). As observed previously (**Supplementary Figure S2),** AAV8-derived barcodes accounted for the majority of signal from murine liver cDNA (**fig. S23**). NGS results were confirmed by immunohistochemical analysis of the three AAV2 variants in individual hFRG mice (**Fig. 4C-D**).

Discussion

During the sixteen years between isolation of prototypical AAV2 and its sequencing and genome annotation, the replicating virus was propagated in vitro on KB and HeLa cells in the presence of Ad12 and Ad2, respectively, for an unknown period of time (*11, 35*). Here we showed that during replication, AAVs that acquire mutations leading to capsid amino acid changes that markedly enhance interaction with HSPG, can rapidly achieve population dominance in vitro. Importantly, while advantageous in tissue culture, clinical(*20*) and preclinical data in chimeric mouse-human livers(*19*) indicate that this property is detrimental to human liver transduction.

Our observation that human liver AAV isolates demonstrate strikingly opposite performance to prototypical AAV2 in vitro and in vivo supports the previously proposed hypothesis(*36*) that the ability to bind HSPG with high affinity might not be a requirement for natural infection in humans, as recently reviewed in the context of AAV and other viruses (*37*). Importantly, while deficient HSPG binding has been shown to markedly reduce AAV2 transduction of the murine liver(*15, 38*), we showed that reduction of heparin binding strength improves functional transduction of human hepatocytes in vivo in a xenograft model.

There are some caveats in the interpretation of these findings. The first is dictated by the difficulty in isolating heparan sulfate (HS) from natural systems and thus the use of structurally similar heparin as a surrogate for HSPG in the binding and competition assays (*39*). Heparin might not perfectly recapitulate biological interactions between HS and AAVs, and it is reasonable to speculate that a transient interaction would be preferential for viral attachment, which is not easily replicated with HiTrap Heparin chromatography. This could explain why a small proportion of the AAV-hu.Lvr variants was detected in the elution fraction. Further

studies will be required to understand if this residual binding could be indicative of a low affinity interaction of AAVs with HS.

Another caveat is extrapolating AAV biology and AAV-liver interactions from studies in a xenograft model. Although the hFRG model allows study of the interactions of AAV with human hepatocytes in vivo, these grow in clusters on the murine scaffold, where the nature of the extracellular environment remains of murine origin. Other limitations associated with this model include the potential effect of cluster size and levels of engraftment on AAV-transduction. In the present study we attempted to minimize these effects by using similarly engrafted hFRGs for vector comparison and analyzing multiple barcoded variants in the same animal. However, barcoded NGS analysis of vector function provides information on the average net transduction in the tested sample and not the vector biodistribution within a hepatic lobule.

We hypothesize that the observed in vivo attenuation of AAVs could be caused by impaired hepatic biodistribution. Under our current proposed model, the culture-adapted variants retain the ability to transduce human cells, as evident from in vitro studies, but excessive HS attachment prevents the vectors from reaching the human clusters. In the context of non-human primates or humans, we hypothesize that acquisition of strong HS binding would direct vector transduction to periportal hepatocytes. In fact, we have recently observed this phenomenon for AAV8 using murine liver (*27*).

The fast kinetics for both the in vitro and in vivo AAV adaptations resembles the mutation rates of the canine parvovirus(*40*) and the human B19 erythrovirus (*41*). This could relate to the fact that AAV replication depends on, and induces, the DNA damage response signaling pathway(*42*), which comprises error-prone DNA polymerases (*43*). It is reasonable to assume that during the replicative cycle of infection numerous viral distributions or clouds, rather than

a uniform population, coexist in infected cells and organisms, a phenomena known as viral quasispecies (*3, 44*). Thus, any AAV variant isolated from an infected host does not necessarily represent a wild-type AAV, but rather a single random individual that coexisted within a pool of other genetically related variants. Importantly, this finding presents novel opportunities for AAV engineering, a method we propose to designate as Quasispecies-Enabled (QE) AAV engineering.

In the absence of sequence data for the original 1966 AAV2 isolate(*11*) we may never know whether the prototypical AAV2 adapted to the tissue culture environment via *de novo* mutations or expanded from a naturally occurring rare variant harboring the HSPG binding motif. Interestingly, HSPG de-targeted variants resulted from single nucleotide mutations. This could represent an intrinsic mechanism to allow rapid adaptation to changing conditions with minimal corresponding changes to the genome. It can be further postulated that during an AAV infection, individual clones arising from random mutagenesis may acquire properties allowing infection of other tissues or surrounding cells. This could explain why not all natural AAV2-like variants published to date lack the HSPG binding motif (*29*).

The paradigm of AAV culture adaptation appears to extend to at least some other prototypical variants. The five liver-derived AAV clade C capsid sequences presented herein resembled AAV13 in terms of heparin-binding residues (*16*), with the key change of lysine 528 to a negatively charged glutamate (*2, 30*). We postulate that all the prototypical variants reported to bind to HSPG, such as AAV3B(*45*), AAV6(*45*) and AAV13(*16*), which like AAV2 were passaged in vitro in the presence of adenovirus before their initial sequencing, may represent culture-adapted variants. Importantly, the principle of preferential periportal zonation would also apply to engineered AAVs harboring HSPG binding regions from these variants, such as AAV-DJ (*18*), AAV-LK03 (*19*) and AAV-KP1 (*46*).

In summary, the studies reported here support the hypothesis that prototypical AAV2 is unlikely to be a primary human isolate, but rather an HSPG-binding variant that achieved population dominance in in vitro culture. The results provide a direct link between increased HSPG binding and resultant attenuation of tropism for primary human hepatocytes in vivo and suggest a mechanism explaining the low efficacy of the AAV2 capsid in the human liver in a clinical trial targeting hemophilia B (*20*). Our study shows that AAV vectors derived directly from human liver transduce primary human hepatocytes in vivo in a xenograft model of human liver with high efficiency and have therapeutic potential. Importantly, the findings presented herein impact the future of AAV-based liver-targeted gene therapy and warrant the re-evaluation of natural AAV human isolate-derived vectors in clinical applications.

