Promoterless gene targeting without nucleases ameliorates haemophilia B in mice

A. Barzel1, N. K. Paulk1, Y. Shi2, Y. Huang1, K. Chu1, F. Zhang1, P. N. Valdmanis1, L. P. Spector1, M. H. Porteus3, K. M. Gaensler2 & M. A. Kay1

Site-specific gene addition can allow stable transgene expression for gene therapy. When possible, this is preferred over the use of promiscuously integrating vectors, which are sometimes associated with clonal expansion1 and oncogenesis2. Site-specific endonucleases that can induce high rates of targeted genome editing are finding increasing applications in biological discovery and gene therapy3. However, two safety concerns persist: endonuclease-associated adverse effects, both on-target4 and off-target5;6;7, and oncogene activation caused by promoter integration, even without nucleases8. Here we perform recombinant adeno-associated virus (rAAV)-mediated promoterless gene targeting without nucleases and demonstrate amelioration of the bleeding diathesis in haemophilia B mice. In particular, we target a promoterless human coagulation factor IX (F9) gene to the liver-expressed mouse albumin (Alb) locus. F9 is targeted, along with a preceding 2A-peptide coding sequence, to be integrated just upstream to the Alb stop codon. While F9 is fused to Alb at the DNA and RNA levels, two separate proteins are synthesized by way of ribosomal skipping. Thus, F9 expression is linked to robust hepatic albumin expression without disrupting it. We injected an AAV8-F9 vector into neonatal and adult mice and achieved on-target integration into ~0.5% of the albumin alleles in hepatocytes. We established that F9 was produced only from on-target integration, and ribosomal skipping was highly efficient. Stable F9 plasma levels at 7–20% of normal were obtained, and treated F9-deficient mice had normal coagulation times. In conclusion, transgene integration as a 2A-fusion to a highly expressed endogenous gene may obviate the requirement for nucleases and/or vector-borne promoters. This method may allow for safe and efficacious gene targeting in both infants and adults by greatly diminishing off-target effects while still providing therapeutic levels of expression from integration.

Site-specific gene targeting is one of the fastest growing fields in gene therapy and genome engineering. The rise in popularity of gene targeting can be attributed in large part to the development of readily engineered and easy to use site-specific endonucleases (for example, TAL- or CRISPR-based9) that can increase rates of gene disruption, gene correction or gene addition by as much as four orders of magnitude. However, these endonucleases may have significant adverse effects including immunogenicity, uncontrolled DNA damage response, off-target cleavage and mutagenesis, induction of chromosomal aberrations, as well as off-target integration of the transgene and endonuclease vectors (if DNA-based10-12). When a vector-borne promoter, driving expression of the therapeutic transgene and/or the nuclease, is integrated either on- or off-target, it may lead to undesired activation of nearby genes, including oncogenes. The use of endonucleases in vivo would require their vectorization, delivery and expression in a transient manner to minimize long-term side effects. It is unclear how integration of the vectored endonuclease gene could be strictly avoided.

Our promoterless, endonuclease-independent method harnesses the efficient transduction, favourable safety profile and high gene targeting rates associated with rAAV9-11, as well as the robust liver-specific expression of the Alb locus12. Different rAAV serotypes can efficiently transduce various cell types in vitro or in vivo, while other serotypes have been designed or selected for desired phenotypes13-17. rAAV is in use in several clinical trials18,19. Notably, rAAV transduction allows high gene targeting rates in vitro9 and in vivo20,21. The increased recombination rates may be due to the viral inverted terminal repeats, the encapsidation of single-stranded DNA, or the timing and subcellular localization of capsid uncoating.

The safety of rAAV stems from its lack of pathogenicity, as well as being devoid of viral genes. Nevertheless, non-targeted genomic integration of rAAV occurs at a low but notable rate, and there are reports of such integrations inducing hepatocellular carcinoma in mice22. Transformation was attributed to vector integration at a chromosome 12 loci for imprinted genes and small RNAs. Integration of rAAV-borne promoters may be the leading cause of aberrant expression, as was established for lentiviral and retroviral vector integration leading to clonal expansion and oncogenesis23,24. Vector-borne promoters are in use in many continuing clinical trials. By contrast, the rAAV used in our strategy encodes no promoter, thus diminishing the chance of neighbouring oncogene activation in rare off-target integrations.

As proof of concept, we targeted the human F9 gene, deficient in the X-linked recessive disease haemophilia B, which affects 1 in 30,000 males. Affected individuals suffer from serious spontaneous bleeding owing to a deficiency of plasma coagulation Factor F9 produced from the liver. Reconstitution with as little as 1–2% clotting factor can considerably improve quality of life, while 5–20% will markedly ameliorate the bleeding diathesis. Here we used the liver tropic rAAV8 serotype to target human F9 for expression after integration from the robust liver-specific mouse Alb promoter. We postulated that: the Alb promoter should allow high levels of coagulation factor production even if integration takes place in only a small fraction of hepatocytes; and the high transcriptional activity at the Alb locus might make it more susceptible to transgene integration by homologous recombination.

