Ribosomal DNA Integrating rAAV-rDNA Vectors Allow for Stable Transgene Expression

Leszek Lisowski[1], Ashley Lau[1,2], Zhongya Wang[3], Yue Zhang[1], Feijie Zhang[1], Markus Grompe[3] and Mark A Kay[1]

[1]Stanford University, Departments of Pediatrics and Genetics, Stanford, California, USA; [2]Harvard Medical School, Boston, Massachusetts, USA; [3]Oregon Stem Cell Center, Oregon Health and Science University, Portland, Oregon, USA

Although recombinant adeno-associated virus (rAAV) vectors are proving to be efficacious in clinical trials, the episomal character of the delivered transgene restricts their effectiveness to use in quiescent tissues, and may not provide lifelong expression. In contrast, integrating vectors enhance the risk of insertional mutagenesis. In an attempt to overcome both of these limitations, we created new rAAV-rDNA vectors, with an expression cassette flanked by ribosomal DNA (rDNA) sequences capable of homologous recombination into genomic rDNA. We show that after *in vivo* delivery the rAAV-rDNA vectors integrated into the genomic rDNA locus 8–13 times more frequently than control vectors, providing an estimate that 23–39% of the integrations were specific to the rDNA locus. Moreover, a rAAV-rDNA vector containing a human factor IX (hFIX) expression cassette resulted in sustained therapeutic levels of serum hFIX even after repeated manipulations to induce liver regeneration. Because of the relative safety of integration in the rDNA locus, these vectors expand the usage of rAAV for therapeutics requiring long-term gene transfer into dividing cells.

Received 20 April 2012; accepted 20 July 2012; advance online publication 18 September 2012. doi:10.1038/mt.2012.164

INTRODUCTION

The permanency of gene transfer for the treatment of human disease is dependent on the type of vector and target tissue. Episomal vectors by definition do not integrate into the host genome but can be effective at transgene expression in quiescent tissues. On the other hand, replication defective vectors based on onco/retro/lentiviral vector can stably integrate into the genomic DNA,[1] a property allowing for persistent transgene expression from dividing cells but also increasing the risk of insertional mutagenesis.[2–4]

In recent years, adeno-associated virus (AAV)-derived vectors gained much popularity for use in gene therapy approaches as early clinical success has been achieved in specific types of blindness[5] and hemophilia B.[6] The large number of natural AAV variants,[7–16] as well as different strategies for the generation of new variants (reviewed in refs. 17,18), provide diverse tissue tropism.

Our laboratory has shown that almost all episomal rAAV genomes are lost after just one hepatocyte cell division, while a very small percentage (<<1%) of the stable rAAV genomes are integrated into different portions of the genome.[19–21] Although the rAAV integration is a rare event, there is still some potential risk of insertional oncogenesis, as suggested by results from neonatal mice given high doses of rAAV.[22]

The nonintegrating aspect of rAAV vectors is a double-edged sword. rAAV vectors are very useful for gene therapy delivery into quiescent cells but when delivery is intended into dividing cells, the vector genome, and thus transgene expression, is quickly lost making the effect transient. Multiple approaches taken in order to integrate rAAV into site-specific genomic loci or in a restricted manner have not proven to be very successful *in vivo* and often required transient coexpression of other protein factors, such as AAV Rep protein which can be toxic.[23–27]

An AAV vector capable of safe integration into a specific or limited number of genomic loci without a risk of causing insertional mutagenesis would ensure long-term transgene expression and allow to move the well-established resources of AAV vectors into use in dividing cells both *in vitro* and *in vivo*/*ex vivo*.