Materials and Methods

Study design

The study was designed to investigate the likelihood of replicative adaptation for Adeno-Associated Viruses (AAV). The studies were performed using natural AAV variants isolated directly from human liver samples as well as previously published natural AAVs. For in vitro characterization assays, four replicates for each AAV variant were used to allow defining of statistical significance. Replicative adaptation of AAV in vitro was performed for two independent natural AAV variants, AAV-hu.lvr02 and AAV-hu.lvr07, through four and five passages in HuH-7 cells, respectively. Replicative adaptation of AAV2 in vivo was performed in biological duplicate using three and two humanized FRG mice per study, respectively, each starting from an independent viral preparation. Acquired mutations were first identified by Sanger Sequencing of no less than n=20 random clones, and population shift was then analyzed with Next Generation Sequencing by designing custom primers flanking the regions of interest. Evaluation of heparin affinity of all reported AAV variants was determined using the same 1-mL Heparin HP column, which was cleaned and reconstituted for binding after each sample.

The in vivo studies in humanized FRG mice were performed in 19- to 31-week-old FRG female mice engrafted with human hepatocytes. Animals with similar levels of engraftment were randomly assigned to experimental groups involving NGS and immunohistochemistry-based comparisons. Investigators were not blinded during monitoring and transduction evaluation of AAVs. For NGS-based studies, six random 6-mer barcoded transgenes were packaged per AAV variant, allowing us to statistically distinguish the performance of variants at both DNA and cDNA levels. For studies that employed immunochemical analysis, AAV transduction efficiency was evaluated in one to two mice per AAV variant, and transduction efficiency was calculated by counting GFP positive cells in ten to twenty human clusters.

Human Ethics. Human liver samples were obtained under Royal Prince Alfred approval ethics proposals X16-0490 & HREC/16/RPAH/701 (“Pathogenesis of Human Liver Disease and Liver Cancer”) and X13-0436 & HREC/10/RPAH/130 (“Pathogenesis of human liver disease”), and Sydney Children’s Hospitals Network (SCHN) Human Research Ethics Committee proposal 18/SCHN/236 (“Establishment of a repository of liver cells to be used in researching gene-based approaches to the treatment of liver disorders”). The samples consisted of human liver donor tissues obtained at the time of liver transplantation.

AAV DNA isolation from primary tissues

DNA was extracted from human liver tissue, using a QIAamp Fast DNA Tissue Kit (Qiagen) according to the manufacturer’s instructions, with the following modifications: (i) Frozen liver chunks (~15 mg each) were homogenized in the digestion buffer using a plastic pestle and manual homogenization to minimize DNA shearing. (ii) Digestion at 56°C was carried out with intermittent agitation (5 sec/min) at 1,000 rpm for 30 minutes. (iii) Samples were triturated through a 22G needle to complete homogenization. (iv) Samples were incubated at 56°C for a further 90 minutes with intermittent agitation. PCR primers were designed to amplify full AAV2 capsid and 3’UTR sequences, by annealing a forward primer (Rep_mpx1_F) in the *rep* gene and a reverse primer (3UTR_mpx2_R) in the 3’ITR, spanning the A/D region junction (**table S1**). All amplification reactions were performed using Q5 Hot Start High-Fidelity DNA Polymerase (New England Biolabs) as per the manufacturer’s instructions using 300 ng template DNA and the following thermal cycling conditions: 98°C for 30 sec; 40 cycles

of 98°C for 10 sec, 67°C for 30 sec, 72°C for 2 min 30 sec; 72°C for 10 min. Capsid amplicons were cloned using the StrataClone Blunt PCR Cloning Kit (Agilent) according to the manufacturer's instructions, and Sanger-sequenced at the Australian Genome Research Facility (Westmead, Australia) using standard M13 (-20) and M13 Reverse sequencing primers, as well as amplicon-specific internal sequencing primers. Samples were named similarly to the nomenclature used by Chen *et al.* (*30*), denoting the species and organ of origin, suffixed by ascending numbers in temporal order of their amplification from DNA and thus not linked to patient or sample numbers (AAV-hu.Lvr01).

Mouse studies

All animal care and experimental procedures were approved by the joint Children's Medical Research Institute and The Children's Hospital at Westmead Animal Care and Ethics Committee. CMRI's established *Fah*$^{-/-}$*Rag2*$^{-/-}$*Il2rg*$^{-/-}$ (FRG) mouse(*25*) colony was used to breed recipient animals. FRG mice were housed in individually ventilated cages with 2-(2-nitro-4-trifluoro-methylbenzoyl)-1,3-cyclohexanedione (NTBC)-supplemented in drinking water (8 mg / mL). FRG mice, 6 to 8 weeks old, were engrafted with human hepatocytes (Lonza Group Ltd., Basel, Switzerland) as described previously (*25*). Humanized FRG (hFRG) mice were placed on 10 % NTBC prior to transduction with vectors and were maintained on 10 % NTBC until harvest. Detailed information on all mice used in the study, including individual estimated repopulation, can be found in **table S2**.

The vector for injection was made up to a final volume of 150 µL using saline. Mice were randomly selected and transduced by intravenous injection (lateral tail vein) with the indicated vectors at a dose of 5×10^{10} vg/vector for NGS comparison, and at a dose of 2×10^{11} vg/vector for immunohistochemistry (with the exception of the *in vivo* adaptation, see "*in vivo* adaptation" subsection). Mice were euthanized by CO_2 inhalation 2 weeks after transduction for immunohistochemistry and 1 week after transduction for barcoded Next-Generation Sequencing analysis. Hepatocytes for flow cytometry analysis were obtained by collagenase perfusion of the liver (see below).

Isolation of human hepatocytes by collagenase perfusion

To perfuse mouse liver and obtain single-cell suspension, the inferior vena cava (IVC) was cannulated, and the solutions were pumped with an osmotic minipump (Gilson Minipuls 3) in the following order: 25 mL of Hank's balanced salt solution (HBSS) (−/−) (cat # H9394; Sigma), 25 mL of HBSS (−/−) supplemented with 0.5 mM EDTA, 25 ml HBSS (−/−), and 25 mL of HBSS (−/−) supplemented with 5 mM $CaCl_2$, 0.05 % wt/vol collagenase IV (Sigma) and 0.01 % wt/vol DNase I (Sigma).