Gene targeting without nucleases should affect only a small fraction of Alb alleles in the liver. Nevertheless, we opted to minimize disruption and dysregulation of the Alb gene by targeting F9 as a 2A-fusion at the end of the Alb reading frame (Fig. 1a). 2A-peptides, derived from plus-strand RNA viruses, allow the production of several proteins from a single reading frame by means of ribosomal skipping25. This process leaves the first translated protein tagged with ~20 carboxy-terminal amino acids, and the second protein with just one additional amino-terminal proline. Functionality of both proteins is typically retained, and clinical trials using 2A-peptides did not report immunogenicity26. We used single-stranded AAV to target a codon-optimized F9 coding sequence, preceded by a sequence coding for a porcine teschovirus-1 2A-peptide (P2A)27, to be integrated just 5′ of the Alb stop codon. After integration, Alb and F9 are co-transcribed from the strong Alb promoter, and should thus be co-regulated at the levels of splicing, nuclear export, messenger RNA stability, translation initiation and endoplasmic reticulum localization. Two separate proteins are translated, both containing a signal

1Departments of Pediatrics and Genetics, 269 Campus Drive, CCSB Building, Room 2105, Stanford, California 94305-5164, USA. 2Department of Medicine, Box 1270, UCSF, San Francisco, California 94143-1270, USA. 3Department of Pediatrics, 269 Campus Drive, Lorry Lokey Stem Cell Research Building, Room G3045, Stanford, California 94305-5164, USA.

LETTER

doi:10.1038/nature13864

©2015 Macmillan Publishers Limited. All rights reserved
Plasmid F9
AAV inverse
RESEARCH

(3.5 ng ml⁻¹ detection was 20 ng ml⁻¹ or inverse control (ELISA after tail vein injections of 9-week-old female B6 mice with 1 vector genome per mouse of the AAV8-3 F9 measured by ELISA following intraperitoneal injections of 2-day-old B6 mice with 2.5 × 10¹³ vector genomes per mouse (~1.25 × 10¹⁴ per kg) of a rAAV8 coding for the human F9 targeting cassette or a vector with an inverted cassette, controlling for off-target expression (Fig. 1b). The fragment inverted in the control with respect to the Alb homology arms includes not only the F9 gene, but also the P2A coding sequence, the adjacent Alb exon, and the preceding splice junction. The inverse control should not allow notable F9 expression after on-target integration, but would allow levels of off-target expression similar to that from the experimental construct. We measured plasma F9 protein levels each week by enzyme-linked immunosorbent assay (ELISA), starting at 4 weeks of life (Fig. 2a). For the experimental group, levels of plasma F9 expression further, our vector has neither an ATG start codon before the F9 signal peptide, nor a start codon in the 2A-peptide coding sequence or preceding Alb exon.

First, we performed intraperitoneal injections of 2-day-old C57BL/6 (B6) mice with 2.5 × 10¹³ vector genomes per mouse (~1.25 × 10¹⁴ per kg) of a rAAV8 coding for the human F9 targeting cassette or a vector with an inverted cassette, controlling for off-target expression (Fig. 1b). The fragment inverted in the control with respect to the Alb homology arms includes not only the F9 gene, but also the P2A coding sequence, the adjacent Alb exon, and the preceding splice junction. The inverse control should not allow notable F9 expression after on-target integration, but would allow levels of off-target expression similar to that from the experimental construct. We measured plasma F9 protein levels each week by enzyme-linked immunosorbent assay (ELISA), starting at 4 weeks of life (Fig. 2a). For the experimental group, levels of plasma F9 expression further, our vector has neither an ATG start codon before the F9 signal peptide, nor a start codon in the 2A-peptide coding sequence or preceding Alb exon.

For the experimental group, levels of plasma F9 expression further, our vector has neither an ATG start codon before the F9 signal peptide, nor a start codon in the 2A-peptide coding sequence or preceding Alb exon.

First, we performed intraperitoneal injections of 2-day-old C57BL/6 (B6) mice with 2.5 × 10¹³ vector genomes per mouse (~1.25 × 10¹⁴ per kg) of a rAAV8 coding for the human F9 targeting cassette or a vector with an inverted cassette, controlling for off-target expression (Fig. 1b). The fragment inverted in the control with respect to the Alb homology arms includes not only the F9 gene, but also the P2A coding sequence, the adjacent Alb exon, and the preceding splice junction. The inverse control should not allow notable F9 expression after on-target integration, but would allow levels of off-target expression similar to that from the experimental construct. We measured plasma F9 protein levels each week by enzyme-linked immunosorbent assay (ELISA), starting at 4 weeks of life (Fig. 2a). For the experimental group, levels of plasma F9 expression further, our vector has neither an ATG start codon before the F9 signal peptide, nor a start codon in the 2A-peptide coding sequence or preceding Alb exon.

First, we performed intraperitoneal injections of 2-day-old C57BL/6 (B6) mice with 2.5 × 10¹³ vector genomes per mouse (~1.25 × 10¹⁴ per kg) of a rAAV8 coding for the human F9 targeting cassette or a vector with an inverted cassette, controlling for off-target expression (Fig. 1b). The fragment inverted in the control with respect to the Alb homology arms includes not only the F9 gene, but also the P2A coding sequence, the adjacent Alb exon, and the preceding splice junction. The inverse control should not allow notable F9 expression after on-target integration, but would allow levels of off-target expression similar to that from the experimental construct. We measured plasma F9 protein levels each week by enzyme-linked immunosorbent assay (ELISA), starting at 4 weeks of life (Fig. 2a). For the experimental group, levels of plasma F9 expression further, our vector has neither an ATG start codon before the F9 signal peptide, nor a start codon in the 2A-peptide coding sequence or preceding Alb exon.