To this end, we wanted to create a rAAV vector capable of safe and site-specific integration into the host genome. As a target sequence we selected ribosomal DNA (rDNA) locus for two reasons: multiple large-scale analyses have shown that rAAV integrations occur at the rDNA locus at a higher frequency than random,[20,21,28] possibly because rAAV is processed at the nucleolus[29] in proximity to rDNA loci. In addition because rDNA exists in numerous tandem repeats at multiple chromosomal locations throughout the genome, with ~400 copies in human genome,[30,31] we hypothesized that targeted integration into this locus, which will disrupt some of the *rDNA* genes, should not have any significant effect on cell viability. Furthermore, studies in other species have demonstrated that cells can very well tolerate disruptions in the *rDNA* genes. In *Drosophila melanogaster* nonlong–terminal repeats retrotransposons disrupt ~50% of the rDNA transcription units,[32] while studies in *Saccharomyces cerevisiae* show that the number of *rDNA* genes fluctuates through natural mechanisms of rDNA deletions and additions,[33] without any effect on cell viability. Experiments in tyrosinemia type 1 (accompanying paper, see Wang *et al.*[34]), a liver disease model, which selects for integrated rAAV genomes carrying human fumarylacetoacetate hydrolase (FAH), indicated that rDNA AAV vectors indeed have higher integration frequencies. In order to determine whether rAAV-rDNA vector has general utility even without transgene selection, we developed

Correspondence: *Mark A Kay, Stanford University, Department of Pediatrics and Genetics 269 Campus Drive, CCSR 2105, Stanford, California 94305-5164, USA. E-mail: markay@stanford.edu*

human factor IX (hFIX) rDNA-AAV vectors. Here, we demonstrate that rDNA might be an ideal safe locus for transgene integration.

RESULTS

Animals treated with rAAV-rDNA vectors are more resistant to the loss of transgene expression

Experiments performed in Fah knockout mice indicated increased transgene persistence with rDNA-containing vectors when compared to historical controls (accompanying paper, see Wang *et al.*[34]). In order to determine whether rDNA vectors were truly superior for integration into the 28S rDNA locus (**Figure 1a**), we compared two rAAV vectors: rAAV-rDNA-hFIX and a size-matched rAAV-stuffer-hFIX control (**Figure 1b**). The sequences chosen as rDNA flanks were the immediate upstream and downstream sequences around I-PpoI endonuclease recognition site. In this first experiment, we intravenously injected six C57BL/6