Following perfusion, the liver was carefully removed and placed in a Petri dish containing 25 ml of Dulbecco's modified Eagle's medium (DMEM) supplemented with 10 % fetal bovine serum (FBS). The blunt end of a scalpel blade was used to break the liver capsule to release the cells into the medium. After collection, the cells were spun down at $50 \times g$ for 3 min at 4 °C. The pellet was resuspended in 21 mL of DMEM and passed through a 100-µm nylon cell strainer. Isotonic Percoll (9 mL) (1 part of 10 × PBS (−/−) with 9 parts of Percoll; GE Healthcare) was added to the cell suspension to separate live and dead cells. Live cells were pelleted at $650 \times g$ for 10 min at 4 °C and the pellet was resuspended in FACS buffer (PBS (−/−) with 5 % FBS and 5 mM EDTA). To delineate between mouse liver cells and human hepatocytes, cells were labelled with phycoerythrin (PE)-conjugated anti-human-

HLA-ABC (clone W6/32; Invitrogen 12-9983-42; 1:20), biotin-conjugated anti-mouse-H2Kb (clone AF6-88.5, BD Pharmigen 553568; 1:100) and allophycocyanin (APC)-conjugated streptavidin (eBioscience 17-4317-82; 1:500). GFP-positive labelled samples were sorted to a minimal 95 % purity using a BD Influx cell sorter. Flow cytometry was performed in the Flow Cytometry Facility, Westmead Institute for Medical Research, Westmead, NSW, Australia. The data were analyzed using FlowJo 7.6.1 (FlowJo, LLC).

Adeno-associated virus transgene constructs

AAV transgene constructs were cloned using standard molecular biological techniques. All of the vectors used in the study contain AAV2 ITR sequences. The AAV construct pLSP1-eGFP-WPRE-BGHpA, which encodes eGFP under the transcriptional control of a heterologous promoter containing one copy of the *SERPINA1* (hAAT) promoter and two copies of the *APOE* enhancer element, has been previously reported (*47*). Eighty four (n=84) versions of the pLSP1-eGFP-BC-WPRE-BGHpA construct were produced by cloning n=84 unique 6-mer barcodes downstream of eGFP.

AAV vector packaging and viral production

AAV constructs were packaged into AAV capsids using HEK-293 cells and a helper-virus-free system as previously described (*48*). Genomes were packaged in capsid serotypes AAV2, AAV8, LK03 and AAV13 using packaging plasmid constructs pAAV2, pAAV8, pAAV-LK03 and pAAV13, respectively. Replication-competent (RC) AAV variants were packaged by co-transfection of a corresponding p-AAV plasmid containing the full-length AAV genome (ITR2-*rep2*-*cap*-ITR2) and pAd5 into HEK-293T cells.

All vector/virus were purified using iodixanol gradient ultracentrifugation as previously described (*49*). AAV preparations were titred using real-time quantitative PCR (qPCR) using *eGFP*-specific qPCR primers GFP-qPCR-For/Rev or AAV2-*rep*-specific qPCR primers Rep-qPCR-For/Rev (**table S1).** For *in vivo* testing of hu.Lvr capsids (study reported in **Fig. 1B-C**), culture adapted hu.Lvr07 (study reported in **Fig. 2G-H**) and AAV2 variants (study reported in **Fig4. A-B**) n=6 independent barcoded transgenes were packaged per capsid using two different concentrations (n=3 barcoded transgenes at high dose: 13 μg / transgene per preparation, and n=3 barcoded transgenes at low dose: 1.3 μg / transgene per preparation). The presence of the two distinct populations was confirmed by next-generation sequencing of the pre-injection mix. This approach effectively allows the simultaneous test of the same capsid at two doses in a single hFRG mouse. Following normalization to the pre-injection mix, no difference was found between results obtained with transgenes at high and low doses. Thus, data are presented normalized.

In vitro adaptation

The isolated AAV-hu.lvr07 was produced to package an ITR2-*rep2*-*caphu.lvr07*-ITR2 replication-competent AAV2-based genome using the crude lysate production method previously described (*50*). The adaptation was performed following a previously published protocol for selection of a replication-competent library in the presence of wtAd5 (ATCC, Lot# 70010153) (*50*). Briefly, HuH-7 cells were infected in duplicates at five 10-fold serial dilutions. Cells were washed with 1× PBS 6 hours after infection and one of the duplicate series was co-infected with wtAd5, while the other dilution series served as a non-Ad5 replication control. No AAV genomes were detected in hAd5 only control (**fig. S12**).

Replication was measured 72 hours after wtAd5 co-infection by SYBR green qPCR quantification of AAV genomes in wtAd5 co-infected vs control wells, using *rep2* specific oligos (Rep-F1/Rep-F2). The sample with the highest AAV dilution that led to a 100-fold Cq increase over the respective non-Ad5 control was chosen for the subsequent round of selection. To monitor the progress of natural mutagenesis/culture adaptation, after each round of selection, n=24 random AAV genomes were recovered and sequenced with External_Seq F/R primers (**table S1**).

In vivo adaptation of RC-AAV2

Humanized FRG mice were injected with 1.2×10^{12} vg of replication-competent RC-AAV2 by i.v tail vein administration. 5×10^{9} PFUs of wild-type human adenovirus-5 (ATCC, VR-5, Lot# 70010153**)** was administered intraperitoneally (i.p.) 24 h later. Xenograft livers were harvested 72 hrs after hAd5 administration, homogenized and snap frozen in liquid nitrogen. To extract AAV particles, approximately 0.3 g fragment of liver was subjected to three freeze-thaw cycles and mechanical homogenization in the presence of 2× w/v of PBS. Sample was subsequently centrifuged for 30 min at 4°C at top speed in a table-top centrifuge to separate the virus-containing supernatant from cellular debris. To inactivate wtAd5, the virus-containing supernatant was incubated at 65°C for 30 min. Following titration by qPCR, 200 µL of the virus-containing supernatant was administrated i.p. into hFRG mouse for subsequent round of selection. Non-engrafted naïve FRG mice were used as control. Consistent with previous findings(*19*), hAd5 did not support effective AAV replication in murine hepatocytes. A similarly engrafted hFRG (#445) was injected with hAd5 only and processed as described above. No detectable AAV genomes were found on hAd5 only control (**fig. S26**)

Barcode amplification, next-generation sequencing and distribution analysis

The 150 base pair region surrounding the 6-mer barcode was amplified with Q5 High-Fidelity DNA Polymerase (NEB, Cat# M0491L) using BC_F and BC_R primers (**table S1**). Next-generation sequencing library preparations and sequencing using a 2×150 paired-end configuration were performed by Genewiz (Suzhou, China) using an Illumina MiSeq instrument. A workflow was written in Snakemake (5.6)(*51*) to process reads and count barcodes. Paired reads were merged using BBMerge and then filtered for reads of the expected length in a second pass through BBDuk, both from BBTools 38.68 (https://sourceforge.net/projects/bbmap/). The merged, filtered fastq files were passed to a Perl (5.26)(*52*) script that identified barcodes corresponding to AAV variants. NGS reads from the DNA and cDNA populations were normalized to the reads from the pre-injection mix. The analyzed NGS data per sample can be found in **tables S3-S5 and is detailed in Data File S1.**

Heparin binding assay

The heparin affinity of various AAV vector variants was determined using a 1-mL HiTrap Heparin HP column (GE Healthcare, Cat #1704601, Lot #10276193) on an ÄKTA pure 25 M2 (GE Healthcare) fast protein liquid chromatography system. A detailed protocol can be found at the Supplementary Materials and Methods Section.