First, we performed intraperitoneal injections of 2-day-old C57BL/6 (B6) mice with 2.5 × 10¹³ vector genomes per mouse (~1.25 × 10¹⁴ per kg) of a rAAV8 coding for the human F9 targeting cassette or a vector with an inverted cassette, controlling for off-target expression (Fig. 1b). The fragment inverted in the control with respect to the Alb homology arms includes not only the F9 gene, but also the P2A coding sequence, the adjacent Alb exon, and the preceding splice junction. The inverse control should not allow notable F9 expression after on-target integration, but would allow levels of off-target expression similar to that from the experimental construct. We measured plasma F9 protein levels each week by enzyme-linked immunosorbent assay (ELISA), starting at 4 weeks of life (Fig. 2a). For the experimental group, levels of plasma F9 expression further, our vector has neither an ATG start codon before the F9 signal peptide, nor a start codon in the 2A-peptide coding sequence or preceding Alb exon.
F9 plateaued at 350–1,000 ng ml\(^{-1}\), which corresponds to 7–20% of normal. For the inverse control group, F9 plasma levels were at or below the level of detection (20 ng ml\(^{-1}\)), suggesting that in the experimental group, F9 expression does indeed originate from on-target integration. Importantly, F9 retains the original plasma protein levels after a two-thirds partial hepatectomy, a surgical procedure known to reduce episomal AAV transgene expression by >90%, further establishing stable transgene integration.

To determine whether liver growth, as seen with neonates, is essential for therapeutic levels of gene targeting, we targeted F9 to the Alb locus using the same vector in adult mice. Adult B6 mice were injected with 1 \(\times\) \(10^{12}\) vector genomes per mouse (\(\sim 5 \times 10^{13}\) vector genomes per kilogram) by tail vein with the AAV8-F9 vector, or the inverse control. A third group of control mice received hydrodynamic tail vein injections of a plasmid coding for the promotercless F9 construct in the ‘correct’ orientation. For the AAV8-F9 mice group, plasma F9 levels were found to be stable at 7–20% of normal (Fig. 2b). Vector injections at lower dose led to lower plasma F9 levels without reaching a plateau at the doses tested (Fig. 2c). For adults as well as neonates, the F9 plasma levels of the inverse control group were at or below the limit of detection. Diminished F9 plasma levels were also associated with mice hydrodynamically injected with plasmid (Fig. 2b). Thus, targeting is dependent on rAAV vectorization. Finally, we performed rAAV injections in adult F9 knock-out haemophilia B mice. The functional coagulation, as determined by the activated partial thromboplastin time in treated knockout mice, was restored to levels similar to that of wild-type mice (Fig. 2d). The F9 biological activity correlated with plasma protein levels of 709 ± 91 ng ml\(^{-1}\), similar to levels in wild-type mice (Fig. 2b–d).

F9 expression from the liver was confirmed by immunohistochemistry (Extended Data Fig. 1). Western blot analysis of liver samples detected F9 at the expected molecular mass, testifying that ribosomal skipping was efficient, and suggesting that both the ELISA and immunohistochemistry signals correspond to accurately processed F9 (Fig. 2d).

We opted to quantify targeting rates by quantitative PCR (qPCR). To avoid false signals from episomal rAAV, we first amplified a 3′ segment of the genomic Alb locus in a manner not affected by presence or absence of an integrated F9 sequence (Fig. 3a and Extended Data Fig. 2). The unbiased amplification was made possible by presence of a common restriction site at a roughly equal distance 3′ of the stop codon in targeted and wild-type alleles. We then used the PCR amplicon as a template for two different qPCR assays: one quantifying the abundance of targeted Alb alleles, and the other quantifying the abundance of untargeted wild-type alleles. In the liver, only hepatocytes are targeted by rAAV8 (ref. 24). Therefore, we conservatively corrected for a 70% hepatocyte frequency\(^{23}\) and found the rate of Alb alleles targeted by F9 to be 0.5% on average for mice injected as either neonates or adults at the highest dose (Fig. 3c and Extended Data Fig. 3). Actual rates of targeting in neonates and adults might differ because AAV distribution to the liver may vary based on the different methods used for vector infusion. We then examined the proportion of fused Alb-F9 mRNAs to wild-type Alb mRNAs by comparing two respective qPCR assays performed on an unbiased cDNA template (Fig. 3b). The proportion was found to be 0.1% on average for mice injected as either neonates or adults (Fig. 3c). This value tended to be lower than the rate of integration at the DNA level, although the difference was not statistically significant. It is possible that the production, processing and/or stability of chimaeric Alb-F9 mRNA transcripts were reduced compared to wild-type Alb mRNA. While AAV8 has been shown to target only hepatocytes in the mouse liver\(^{24}\), here we did not rule out the possibility that some integration occurred in non-parenchymal cells that do not express albumin. Our observed targeting rate is higher than other reports\(^{9,12,20}\), and is particularly noteworthy in adult mice in which non-proliferating hepatocytes were expected to allow for a low rate of homologous recombination.