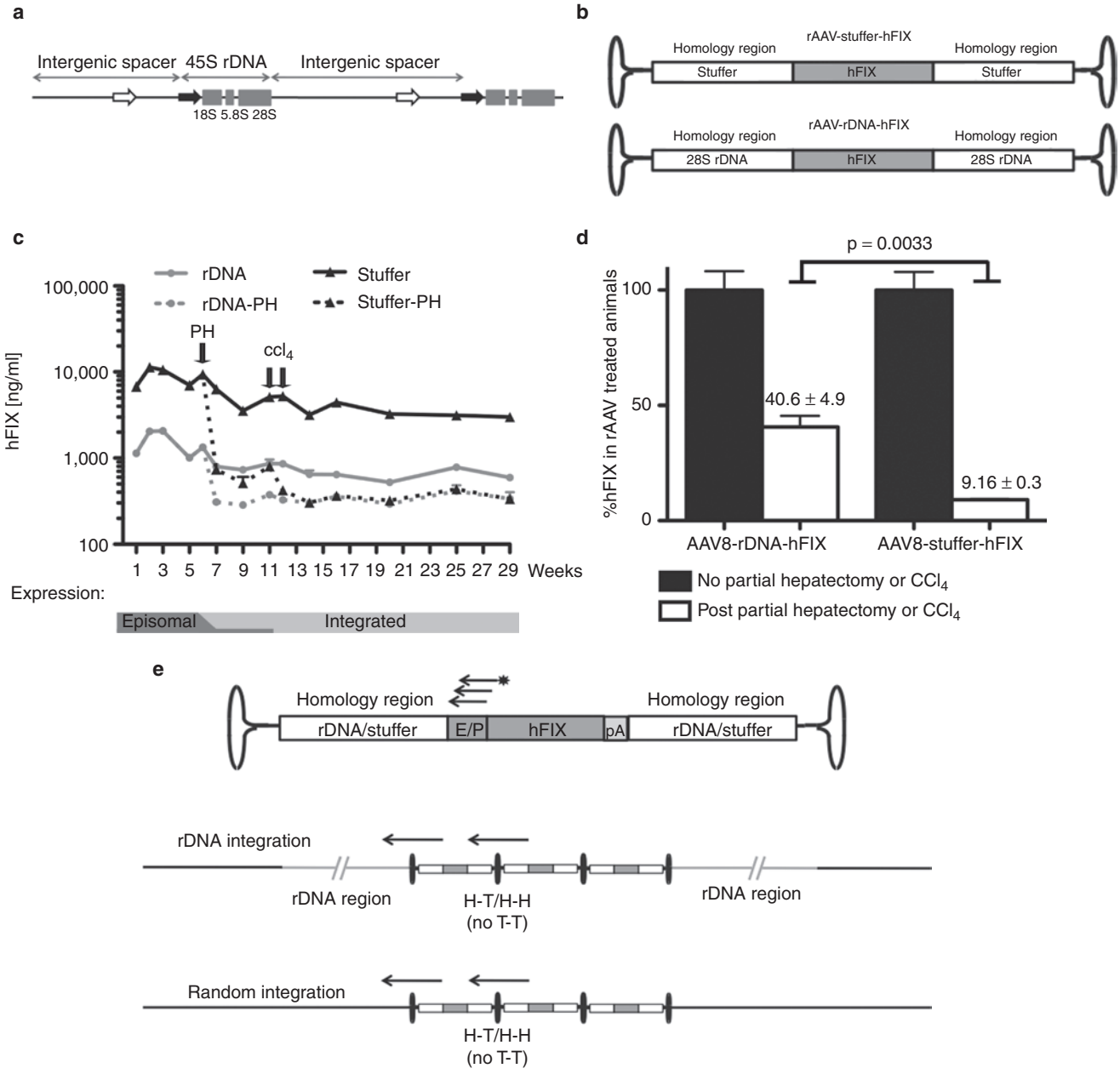


Figure 1 Experimental design and preliminary data. (**a**) Schematic representation of the organization of the *rRNA* genes in eukaryotes. (**b**) rAAV-rDNA-hFIX and rAAV-stuffer-hFIX vector design. The white boxes represent the 28S ribosomal DNA (rDNA) and the corresponding "stuffer" sequences. The gray boxes represent hFIX expression cassette (TTR promoter, hFIX cDNA+1st intron, and bovine growth hormone pA). (**c**) hFIX enzyme-linked immunosorbent assay (ELISA) data throughout the course of the study. Solid lines—data for animals that did not undergo liver regeneration. Dotted lines—data for animals that underwent partial hepatectomy and CCl_4 treatment. Time points of partial hepatectomy and CCl_4 treatments are indicated. The expected form of rAAV that expresses transgene at different times during the study is indicated in the lower panel. Error bars represent SD $n = 3$. (**d**) Quantification of hFIX levels at the end of the study (week 29). Error bars represent SD $n = 3$. (**e**) Linear amplification-mediated-PCR (LAM–PCR) design. The biotinylated primer used for primer extension step is shown as arrow with a star. Long arrows show two possible amplicons—internal noninformative amplicon and external informative amplicon-containing genomic DNA.

female mice with 5 × 10 vg/mouse of the corresponding vectors packaged into AAV8 capsids. Five weeks after vector administration, three animals from each group underwent a surgical 2/3 partial hepatectomy (PH), a procedure previously shown to cause loss of over 95% of episomal rAAV proviral genomes.[19] To further ensure that all the episomal rAAV sequences were lost, 5 weeks later, animals received two sublethal injections of CCl_4. Serum hFIX levels in animals administered the rAAV-rDNA-hFIX and rAAV-stuffer-hFIX vectors (with and without PH/CCl_4) are shown in **Figure 1c**. After liver regeneration, the serum hFIX levels in rAAV-rDNA-hFIX-treated mice were 23% of what was observed in the first few weeks post-treatment (~1,300 ng hFIX/ml before PH and ~300 ng hFIX/ml after PH and CCl_4) or 41% of non–liver-regeneration controls followed for the same period of time at week 29 (**Figure 1c** and **d**). In contrast, rAAV-stuffer-hFIX-treated mice had only 3.8% of their original serum hFIX after liver regeneration or 9% of non–liver regeneration controls followed for the same period of time (**Figure 1c** and **d**). This data was originally interpreted to suggest that the rAAV-rDNA-hFIX-injected mice had substantially enhanced vector integration compared to animals that received the control, non-rDNA containing vector. However, we noticed lower levels of pre-liver regeneration transgene expression from rAAV-rDNA versus rAAV-stuffer vector-injected animals. Because the average vector copy number (VCN) and provector structures were similar between the rDNA and stuffer groups (**Supplementary Figure S1**) the discordant transgene expression from the two vectors most probably resulted from differences in vector expression and not gene transfer efficiencies (see also below and Discussion section). Of note, the rAAV-rDNA-hFIX mice maintained a therapeutic level of serum hFIX even after the induction of robust liver regeneration.