AAV structural visualizations

To visualize the location of the residues of interest on the AAV-hu.Lvr07 capsid, a 3D homology model of a VP3 monomer was created by supplying the isolated sequence to the online the SWISS-MODEL server (*53*). A 60mer of the VP3 was then generated using the oligomer generator subroutine of the online Viperdb server(*54*), and visualized with the COOT application (*55*). The surface maps were generated in Chimera(*56*) and the stereographic roadmap was created in the program RIVEM (*57*). The positions of the three AAV2 amino acid changes of interest were visualized on the AAV2 capsid surface using VP3 crystal structure coordinates, RCSB PDB # 1LP3, following icosahedral matrix multiplication as described for AAV-hu.Lvr07. The viral asymmetric unit (**Fig. 3E-F**) is shown as a black triangle bounded by 2-, 3-, and 5-fold axes, depicted by an oval, triangle, and pentagon respectively, in both the surface map and the stereographic roadmap.

Statistical analysis

Nonparametric statistical analyses were performed using the two-tailed Mann–Whitney test with the specified biological and technical replicates in each experimental group. (* P value ≤ 0.05; ** P value ≤ 0.01; **** P ≤0.0001, n.s. P value > 0.05).

List of Supplemental Materials

Supplementary Materials and Methods

References:

1. R. W. Atchison, B. C. Casto, W. M. Hammon, Adenovirus-Associated Defective Virus Particles. *Science* **149**, 754-756 (1965).
2. G. Gao, L. H. Vandenberghe, M. R. Alvira, Y. Lu, R. Calcedo, X. Zhou, J. M. Wilson, Clades of Adeno-associated viruses are widely disseminated in human tissues. *J Virol* **78**, 6381-6388 (2004).
3. G. Gao, M. R. Alvira, S. Somanathan, Y. Lu, L. H. Vandenberghe, J. J. Rux, R. Calcedo, J. Sanmiguel, Z. Abbas, J. M. Wilson, Adeno-associated viruses undergo substantial evolution in primates during natural infections. *Proc Natl Acad Sci U S A* **100**, 6081-6086 (2003).
4. P. L. Hermonat, N. Muzyczka, Use of adeno-associated virus as a mammalian DNA cloning vector: transduction of neomycin resistance into mammalian tissue culture cells. *Proc Natl Acad Sci U S A* **81**, 6466-6470 (1984).
5. J. R. Mendell, S. Al-Zaidy, R. Shell, W. D. Arnold, L. R. Rodino-Klapac, T. W. Prior, L. Lowes, L. Alfano, K. Berry, K. Church, J. T. Kissel, S. Nagendran, J. L'Italien, D. M. Sproule, C. Wells, J. A. Cardenas, M. D. Heitzer, A. Kaspar, S. Corcoran, L. Braun, S. Likhite, C. Miranda, K. Meyer, K. D. Foust, A. H. M. Burghes, B. K. Kaspar, Single-Dose Gene-Replacement Therapy for Spinal Muscular Atrophy. *N Engl J Med* **377**, 1713-1722 (2017).
6. J. W. Bainbridge, A. J. Smith, S. S. Barker, S. Robbie, R. Henderson, K. Balaggan, A. Viswanathan, G. E. Holder, A. Stockman, N. Tyler, S. Petersen-Jones, S. S. Bhattacharya, A. J. Thrasher, F. W. Fitzke, B. J. Carter, G. S. Rubin, A. T. Moore, R. R. Ali, Effect of gene therapy on visual function in Leber's congenital amaurosis. *N Engl J Med* **358**, 2231-2239 (2008).
7. L. A. George, S. K. Sullivan, A. Giermasz, J. E. J. Rasko, B. J. Samelson-Jones, J. Ducore, A. Cuker, L. M. Sullivan, S. Majumdar, J. Teitel, C. E. McGuinn, M. V. Ragni, A. Y. Luk, D. Hui, J. F. Wright, Y. Chen, Y. Liu, K. Wachtel, A. Winters, S. Tiefenbacher, V. R. Arruda, J. C. M. van der Loo, O. Zelenaia, D. Takefman, M. E. Carr, L. B. Couto, X. M. Anguela, K. A. High, Hemophilia B Gene Therapy with a High-Specific-Activity Factor IX Variant. *N Engl J Med* **377**, 2215-2227 (2017).
8. A. C. Nathwani, U. M. Reiss, E. G. Tuddenham, C. Rosales, P. Chowdary, J. McIntosh, M. Della Peruta, E. Lheriteau, N. Patel, D. Raj, A. Riddell, J. Pie, S. Rangarajan, D. Bevan, M. Recht, Y. M. Shen, K. G. Halka, E. Basner-Tschakarjan, F. Mingozzi, K. A. High, J. Allay, M. A. Kay, C. Y. Ng, J. Zhou, M. Cancio, C. L. Morton, J. T. Gray, D. Srivastava, A. W. Nienhuis, A. M. Davidoff, Long-term safety and efficacy of factor IX gene therapy in hemophilia B. *N Engl J Med* **371**, 1994-2004 (2014).
9. A. C. Nathwani, E. G. Tuddenham, S. Rangarajan, C. Rosales, J. McIntosh, D. C. Linch, P. Chowdary, A. Riddell, A. J. Pie, C. Harrington, J. O'Beirne, K. Smith, J. Pasi, B. Glader, P. Rustagi, C. Y. Ng, M. A. Kay, J. Zhou, Y. Spence, C. L. Morton, J. Allay, J. Coleman, S. Sleep, J. M. Cunningham, D. Srivastava, E. Basner-Tschakarjan, F. Mingozzi, K. A. High, J. T. Gray, U. M. Reiss, A. W. Nienhuis, A. M. Davidoff, Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. *N Engl J Med* **365**, 2357-2365 (2011).
10. L. Wang, H. Morizono, J. Lin, P. Bell, D. Jones, D. McMenamin, H. Yu, M. L. Batshaw, J. M. Wilson, Preclinical evaluation of a clinical candidate AAV8 vector for ornithine transcarbamylase (OTC) deficiency reveals functional enzyme from each persisting vector genome. *Mol Genet Metab* **105**, 203-211 (2012).
11. M. D. Hoggan, N. R. Blacklow, W. P. Rowe, Studies of small DNA viruses found in various adenovirus preparations: physical, biological, and immunological characteristics. *Proc Natl Acad Sci U S A* **55**, 1467-1474 (1966).