We propose that the high expression rate at the Alb locus and the associated chromatin status may contribute to the high rates of targeting. Damage-induced proliferation cannot be strictly ruled out, but no increase in alanine transaminase (ALT) levels was seen after injection (Extended Data Fig. 4).

AAV genomes may be present in cells as episomes, or as on- or off-target integrants. Total vector copy number was assessed by qPCR (Extended Data Fig. 5). The minor change in vector copy number after partial hepatectomy in mice injected as neonates may suggest that episomal vectors had already been greatly diluted during normal liver growth and development. In which case, vector copy number can be used as an approximate lower bound on the rate of off-target to on-target integration. However, most importantly, in the absence of a vector-borne promoter, F9 should only be expressed from on-target integration. The reconstituted high F9 levels after partial hepatectomy (Fig. 2a) support this assumption, as only stably integrated transgenes could rebound after such a procedure, unlike that seen with transient episomal expression\(^{26}\). Lack of notable F9 plasma levels after treatment with the inverse control vector further demonstrated reduced off-target expression. We used quantitative reverse transcription PCR (qRT–PCR) to assess the ratio of fused Alb-F9 mRNAs directly among the total F9 mRNA pool (Fig. 4a). The ratio was found to be 1:1 for mice injected as neonates and as adults (Fig. 4b). This suggests that F9 is expressed almost exclusively from

---

**Figure 3** | Rate of Alb targeting at the DNA and RNA levels. a, Assessment of on-target integration rate begins using linear amplification with biotinylated primer 1 (black), annealing to the genomic locus but not to the vector. Linear amplicons are then bound to streptavidinylated beads and washed to exclude episomal vectors. Subsequent second-strand DNA synthesis with random primers was followed by CviQI restriction digestion. A compatible linker is then ligated, followed by two rounds of nested PCR (primers 2–3 in blue, and then primers 4–5 in red). CviQI cleaves at the same distance from the homology border in both targeted and wild-type alleles, thus allowing for unbiased amplification. The amplicons of the second nested PCR then serve as a template for qPCR assays with either primers 6–7 (green) or 8–9 (orange). b, For mRNA quantification, primers 10–11 or 11–12 were used to generate a cDNA for qPCR assays. Shape and fill code as in Fig. 1. c, Black bars represent the targeting rate of Alb alleles as the ratio between the abundance of the DNA template amplified by primers 6–7 to the abundance of the DNA template amplified by primers 8–9, corrected by a factor of 0.7 to account for hepatocyte frequency. Grey bars represent the expression rate of targeted Alb alleles as the ratio between the abundance of the cDNA template amplified by primers 10–11 to the abundance of the cDNA template amplified by primers 11–12. n = 3 for each group, error bars represent s.d.
on-target integration. Indeed, the only specific signal from a northern blot with a P2A probe corresponded to the expected fused Alb-P2A-F9 mRNA (Fig. 4c). Finally, a western blot with an anti-2A-peptide antibody indicated that the 2A-peptide is associated with a single species at the expected molecular mass of Alb (Fig. 4d), as would be expected only if expression was restricted to on-target integration and followed by efficient ribosomal skipping.

rAAV has become a popular vector for clinical therapy. Although the period of transgene expression in adults can last for several years, it is not yet known whether lifelong expression, as required for many genetic disorders, can be obtained with routine promoter-containing vectors. Episomal expression from AAV vectors is rapidly lost in dividing cells, even after just one round of cell division.26 This makes it likely that diseases that induce regeneration and/or are treated in infancy while tissues continue to grow, will have limited durability. Secondary infusion of an AAV vector will be unlikely to result in a successful transduction, owing to the robust humoral immunity resulting from primary vector administration.27 By contrast, our approach results in vector insertion of an AAV vector will be unlikely to result in a successful transduction, owing to the robust humoral immunity resulting from primary vector administration.27 By contrast, our approach results in vector insertion of an AAV vector will be unlikely to result in a successful transduction, owing to the robust humoral immunity resulting from primary vector administration.27 By contrast, our approach results in vector insertion of an AAV vector will be unlikely to result in a successful transduction, owing to the robust humoral immunity resulting from primary vector administration.27 By contrast, our approach results in vector insertion of an AAV vector will be unlikely to result in a successful transduction, owing to the robust humoral immunity resulting from primary vector administration.27 By contrast, our approach results in vector insertion of an AAV vector will be unlikely to result in a successful transduction, owing to the robust humoral immunity resulting from primary vector administration.27 By contrast, our approach results in vector insertion of an AAV vector will be unlikely to result in a successful transduction, owing to the robust humoral immunity resulting from primary vector administration.27 By contrast, our approach results in vector insertion of an AAV vector will be unlikely to result in a successful transduction, owing to the robust humoral immunity resulting from primary vector administration.27 By contrast, our approach results in vector insertion of an AAV vector will be unlikely to result in a successful transduction, owing to the robust humoral immunity resulting from primary vector administration.27