Detailed Southern blot analysis of provector integration profiles

To study the provector integration profile, we performed Southern blot analyses on genomic DNA from the livers of animals that had received rAAV-rDNA-hFIX or rAAV-control-hFIX with or without induction of liver regeneration (**Supplementary Figure S1**). As we have previously shown, before PH/CCl_4 the majority of rAAV proviral genomes persist as monomeric or concatemeric episomes, whereas after liver regeneration only the integrated provector remains.[19] Using a restriction enzyme that cuts twice in the vector sequence, we determined and found as expected[19] that the average total VCN was significantly higher in animals that did not undergo liver regeneration compared to those that did (**Supplementary Figure S1a**). The presence of specific bands corresponding to head-to-head (H-H) and head-to-tail (H-T) orientations of vectors within concatemers when using a single cutter enzyme (**Supplementary Figure S1b**) indicated that a proportion of both episomal and integrated vectors existed as concatemers. Based on homology arms present in the rAAV-rDNA vectors, an *Spe*I digest would produce a 8.3-kb band if a significant fraction of rDNA vectors integrated into 28S rDNA *via* homologous recombination (**Supplementary Figure S1d**). The lack of additional bands in samples from animals after liver regeneration indicated, however, that majority of provectors did not integrate through homologues recombination but integrated either as large concatemers in a limited number of genomic loci, or that the provectors existed as small concatemers at many different loci. The monomeric and concatemeric forms of the episomal vector in non–liver-regeneration animals were also detected using a zero-cutter restriction enzyme (**Supplementary Figure S1c**). The lack of detectable bands in the PH/CCl_4 animals treated with rAAV-rDNA or rAAV-stuffer vectors after digestion with a zero-cutter, indicated that, as predicted,[19] the rAAV genomes remaining after liver regeneration were likely integrated into genomic DNA. The vector copy controls established the sensitivity of our Southern blot at <0.5 provector copies per diploid genome. The absence of any detectable bands using zero-cutter suggested that (i) most of the integration did not occur at same specific site and/or (ii) the provectors integrated as various sized concatemers, either at a multiple loci or at the same loci in different cells, but due to significant size they were not detected by Southern blot.

rAAV-rDNA vectors integrate into rDNA loci more frequently than control vectors

In order to investigate whether the rDNA vectors not only integrated more frequently into genomic DNA, but more importantly, that integration was locus specific, we performed extensive linear amplification-mediated-PCR (LAM-PCR) analyses. The design of LAM-PCR is shown in **Figure 1e**. Multiple biotinylated oligos, endonucleases, and polymerases were tested and compared side-by-side in order to derive the optimal combination (see **Supplementary Data**). It is important to mention, that due to the fact that the exact mechanism of integration was difficult to predict, it was not possible to determine which vector sequences would be present in the provector (see **Supplementary Data**). The size of the PCR amplicon required in order to obtain genomic location information was thus unknown, making it impossible to use any of currently available deep-sequencing techniques.