12. G. J. Logan, A. P. Dane, C. V. Hallwirth, C. M. Smyth, E. E. Wilkie, A. K. Amaya, E. Zhu, N. Khandekar, S. L. Ginn, S. H. Y. Liao, S. C. Cunningham, N. Sasaki, M. Cabanes-Creus, P. P. L. Tam, D. W. Russell, L. Lisowski, I. E. Alexander, Identification of liver-specific enhancer-promoter activity in the 3' untranslated region of the wild-type AAV2 genome. *Nat Genet* **49**, 1267-1273 (2017).
13. C. Summerford, R. J. Samulski, Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions. *J Virol* **72**, 1438-1445 (1998).
14. Q. Xie, W. Bu, S. Bhatia, J. Hare, T. Somasundaram, A. Azzi, M. S. Chapman, The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. *Proc Natl Acad Sci U S A* **99**, 10405-10410 (2002).
15. A. Kern, K. Schmidt, C. Leder, O. J. Muller, C. E. Wobus, K. Bettinger, C. W. Von der Lieth, J. A. King, J. A. Kleinschmidt, Identification of a heparin-binding motif on adeno-associated virus type 2 capsids. *J Virol* **77**, 11072-11081 (2003).
16. M. Schmidt, L. Govindasamy, S. Afione, N. Kaludov, M. Agbandje-McKenna, J. A. Chiorini, Molecular characterization of the heparin-dependent transduction domain on the capsid of a novel adeno-associated virus isolate, AAV(VR-942). *J Virol* **82**, 8911-8916 (2008).
17. S. R. Opie, K. H. Warrington, Jr., M. Agbandje-McKenna, S. Zolotukhin, N. Muzyczka, Identification of amino acid residues in the capsid proteins of adeno-associated virus type 2 that contribute to heparan sulfate proteoglycan binding. *J Virol* **77**, 6995-7006 (2003).
18. D. Grimm, J. S. Lee, L. Wang, T. Desai, B. Akache, T. A. Storm, M. A. Kay, In vitro and in vivo gene therapy vector evolution via multispecies interbreeding and retargeting of adeno-associated viruses. *J Virol* **82**, 5887-5911 (2008).
19. L. Lisowski, A. P. Dane, K. Chu, Y. Zhang, S. C. Cunningham, E. M. Wilson, S. Nygaard, M. Grompe, I. E. Alexander, M. A. Kay, Selection and evaluation of clinically relevant AAV variants in a xenograft liver model. *Nature* **506**, 382-386 (2014).
20. C. S. Manno, G. F. Pierce, V. R. Arruda, B. Glader, M. Ragni, J. J. Rasko, M. C. Ozelo, K. Hoots, P. Blatt, B. Konkle, M. Dake, R. Kaye, M. Razavi, A. Zajko, J. Zehnder, P. K. Rustagi, H. Nakai, A. Chew, D. Leonard, J. F. Wright, R. R. Lessard, J. M. Sommer, M. Tigges, D. Sabatino, A. Luk, H. Jiang, F. Mingozzi, L. Couto, H. C. Ertl, K. A. High, M. A. Kay, Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. *Nat Med* **12**, 342-347 (2006).
21. G. P. Gao, M. R. Alvira, L. Wang, R. Calcedo, J. Johnston, J. M. Wilson, Novel adeno-associated viruses from rhesus monkeys as vectors for human gene therapy. *Proc Natl Acad Sci U S A* **99**, 11854-11859 (2002).
22. A. M. Davidoff, J. T. Gray, C. Y. Ng, Y. Zhang, J. Zhou, Y. Spence, Y. Bakar, A. C. Nathwani, Comparison of the ability of adeno-associated viral vectors pseudotyped with serotype 2, 5, and 8 capsid proteins to mediate efficient transduction of the liver in murine and nonhuman primate models. *Mol Ther* **11**, 875-888 (2005).
23. A. C. Nathwani, J. T. Gray, J. McIntosh, C. Y. Ng, J. Zhou, Y. Spence, M. Cochrane, E. Gray, E. G. Tuddenham, A. M. Davidoff, Safe and efficient transduction of the liver after peripheral vein infusion of self-complementary AAV vector results in stable therapeutic expression of human FIX in nonhuman primates. *Blood* **109**, 1414-1421 (2007).
24. A. C. Nathwani, J. T. Gray, C. Y. Ng, J. Zhou, Y. Spence, S. N. Waddington, E. G. Tuddenham, G. Kemball-Cook, J. McIntosh, M. Boon-Spijker, K. Mertens, A. M. Davidoff, 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 (2006).