Previous work demonstrating targeting of F9 to a chimaeric locus in a transgenic mouse relied on the co-expression of nucleases that may be associated with immunological and genotoxic side effects. Interestingly, the same reliance on endonucleases held true even when F8 was targeted to the Alb locus in mice and non-human primates, probably because no homology arms were provided and integration relied instead on non-homologous end-joining. rAAV has already been used in clinical gene therapy trials to treat haemophilia B.64 However, the transgene in these clinical trials was expressed from a vector-borne promoter that might induce oncogene activation, as has been reported in mice.7 Measuring levels of alanine transaminases, we observed no liver toxicity with the injection of our human F9 targeting vector (Extended Data Fig. 4). However, it remains to be determined whether the transgene overexpression associated with our method will lead to toxicity when different therapeutic transgenes are targeted. 2A-induced immunogenicity could not be strictly ruled out, but notably no immune effects were reported in clinical trials when vector coding for a 2A peptide was targeted to lymphocytes.25 Although we found no evidence of off-target expression and no ALT increase, the high vector dose we used may lead to other undesired outcomes such as increased off-target integration and increased immunogenicity. In the future, this could be mitigated by the use of AAV serotypes having better tropism and/or by use of hyperactive F9 variants. Genetic polymorphisms at the target locus in the human patient population may lead to variable therapeutic efficacy owing to reduced homology. However, we found that ~95% of a 1000 Genomes Project (http://www.1000genomes.org) sample of the human population have no more than just two haplotypes at the relevant ALB sequence, which may enable broad applicability (Extended Data Table 1). Our work demonstrates a therapeutic effect for in vivo gene targeting without nucleases and without a vector-borne promoter. The favourable safety profile of our promoterless and nuclease-free gene targeting strategy for rAAV makes it a prime candidate for clinical assessment in the context of haemophilia and other genetic deficiencies. More generally, this strategy could be applied whenever the therapeutic effect is conveyed by a secreted protein (for example, broadly neutralizing antibodies) or when targeting confers a selective advantage.22

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 April; accepted 12 September 2014.

Published online 29 October 2014.

1. Aiuti, A. et al. Lentiviral hematopoietic stem cell gene therapy in patients with Wiskott–Aldrich syndrome. Science 341, 1233151 (2013).
2. Hacein-Bey-Abina, S. et al. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. J. Clin. Invest. 118, 3132–3142 (2008).
3. Rajagopalan, S. et al. Gensbach, C. A. & Barbas, C. F. III. ZFN, TALEN, and CRISPR-Cas-based methods for genome engineering. Trends Biotechnol. 31, 397–405 (2013).
4. Hendel, A. et al. Quantifying genome-editing outcomes at endogenou loci with SMRT sequencing. Cell Rep. 7, 293–305 (2014).
5. Cho, S. W. et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res. 24, 132–141 (2014).
6. Fu, Y. et al. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. Nature Biotechnol. 31, 822–826 (2013).
7. Donsante, A. et al. AAV vector integration sites in mouse hepatocellular carcinoma. Science 317, 477 (2007).
8. Lisowski, L. et al. Ribosomal DNA integrating rAAV-rDNA vectors allow for stable transgene expression. Mol. Ther. 20, 1912–1923 (2012).
9. Miller, D. G. et al. Gene targeting in vivo by adeno-associated virus vectors. Nature Methods 4, 1022–1026 (2006).
10. Li, H. et al. In vivo genome editing restores haemostasis in a mouse model of haemophilia. Nature 475, 217–221 (2011).
11. Shi, Y., Falahahi, R., Zhang, J., Flebbe-Rehwoldt, L. & Gansler, K. K. Role of antihypertensive-specific regulatory CD4 ‘CD29’ T cells in tolerance induction after neonatal IP administration of AAV-H-FX. Gene Ther. 20, 987–996 (2013).
12. Paull, N. K., Loza, L. M., Finegold, M. J. & Grompe, M. AAV-mediated gene targeting is significantly enhanced by transient inhibition of nonhomologous end joining or the proteasome in vivo. Hum. Gene Ther. 23, 658–665 (2012).
13. Garcia-Martinez, R. et al. Albumin: pathophysiologic basis of its role in the treatment of cirrhosis and its complications. Hepatology 58, 1836–1846 (2013).
14. Grimm, D. et al. In vitro and in vivo gene therapy vector evolution via multispecies interbreeding and retargeting of adeno-associated viruses. J. Virol. 82, 5887–5911 (2008).
15. Lisowski, L. et al. Selection and evaluation of clinically relevant AAV variants in a xenograft liver model. Nature 506, 382–386 (2014).
16. Li, C. et al. Single amino acid modification of adeno-associated virus capsid changes transduction and humoral immune profiles. J. Virol. 86, 7752–7759 (2012).
17. Daikara, D. et al. In vivo-directed evolution of a new adeno-associated virus for therapeutic outer retinal gene delivery from the vitreous. Sci. Transl. Med. 5, 189ra176 (2013).
18. Nathwani, A. C. et al. Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. N. Engl. J. Med. 365, 2357–2365 (2011).
19. MacLaren, R. E. et al. Retinal gene therapy in patients with choroideremia: initial findings from a phase 1/2 clinical trial. Lancet 383, 1129–1137 (2014).
20. Paulk, N. K. et al. Adeno-associated virus gene repair corrects a mouse model of hereditary tyrosinemia in vivo. Hepatology 51, 1200–1208 (2010).
21. Cavazza, A., Moiani, A. & Mavilio, F. Mechanisms of retroviral integration and mutagenesis. Hum. Gene Ther. 24, 119–131 (2013).
22. Kim, J. H. et al. High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice. PLoS ONE 6, e18556 (2011).