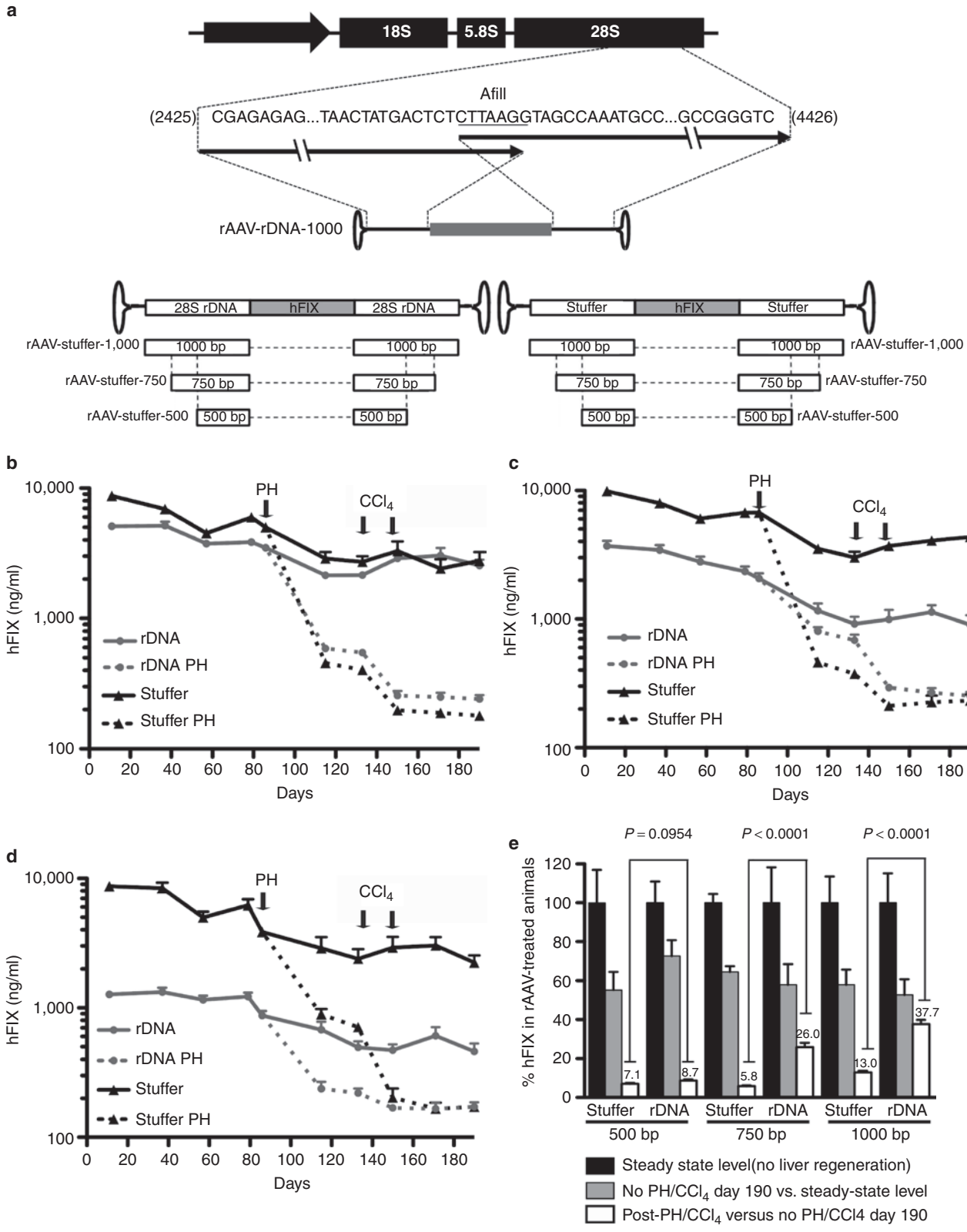
LAM-PCR analysis of liver DNA samples from three separate animals in each group demonstrated that out of 38 analyzed informative clones in the rAAV-rDNA-hFIX group, 32 (84%) were integrated in the rDNA genomic loci, whereas the remaining 6 (16%) were integrated into non-rDNA loci or could not be analyzed as the specific sequence returned no matches to the mouse genome. In the rAAV-stuffer-hFIX group, out of 44 informative clones 1 (2%) was found in the rDNA locus, 26 (60%) were found to be in non-rDNA loci, while the remaining 17 (38%) were integrated in a genomic region that could not be annotated (data not shown).