25. H. Azuma, N. Paulk, A. Ranade, C. Dorrell, M. Al-Dhalimy, E. Ellis, S. Strom, M. A. Kay, M. Finegold, M. Grompe, Robust expansion of human hepatocytes in Fah-/-/Rag2-/-/Il2rg-/- mice. *Nat Biotechnol* **25**, 903-910 (2007).
26. N. K. Paulk, K. Pekrun, E. Zhu, S. Nygaard, B. Li, J. Xu, K. Chu, C. Leborgne, A. P. Dane, A. Haft, Y. Zhang, F. Zhang, C. Morton, M. B. Valentine, A. M. Davidoff, A. C. Nathwani, F. Mingozzi, M. Grompe, I. E. Alexander, L. Lisowski, M. A. Kay, Bioengineered AAV Capsids with Combined High Human Liver Transduction In Vivo and Unique Humoral Seroreactivity. *Mol Ther* **26**, 289-303 (2018).
27. M. Cabanes-Creus, A. Westhaus, R. G. Navarro, G. Baltazar, E. Zhu, A. K. Amaya, S. H. Y. Liao, S. Scott, E. Sallard, K. L. Dilworth, A. Rybicki, M. Drouyer, C. V. Hallwirth, A. Bennett, G. Santilli, A. J. Thrasher, M. Agbandje-McKenna, I. E. Alexander, L. Lisowski, Attenuation of Heparan Sulfate Proteoglycan Binding Enhances In Vivo Transduction of Human Primary Hepatocytes with AAV2. *Mol Ther Methods Clin Dev* **17**, 1139-1154 (2020).
28. J. C. Nault, S. Datta, S. Imbeaud, A. Franconi, M. Mallet, G. Couchy, E. Letouze, C. Pilati, B. Verret, J. F. Blanc, C. Balabaud, J. Calderaro, A. Laurent, M. Letexier, P. Bioulac-Sage, F. Calvo, J. Zucman-Rossi, Recurrent AAV2-related insertional mutagenesis in human hepatocellular carcinomas. *Nat Genet* **47**, 1187-1193 (2015).
29. T. La Bella, S. Imbeaud, C. Peneau, I. Mami, S. Datta, Q. Bayard, S. Caruso, T. Z. Hirsch, J. Calderaro, G. Morcrette, C. Guettier, V. Paradis, G. Amaddeo, A. Laurent, L. Possenti, L. Chiche, P. Bioulac-Sage, J. F. Blanc, E. Letouze, J. C. Nault, J. Zucman-Rossi, Adeno-associated virus in the liver: natural history and consequences in tumour development. *Gut* **69**, 737-747 (2019).
30. C. L. Chen, R. L. Jensen, B. C. Schnepp, M. J. Connell, R. Shell, T. J. Sferra, J. S. Bartlett, K. R. Clark, P. R. Johnson, Molecular characterization of adeno-associated viruses infecting children. *J Virol* **79**, 14781-14792 (2005).
31. S. C. Cunningham, A. P. Dane, A. Spinoulas, G. J. Logan, I. E. Alexander, Gene delivery to the juvenile mouse liver using AAV2/8 vectors. *Mol Ther* **16**, 1081-1088 (2008).
32. N. Purohit, A. Jain, V. Mathews, G. R. Jayandharan, Molecular characterization of novel Adeno-associated virus variants infecting human tissues. *Virus Res*, 197716 (2019).
33. T. F. Lerch, M. S. Chapman, Identification of the heparin binding site on adeno-associated virus serotype 3B (AAV-3B). *Virology* **423**, 6-13 (2012).
34. F. Zhang, J. Aguilera, J. M. Beaudet, Q. Xie, T. F. Lerch, O. Davulcu, W. Colon, M. S. Chapman, R. J. Linhardt, Characterization of interactions between heparin/glycosaminoglycan and adeno-associated virus. *Biochemistry* **52**, 6275-6285 (2013).
35. A. Srivastava, E. W. Lusby, K. I. Berns, Nucleotide sequence and organization of the adeno-associated virus 2 genome. *J Virol* **45**, 555-564 (1983).
36. M. Nonnenmacher, T. Weber, Intracellular transport of recombinant adeno-associated virus vectors. *Gene Ther* **19**, 649-658 (2012).
37. V. Cagno, E. D. Tseligka, S. T. Jones, C. Tapparel, Heparan Sulfate Proteoglycans and Viral Attachment: True Receptors or Adaptation Bias? *Viruses* **11**, (2019).
38. O. S. Gorbatyuk, K. H. Warrington, Jr., M. S. Gorbatyuk, I. Zolotukhin, A. S. Lewin, N. Muzyczka, Biodistribution of adeno-associated virus type 2 with mutations in the capsid that contribute to heparan sulfate proteoglycan binding. *Virus Res* **274**, 197771 (2019).

39. M. C. Meneghetti, A. J. Hughes, T. R. Rudd, H. B. Nader, A. K. Powell, E. A. Yates, M. A. Lima, Heparan sulfate and heparin interactions with proteins. *J R Soc Interface* **12**, 0589 (2015).
40. L. A. Shackelton, C. R. Parrish, U. Truyen, E. C. Holmes, High rate of viral evolution associated with the emergence of carnivore parvovirus. *Proc Natl Acad Sci U S A* **102**, 379-384 (2005).
41. L. A. Shackelton, E. C. Holmes, Phylogenetic evidence for the rapid evolution of human B19 erythrovirus. *J Virol* **80**, 3666-3669 (2006).
42. R. A. Schwartz, C. T. Carson, C. Schuberth, M. D. Weitzman, Adeno-associated virus replication induces a DNA damage response coordinated by DNA-dependent protein kinase. *J Virol* **83**, 6269-6278 (2009).
43. R. Sanjuan, P. Domingo-Calap, Mechanisms of viral mutation. *Cell Mol Life Sci* **73**, 4433-4448 (2016).
44. R. Andino, E. Domingo, Viral quasispecies. *Virology* **479-480**, 46-51 (2015).
45. E. A. Rutledge, C. L. Halbert, D. W. Russell, Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2. *J Virol* **72**, 309-319 (1998).
46. K. Pekrun, G. De Alencastro, Q. J. Luo, J. Liu, Y. Kim, S. Nygaard, F. Galivo, F. Zhang, R. Song, M. R. Tiffany, J. Xu, M. Hebrok, M. Grompe, M. A. Kay, Using a barcoded AAV capsid library to select for clinically relevant gene therapy vectors. *JCI Insight* **4**, (2019).
47. A. P. Dane, S. C. Cunningham, N. S. Graf, I. E. Alexander, Sexually dimorphic patterns of episomal rAAV genome persistence in the adult mouse liver and correlation with hepatocellular proliferation. *Mol Ther* **17**, 1548-1554 (2009).
48. X. Xiao, J. Li, R. J. Samulski, Production of high-titer recombinant adeno-associated virus vectors in the absence of helper adenovirus. *J Virol* **72**, 2224-2232 (1998).
49. I. F. Khan, R. K. Hirata, D. W. Russell, AAV-mediated gene targeting methods for human cells. *Nat Protoc* **6**, 482-501 (2011).
50. M. Cabanes-Creus, S. L. Ginn, A. K. Amaya, S. H. Y. Liao, A. Westhaus, C. V. Hallwirth, P. Wilmott, J. Ward, K. L. Dilworth, G. Santilli, A. Rybicki, H. Nakai, A. J. Thrasher, A. C. Filip, I. E. Alexander, L. Lisowski, Codon-Optimization of Wild-Type Adeno-Associated Virus Capsid Sequences Enhances DNA Family Shuffling while Conserving Functionality. *Mol Ther Methods Clin Dev* **12**, 71-84 (2019).
51. J. Koster, S. Rahmann, Snakemake-a scalable bioinformatics workflow engine. *Bioinformatics* **34**, 3600 (2018).
52. L. Wall, S. Hay, The Perl Programming Language, version 5.26. (2017).
53. A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F. T. Heer, T. A. P. de Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res* **46**, W296-w303 (2018).
54. M. Carrillo-Tripp, C. M. Shepherd, I. A. Borelli, S. Venkataraman, G. Lander, P. Natarajan, J. E. Johnson, C. L. Brooks, 3rd, V. S. Reddy, VIPERdb2: an enhanced and web API enabled relational database for structural virology. *Nucleic Acids Res* **37**, D436-442 (2009).
55. P. Emsley, B. Lohkamp, W. G. Scott, K. Cowtan, Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* **66**, 486-501 (2010).
56. Z. Yang, K. Lasker, D. Schneidman-Duhovny, B. Webb, C. C. Huang, E. F. Pettersen, T. D. Goddard, E. C. Meng, A. Sali, T. E. Ferrin, UCSF Chimera, MODELLER, and IMP: an integrated modeling system. *J Struct Biol* **179**, 269-278 (2012).