©2015 Macmillan Publishers Limited. All rights reserved

15 JANUARY 2015 | VOL 517 | NATURE | 363
23. Johnson, L. A. et al. Gene therapy with human and mouse T-cell receptors mediates cancer regression and targets normal tissues expressing cognate antigen. *Blood* 114, 535–546 (2009).

24. Nakai, H. et al. Unrestricted hepatocyte transduction with adeno-associated virus serotype 8 vectors in mice. *J. Virol.* 79, 214–224 (2005).

25. Si-Tayeb, K., Lemaigre, F. P. & Duncan, S. A. Organogenesis and development of the liver. *Dev. Cell* 18, 175–189 (2010).

26. Nakai, H. et al. Extrachromosomal recombinant adeno-associated virus vector genomes are primarily responsible for stable liver transduction in vivo. *J. Virol.* 75, 6969–6976 (2001).

27. Calcedo, R. & Wilson, J. M. Humoral immune response to AAV. *Front. Immunol.* 4, 341 (2013).

28. Anguela, X. M. et al. ZFN mediated targeting of albumin “safe harbor” results in therapeutic levels of human factor viii in a mouse model of hemophilia A. *Blood* 122, 720 (2013).

29. Yew, N. S. & Cheng, S. H. Gene therapy for lysosomal storage disorders. *Pediatr. Endocrinol. Rev.* 11 (suppl. 1), 99–109 (2013).

30. Balazs, A. B. et al. Antibody-based protection against HIV infection by vectored immunoprophylaxis. *Nature* 481, 81–84 (2011).

**Acknowledgements**

This work was supported by a grant to M.A.K. from the National Heart Lung & Blood Institute (R01-HL064274). A.B. was supported by a fellowship from the Lucile Packard Foundation for Children’s Health, Stanford NIH-NCATS-CTSA UL1 TR001085 and Child Health Research Institute of Stanford University. N.K.P. was supported by a fellowship from the National Heart Lung & Blood Institute (F32-HL119059), the Hans Popper Memorial Fellowship from the American Liver Foundation, and the Stanford Dean’s Fellowship. M.H.P. was supported by the Laurie Krauss Lacob Faculty Scholar Fund in Pediatric Translational Medicine. The funding organizations played no role in experimental design, data analysis or manuscript preparation.

**Author Contributions**

A.B., N.K.P., M.H.P., K.M.G. and M.A.K. designed the experiments. A.B., N.K.P., Y.S., Y.H., K.C., F.Z., P.N.V. and L.P.S. generated reagents and performed the experiments. A.B., N.K.P. and M.A.K. wrote and edited the manuscript.

**Author Information**

Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.A.K. (markay@stanford.edu).
METHODS

Plasmid construction. A mouse genomic Alb segment (90474003-90476720 in NCBI reference sequence, NC_000071.6) was PCR amplified and inserted between AAV2 inverted terminal repeats into BsrGI and SpeI restriction sites in a modified pTRF backbone31. The genomic segment spans 1.3 kilobases (kb) upstream and 1.4 kb downstream to the Alb stop codon. We then inserted into the BpuI restriction site an optimized P2A coding sequence preceded by a linker coding sequence (glycine–serine–glycine) and followed by an Nhel restriction site. Finally, we inserted a codon-optimized F9 coding sequence into the Nhel site to get a P2B269 that served in the construction of the rAAV8 vector. To construct the inverse control, we first amplified an internal segment from the BspW1 restriction site to the 3′ Nhel restriction site. PCR primers used had 15 base tails to allow subsequent integration of the ampiclon into a BspW1 and Nhel cleaved plasmid in the inverse orientation using an In-Fusion Kit (Clontech). Primers used were: forward, 5′-ATGCAAGGCCGACAGCTGGCATCAGCTTGCATCACC3′ and reverse, 5′-TTTAGGCTTAAGCTTACATGCTAATGCTGATCACC3′. Finally rAAV production plasmids were generated using an EndoFree Plasmid Megaprep Kit (Qiagen).

rAAV vector production and titration. rAAV vectors were produced as previously described using a Ca3(PO4)2 transfection protocol followed by CsCl gradient purification31. Vectors were titrated as previously described using a TaqMan qPCR assay with primers 10: 5′-GCCGGTCGAAAGCTGGATCAAAGAAA-3′, and 11: 5′-ATGCAAGGCCGACAGCTGGCATCAGCTTGCATCACC-3′. For non-injected controls, no qPCR signal was detected with primers 8–9. The abundance of template for each amplicon was calculated as the ratio between the abundance of the template for the 8–9 primer pair to abundance of template for the 10–11 primer pair. We then conservatively corrected this ratio by multiplying it by a factor of 0.7 to account for hepatocyte frequency in the samples32, as only hepatocytes are targeted by rAAV8 in the liver31.