Despite the fact that this was a relatively small pilot study with a small number of animals in each treatment group, these preliminary data prompted us to initiate a more detailed study to compare the effect of rDNA flanking sequences on the frequency of provector integration into the rDNA loci of the genome.

The packaging of rAAV-rDNA vectors is not affected by rDNA flanking sequences ranging between 500 and 1,000 bp

Due to the small packaging capacity of the rAAV, we wanted to determine the shortest rDNA sequences needed for efficient and specific transgene integration into rDNA loci. To do this, we cloned a new set of vectors (**Figure 2a**), in which the hFIX cassette (identical to the one used in the first study, see above) was

flanked with 500, 750, or 1,000 bp of 28S rDNA. Control mammalian stuffer DNA replaced the rDNA sequences in control vectors (**Figure 2a**). As the size of the flanking sequences increased, so did the overall size of the vector, making them 3.6, 4.1, and 4.6 kb in length, respectively. To avoid complicating the direct comparison between rAAV-rDNA and control rAAV-stuffer vectors, we decided not to adjust the overall vector size of the shorter vectors by introducing additional stuffer sequences. Despite the significant differences in vector sizes and presence of up to 2 kb of rDNA sequences per vector, as shown in **Supplementary Figure S2** all



six vectors were packaged into AAV8 capsid at very similar efficiencies resulting in rAAV preparations with dot blot titers ranging between 1.3 and 2.6 × 10^{13} vg/ml.

rDNA flanks of 750 bp or more are required for efficient targeted integration

In order to compare the new constructs *in vivo*, we followed a similar experimental design as in the previously described pilot study (see above). We injected 5 × 10^{10} vg/animal of each construct into fifteen C57BL6 age-matched female mice and monitored the serum hFIX levels at various time points throughout the experiment (**Figure 2b–d**). On day 86, 10 of the animals from each treatment group underwent a 2/3 PH, followed by CCl_4 treatments on days 133 and 150. The detailed comparison of serum hFIX levels at the last time point of the experiment (day 190) is shown in **Figure 2e**. The relative difference in the serum hFIX levels in animals with or without liver regeneration was similar between the 500 bp rAAV-rDNA vector (4.8% before PH/CCl_4, or 8.7% of non–liver-regeneration controls) and the rAAV stuffer controls (5.1% before liver regeneration, or 7.1% of non–liver-regeneration controls) (**Figure 2b** and **e**). In contrast, the difference between the rAAV-rDNA and stuffer control vectors was more substantial in the 750 and 1,000 bp groups. In the 750 bp rAAV-rDNA group after PH/CCl_4 treatment, serum hFIX levels were 11.2% of the levels observed before the induction of liver regeneration or 26.0% of the levels observed in control animals that received the same vectors but were not subjected to liver regeneration, while relative hFIX expression in the 750 bp AAV-stuffer control group were 3.8 and 5.8%, respectively. Thus the relative level of hFIX in rAAV-rDNA-treated animals were about four times higher than in corresponding stuffer control animals. In the 1,000 bp group, the difference in average level of hFIX after liver regeneration between rAAV-rDNA and rAAV-stuffer treatment group was about threefold (38% in rDNA versus 13% in the control group).

rDNA sequences affects transgene expression from episomal rAAV vectors

As noted in the pilot study (see above), during the course of this study it was also noted that despite the fact that all animals were injected with the same total number of vector particles, the expression of serum hFIX differed between rAAV-rDNA and rAAV-stuffer treated animals in each group (**Figure 2b–d**). The repeated dot blot titration of vector preparations used in the study confirmed that the original titer data were correct and that all animals received the intended number of viral particles (data not shown). However, despite the identical vector dose given in different groups, the longer the rDNA insert present in the vector, the greater the difference between AAV-rDNA and AAV-stuffer mediated transgene expression. The biggest difference was observed in the 1,000 bp group at the earliest time point with the rDNA group expressing at ~15% of the "stuffer" control group (1,300 versus 8,700 ng/ml).