57. C. Xiao, M. G. Rossmann, Interpretation of electron density with stereographic roadmap projections. *J Struct Biol* **158**, 182-187 (2007).
58. S. Scott. (Code Ocean, 2020).

Acknowledgements

We thank CMRI Vector and Genome Engineering Facility for help in vector preparation, Rohan Shah from CMRI Cancer Data Science Group for consultations on the statistical analysis and the Cytometry Facility of the Westmead Institute for Medical Research for help with sorting of human/murine hepatocytes. We also would like to thank all the members of the CMRI Bioresources, with special thanks to Shelley Dimech.

Funding:
This work was supported by project grants from Australian National Health and Medical Research Council (NHMRC) to L.L. and I.E.A. (APP1108311, APP1156431 and APP1161583). The work of L.L. was also supported by research grants from the Department of Science and Higher Education of Ministry of National Defense, Republic of Poland ("Kościuszko" k/10/8047/DNiSW/T – WIHE/3) and from the National Science Centre, Republic of Poland *(OPUS 13) (UMO-2017/25/B/NZ1/02790)*. I.E.A. was also supported by Discovery Project (DP150101253) grant from the Australian Research Council (ARC). AJT is a Wellcome Trust Principal Research Fellow. GS and AJT were supported by the NIHR Biomedical Research Centre at Great Ormond Street Hospital. M.A.M. and A.B. were supported by NSF grant DMS 1563234.

Author contributions:
M.C.-C., C.V.H., A.W., I.E.A. and L.L designed the experiments. M.C.-C, S.H.Y.L, E.Z. and R.G.N. performed the in vivo and in vitro vector comparisons and the in vivo replicative adaptations of RC-AAV2. C.V.H. isolated the AAV-hu.Lvr variants from human liver tissue. A.W. performed the tissue culture adaptation of AAV-hu.Lvr07 and M.C.-C and G.B. performed the tissue culture adaptation of AAV-hu.Lvr02. B.H.N., G.B. and M.C.-C generated the chromatography data. A.B. and M.A.M generated the structural visualizations. S.S. generated the script to identify barcodes corresponding to AAV variants. M.C.-C., C.V.H., A.W., B.H.N, S.H.Y.L, E.Z., R.G.N., G.B., S.S., A.B., generated reagents, protocols, performed and analyzed data. M.C.-C., C.V.H., M.D., I.E.A. and L.L. wrote the article and generated the figures. All authors reviewed, edited, and commented on the article.

Competing interests:
M.C-C., C.V.H., S.L.G., I.E.A. and L.L. are inventors on patent applications filed by Children's Medical Research Institute related to AAV capsid sequences and in vivo function of novel AAV variants (PCT/AU2020/050703, Title: Methods and AAV vectors for in vivo transduction). L.L. is a founder and scientific advisor of LogicBio Therapeutics and is the scientific founder of Reformer Therapeutics. L.L. and I.A.E. are co-founders of Exigen Biotherapeutics. A.J.T. is a co-founder and scientific consultant for Orchard Therapeutics, consultant for Rocket Pharmaceuticals, Generation bio, bluebird bio, 4Bio Capital Partners, and Sana Biotechnology. G.J.L. consults for Gyroscope Therapeutics. L.L. and I.A.E. have consulted on technologies addressed in this paper. L.L. has stocks and/or equity in LogicBio Therapeutics, Reformer Therapeutics, Orphinic Scientific and Exigen Biotherapeutics. I.A.E.

has stock and/or equity in Gyroscope Therapeutics and Exigen Biotherapeutics. All other authors declare no conflicts of interest.

Data and materials availability:
All data associated with this study are in the paper or supplementary materials. Computer code has been archived at Code Ocean [DOI: 10.24433/CO.7176285.v1](*58*) DNA sequence data for AAV variants AAV-hu.Lvr01 to AAV-hu.Lvr07) has been deposited in GenBank (Accession numbers MT709008-MT709014).

Figure Legends

Figure 1. AAV capsids isolated from human liver transduced human hepatocytes in vivo with high efficiency (**A**) Phylogram depicting the deduced evolutionary relationships among human liver-derived (AAV-hu.Lvr01 to 07) and related AAV capsid sequences (AAV1, 2, 6, 3B, 8 and 13). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. (**B-C**) In vivo comparison of physical (**B**) and functional (**C**) transduction of AAV-hu.Lvr capsids in the xenograft liver of the hFRG mouse model. Each AAV variant was used to package six unique barcoded ssAAV-LSP1-GFP-BC$^{(1-6)}$-WPRE-BGHpA cassettes and an equimolar mix of all eleven variants was used. Percentage of NGS reads mapped to each capsid in human hepatocytes at the DNA (cell entry, physical transduction, average of 877 vg / diploid human genome) (**B**) and cDNA (expression, functional transduction) level (**C**), normalized to the pre-injection mix, is shown. Statistical significance was calculated by comparing the performance of each AAV-hu.Lvr variant with its closest parental variant (i.e. hu.Lvr02 with AAV2). (**D**) Representative immunohistochemical analysis of hFRG mouse liver transduced with indicated variants. Purple: human GAPDH; green: vector-expressed GFP; blue: DAPI (nuclei). Scale=100 μm. (**E**) Quantification of the percentage of transduced human hepatocytes per human cluster, n=10 human clusters / mouse, N=2 mice / vector for AAV2, hu.Lvr01-02, hu.Lvr06-7 and N=1 mouse / vector for the rest. (**F**) Results of in vitro heparin competition assay, multiplicity of infection (MOI) = 100,000 vector genomes / cell.