Assessing rate of F9-containing Alb mRNAs by qPCR. cDNA produced from reverse transcription with a poly-dT primer served as a template for two different TaqMan qPCR assays. We quantified the abundance of wild-type Alb mRNA by qPCR with primers 10: 5′-CTGACAGGACACGCTCTC-3′ and 11: 5′-TGAACGTCTCAGCTTGCACACC3′. For non-injected controls, no qPCR signal was detected with primers 8–9. The abundance of template for each amplicon was calculated using its own standard curve. We calculated the ratio between the abundance of the template for the 6–7 primer pair to the abundance of the template for the 8–9 primer pair. We then conservatively corrected this ratio by multiplying it by a factor of 0.7 to account for hepatocyte frequency in the samples32, as only hepatocytes are targeted by rAAV8 in the liver31.

Assessing specificity of F9 expression. cDNA produced from reverse transcription with a poly-dT primer served as a template for two different TaqMan qPCR assays. We quantified the abundance of on-target F9 expression using a qPCR with primers 13: 5′-CTTGGGCTGTGTCCTAC-3′ and 14: 5′-AGATCTTGTTGGCGTTTC-3′. For non-injected controls, no qPCR signal was detected with primers 11–12 and 13–14. The rate of F9-containing Alb mRNAs was calculated as the ratio between the abundance of the template for the 10–11 primer pair to the abundance of template for the 11–12 primer pair.

Assessing vector copy number. DNA purified from mouse liver served as a template for two different TaqMan qPCR assays. We quantified the abundance of hepatoid mouse genomes using a qPCR with primers ‘Alb copy number Ref F’: 5′-CAGGCGGTCTTTTCGGGTAT-3′ and ‘Alb copy number Ref R’: 5′-TCTCATCCTGCACAATCCACAAA-3′. We calculated the abundance of Alb genomes (episomal or integrated) using qPCR with primers 8: 5′-GATACGGTGACTCTCAGTGG-3′ and TaqMan probe: 5′-GGTGCCTGTTCTACCATGACG-3′. We calculated total F9-containing mRNA abundance using a qPCR with primers 15: 5′-GCCGGTCGAAAGCTGGATCAAAGAAA-3′ and 16: 5′-CAAGGGCAGGACACGCTCTC-3′. For non-injected controls, no qPCR signal was detected with primers 14–15 primer pair.

Assessing Northern blot analysis. Western blots for the detection of F9, albumin, P2A, and actin used the following antibodies: polyclonal goat anti-human F9 peroxidase-conjugated IgG secondary antibody at 1:4,200 (Enzyme Research GAFIX-APHRP).

Activated partial thromboplastin time assay. The activated partial thromboplastin time assay was carried out using a SCAB2000 veterinary coagulation analyser (Symbiotics) according to manufacturer’s instructions33.

Western blot analysis. Western blots for the detection of F9, albumin, P2A, and actin used the following antibodies: polyclonal goat anti-human F9 peroxidase-conjugated IgG primary antibody at 1:20,000 (Enzyme Research GAFIX-APHRP), polyclonal rabbit anti-mouse serum albumin IgG primary antibody at 1:40,000 (Abcam ab19196), donkey anti-rabbit peroxidase-conjugated IgG secondary antibody at 1:10,000 (ECL NA-9340), polyclonal rabbit anti-2A peptide primary antibody at 1:10,000 (Millipore ABS31), and monoclonal mouse anti-β-actin peroxidase-conjugated IgG primary antibody at 1:50,000 (Sigma A3854).

Northern blot analysis. Northern blots for the detection of P2A-F9 coding mRNAs used the following probe end-labelled anti-2A probe: 5′-GCCGGTCGAAAGCTGGATCAAAGAAA-3′, and TaqMan probe: 5′-GGTGCCTGTTCTACCATGACG-3′. We calculated the abundance of Alb mRNAs, but not undesired vector amplification, began by performing linear amplification with the following dual-biotinylated primer: 5′-TCAATGCGGCTAGTGTTGACAAATGTGGTGCGGCGG-3′.

Assessing rate of Alb locus targeting by qPCR. Amplification of desired genomic Alb, but not undesired vector amplification, was assessed as the ratio between the abundance of template for the 10–11 primer pair to abundance of template for the 11–12 primer pair.
antibody (Life Technologies A11058) at 1:400 in dPBS for 30 min and then washed three times in dPBS for 2 min. Slides were rinsed with distilled water and mounted with 80 μl ProLong Gold Antifade with DAPI nuclear counterstain (Life Technologies P36935) and covered with a #1.5 coverslip (Thermo 12-544-G). Fluorescent images were taken on a Zeiss Observer.Z1 microscope equipped with a Zeiss AxioCam MRc colour camera and Zeiss AxioVision software (version 4.8.2.0). Images were overlaid using Adobe Photoshop CS6 software (version 13 x64). Controls included no-primary secondary-only antibody staining, and comparison to positive control frozen human liver tissue sections (Zyagen HF-314) and negative control frozen untreated mouse liver sections.

**ALT measurements.** Serum ALT measurements were performed on mouse serum obtained via retro-orbital bleeding using an ALT kinetic measurement kit (Teco Diagnostics) compared with a standard curve. AAV8-U6 and H1 promoter short hairpin RNA (shRNA) sequences are derived from shRNA toxicity studies performed previously.**

**Assessing the distribution of Alb haplotypes in the human population.** A selection of the ShapeIt2 phased haplotypes for the 1000 Genomes Phase 1 integrated variant calls, corresponding to a region 1.3-kb upstream and 1.4-kb downstream from the human P9 integration site at the Alb stop codon, was downloaded from ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase1/analysis_results/shapeit2_phased_haplotypes/ using the 1000 Genomes Data Slicer tool (available at http://browser.1000genomes.org/Homo_sapiens/UserData/SelectSlice). Haplotypes consisting of single nucleotide polymorphisms with a substantial frequency of the alternative allele in all populations were combined (here, greater than or equal to 45%, whereas those excluded were less than 1%) and treated as individual strings for calculating population frequency.

31. Grimm, D., Pandey, K., Nakai, H., Storm, T. A. & Kay, M. A. Liver transduction with recombinant adeno-associated virus is primarily restricted by capsid serotype not vector genotype. *J. Virol.* **80,** 426–439 (2006).
32. Lu, J., Zhang, F. & Kay, M. A. A mini-intronic plasmid (MIP): a novel robust transgene expression vector in vivo and in vitro. *Mol. Ther.* **21,** 954–963 (2013).
33. Mitchell, C. & Willenbring, H. A reproducible and well-tolerated method for 2/3 partial hepatectomy in mice. *Nature Protocols* **3,** 1167–1170 (2008).
34. Rogers, G. L. & Hoffman, B. E. Optimal immunofluorescent staining for human factor ix and infiltrating T cells following gene therapy for hemophilia B. *J. Genet. Syndr. Gene Ther. Suppl.* **1,** 012 (2012).
35. Grimm, D. et al. Fatality in mice due to oversaturation of cellular microRNA/short hairpin RNA pathways. *Nature* **441,** 537–541 (2006).
Extended Data Figure 1 | Human F9 liver immunohistochemistry. From top to bottom, panels show human F9 staining (red) with 4′,6-diamidino-2-phenylindole (DAPI) nuclear counterstain (blue) in positive control human liver, negative control untreated mouse liver, and two sets of representative stains from mice treated as neonates or adults with AAV8-F9. Original magnification, ×200.
Extended Data Figure 2 | Scheme of targeting rate assessment. Assessment of on-target integration rate begins using linear amplification with biotinylated primer 1 (black), annealing to the genomic locus but not to the vector (step 1). Linear amplicons are then bound to streptavidinylated beads and washed to exclude episomal vectors (step 2). Subsequent second-strand DNA synthesis with random primers (step 3) was followed by CviQI restriction digestion (step 4). A compatible linker is then ligated (step 5) followed by two rounds of nested PCR amplifications (primers 2–3 in blue (step 6), and then primers 4–5 in red (step 7)). CviQI cleaves at the same distance from the homology border in both targeted and wild-type alleles, thus allowing for unbiased amplification. The amplicons of the second nested PCR then serve as a template for qPCR assays with either primers 6–7 (green) or 8–9 (orange) (step 8).
Extended Data Figure 3 | Standard curves for targeting rate assessment by qPCR. qPCR standard curves for the targeted allele (primers 8 and 9, Fig. 3) and non-targeted allele (primers 6 and 7, Fig. 3). Mass units used are functionally equivalent to molarity because all amplicons used were of equal length.
Extended Data Figure 4 | Toxicity assessment by ALT measurement.
Alanine transaminase levels (ALT) were evaluated 7 days after injection in mice injected with AAV8 coding for our experimental vector ($1 \times 10^{12}$) or a negative control coding for a known non-toxic cassette ($1 \times 10^{12}$ of H1 promoter-driven shRNA), or a positive control coding for a known toxic cassette ($5 \times 10^{11}$ of U6 promoter-driven shRNA). Data represent mean of two measurements of four independent mice for each group. The statistical significance is defined here as having $P < 0.05$ in a one-tailed $t$-test between samples of different variance.
Extended Data Figure 5 | Vector copy number. Vector copy number assessed by qPCR using primers 8 and 9 (Fig. 3). $n = 7$ for mice injected as adults; $n = 6$ for mice injected as neonates and analysed before or after partial hepatectomy (PH). Error bars represent s.d.
Extended Data Table 1 | Haplotypes in the human population at the relevant *ALB* locus as extracted from the 1000 Genomes Project

| Position | ID    | REF | ALT | HAP1 | HAP2 | HAP3 | HAP4 | HAP5 | HAP6 | HAP7 |
|----------|-------|-----|-----|------|------|------|------|------|------|------|
| 74285239 | rs962004 | C   | T   | T    | C    | C    | C    | C    | T    | C    |
| 74285552 | rs4076  | A   | G   | A    | G    | A    | A    | A    | G    | G    |
| 74285567 | rs962005 | C   | A   | C    | A    | C    | C    | C    | A    | A    |
| 74285758 | rs2236766 | G   | T   | G    | T    | T    | G    | G    | T    | T    |
| 74285823 | rs2236767 | G   | A   | G    | A    | G    | G    | G    | A    | A    |
| 74287403 | rs4429703 | T   | C   | C    | T    | T    | C    | T    | T    | C    |
| **Frequency** | | | | | | **50.14%** | **44.69%** | **2.15%** | **1.51%** | **1.37%** | **0.09%** | **0.05%** |