AAV2 served as a control. Data are mean ± SD. Statistical significance was calculated using the two-tailed Mann-Whitney test (* P value ≤ 0.05; ** P value ≤ 0.01; **** P ≤0.0001, n.s. P value > 0.05).

Figure 2. In vitro adaptation of AAV-hu.Lvr07 attenuated in vivo transduction of human hepatocytes (**A**) Graphic representation of the in vitro evolution approach. (**B-C**) Population shift of hu.Lvr07 E530 and G593 during in vitro adaptation. Next-generation sequencing was used to determine the proportion of variants harboring mutations at hu.Lvr07 *cap* nucleotide position 1663 (leading to an E530K substitution) (**B**) and at nucleotide 1777 (leading to a G593R substitution) (**C**). PL, plasmid; RC, replication-competent; R1-R4, rounds 1 to 4. (**D**) In vitro heparin competition assay using HuH-7 cells transduced with culture-adapted variants of AAV-hu.Lvr07 expressing eGFP, with and without soluble heparin in the media. MOI = 40,000 vector genomes / cell. AAV2, AAV3B and AAV13 served as controls. Data are mean ± SD. Statistical significance was calculated using the two-tailed Mann-Whitney test, comparing the in vitro performance of AAV-hu.Lvr07 and the culture-isolated variants as indicated (* P value ≤ 0.05). **(E**) Surface map and (**F**) stereographic roadmap projection of a 3D model of AAV-hu.Lvr07 viewed down the icosahedral 2-fold axis. The viral asymmetric unit is shown as a black triangle and bounded by 2-, 3-, and 5-fold axes, depicted by an oval, triangle, and pentagon respectively, in both the surface map and the stereographic roadmap. The HSPG binding residues K532, R484 and R487 are colored blue, E530 is colored cyan and G593 is colored marine blue. **(G-H**) In vivo comparison of physical (DNA) and functional (RNA/cDNA) transduction of AAV-hu.Lvr07 and its culture adapted version (ca-hu.Lvr07-E530K) capsids in the xenograft liver model. AAV2, AAV8 and AAV13 were used as controls. Each AAV variant was used to package six unique barcoded cassettes as described before. Percentage of NGS reads mapped to each capsid in human hepatocytes at the DNA (cell

entry, physical transduction, average-of 611.97 vgs / diploid human genome) (**G**) and cDNA (expression, functional transduction) level (**H**), normalized to the pre-injection mix, is shown. Data are mean ± SD. Statistical significance was calculated using the two-tailed Mann-Whitney test by comparing the performance hu.Lvr07 with its culture adapted version (** P value ≤ 0.01).

Figure 3. HSPG-detargeting mutations accumulated upon in vivo replication of prototypical AAV2 (**A**) Schematic representation of the in vivo evolution approach. (**B-D**) Population shift of RC-AAV2 during in vivo adaptation experiments. The proportion of AAV2 with acquired mutations leading to amino acid changes at positions R487 (**B**), R588 (**C**) for adaptation experiment #1**,** and R588 (**D**) for experiment #2. RC: replication-competent; R1-R3: rounds 1 to 3. (**E**) Surface map and (**F**) stereographic roadmap projection of a 3D model of evolved AAV2 variants viewed down the icosahedral 2-fold axis. The HSPG binding residues R484, K532 and R585 (blue), R487 (light green) and R588 (yellow) are shown. (**G**) In vitro heparin competition assay. MOI = 100,000 vector genomes / cell. AAV2 served as a control. Data are represented as mean±SD. Statistical significance was calculated using the two-tailed Mann-Whitney test by comparing vector performance with and without the presence of soluble heparin (* P value ≤ 0.05; n.s. P value > 0.05).

Figure 4. AAV2 variants with mutations leading to HSPG-detargeting transduced human hepatocytes in vivo with higher efficiency than prototypical AAV2 (**A-B**) In vivo physical transduction (**A**) and functional transduction (**B**) comparison of AAV2 variants in the xenograft liver model. Each AAV variant was used to package six barcoded ssAAV-LSP1-GFP-BC$^{(1-6)}$-WPRE-BGHpA constructs and an equimolar mix of all seven variants was used in the study. AAV2, AAV8, AAV-LK03 and AAV13 served as controls. Percentage of NGS reads

mapped to each capsid in human cells at the DNA (average of 438 vgs / diploid human genome) (**A**) and RNA/cDNA levels (**B**), normalized to the pre-injection mix, is shown. (**C**) Representative immunohistochemical analysis of hFRG mouse liver transduced with indicated variants. Purple: human GAPDH; green: vector-expressed GFP; blue: DAPI (nuclei). Scale=100 μm. (**D**) Quantification of the percentage of transduced human hepatocytes per human cluster. Data for control serotypes AAV2, AAV13, AAV8 and AAV-LK03 from **Fig. 1E** are shown again for reference. Data shown as mean±SD, n=10 human clusters / mouse, n=1 mouse / vector for AAV2 variants, and N=2 mice for AAV2. Statistical significance was calculated using the two-tailed Mann-Whitney test (** P value ≤ 0.01; **** P ≤ 0.0001).

Table 1. Summary of HiTrap Heparin Column binding studies.

AAV Capsid	AAV predominantly detected in:	[NaCl] at elution peak maxima (mM)
hu.Lvr01-07	Flow Through	(-)
caLvr07^{E530K}	Elution Fraction	320.89
caLvr07^{G593R}	Elution Fraction	341.94
cahu.Lvr02$^{D469N+E555K}$	Elution Fraction	291.8
AAV2	Elution Fraction	454.34
AAV2-R588I	Elution Fraction	58.25
AAV2-R588T	Elution Fraction	98.09
AAV2-R487Q	Elution Fraction	175.97
AAV3B	Elution Fraction	278.69
AAV13	Elution Fraction	331.89

Table 2 Sequence alignment of AAV HSPG binding residues. Underlined are residues important for serotype-specific heparin binding. The numbers in parentheses indicate the sequence and structure-equivalent residue compared to AAV2.

Residue #	AAV2	AAV3b	AAV13	hu.Lvr07
484	R	R	R	R
487	R	R	R	R
530	E	E	K(528)	E
532	K	K	K(530)	K
585	R	S(586)	N(583)	N
588	R	T(589)	A(586)	T
593	A	R(594)	G(591)	G