Southern blot and TaqMan analysis of liver samples taken during PH (**Figure 3a** and **Supplementary Figure S3**) as well as TaqMan analysis of liver samples taken at the end of the study from animals that underwent liver regeneration (**Supplementary Figure S3**) confirmed that the average VCN in all groups were similar.

Based on the vector copies in pre-PH and post-PH samples, and the fact that both copy numbers and hFIX levels after PH/CCl_4 were similar between rAAV-rDNA and rAAV-stuffer-treated animals, we conclude that the inclusion of rDNA sequences did not affect vector DNA transfer efficiency but rather reduced expression from the episomal vectors only. Thus, although the net decrease in VCNs after liver regeneration was similar between corresponding rAAV-rDNA and rAAV-stuffer treated groups, the net decrease in hFIX expression associated with the loss of episomal AAV forms was significantly larger for rAAV-stuffer-treated groups. Thus, determining the relative level of transgene expression, before and after liver regeneration and using these calculations to estimate the relative efficiency of integration between the different vector groups can be misleading. Despite the fact that the rDNA vectors expressed at a lower lever than control vectors prior to liver regeneration, the transgene expression from 500 and 750 bp vectors was still higher when compared with animals treated with the control vectors after liver regeneration was complete. In addition, after liver regeneration was complete, therapeutic serum hFIX levels were maintained from integrated provector genomes.

The presence of rDNA flanking sequences had no effect on rAAV-rDNA proviral genome structure

In order to establish whether the inclusion of rDNA sequences had an effect on rAAV proviral DNA structure, which perhaps could explain the observed differences in transgene epression, Southern blot analysis was performed (**Figure 3b** and **c**). Analysis of samples from animals that did not undergo liver regeneration revealed all expected episomal proviral forms (**Figure 3b**). Single-cutter revealed bands corresponding to H-H and H-T orientations of monomeric vectors in concatemers, whereas zero-cutter revealed bands corresponding to supercoiled circular monomers, monomers and larger concatemers, in all groups treated with rDNA and control vectors. This suggests that the rDNA did not have a major effect on proviral DNA structures and cannot explain the inhibitory effects on transgene expression. After liver regeneration (**Figure 3c**), the single-cutter similarly produced bands corresponding to H-H and H-T orientation originating from concatemers, whereas the zero-cutter gave rise to high molecular weight bands in all treatment groups. Similarly to results discussed earlier (**Supplementary Figure S1**), due to low average VCN, the

Figure 2 ***In vivo*** **comparison of human factor IX (hFIX) levels in animals treated with 500, 750, or 1,000 bp ribosomal DNA (rDNA) or stuffer vectors.** (**a**) Upper panel: Design of rAAV-rDNA vectors in respect to genomic rDNA sequence; Lower panel: Graphic representation of the six vectors used in the study. The gray box represents hFIX expression cassette (TTR promoter, hFIX cDNA+1st intron, and bovine growth hormone pA). The size of each flank is indicated. (**b–d**) hFIX levels throughout the study in animals treated with rAAV-rDNA/stuffer-hFIX (**b**) 500 bp, (**c**) 750 bp, or (**d**) 1,000 bp. Data for animals treated with rAAV-rDNA vectors is shown in gray, and data for animals treated with rAAV-stuffer is shown in black. Dotted lines represent data for animals that underwent partial hepatectomy (PH) and CCl_4 treatments. Error bars represent SD n = 5–10. (**e**) Quantification of data at the last time point shown in (**b–d**). hFIX levels at the last time point before PH were used for normalization and thus were assigned a value of 100%. Shaded bars represent hFIX levels on day 190 in animals that did not undergo PH/CCl_4. White bars represent hFIX levels on day 190 in animals post PH/CCl_4 normalized to animals followed for the same period of time that did not undergo liver regeneration. Error bars represent SD, n = 5–10.

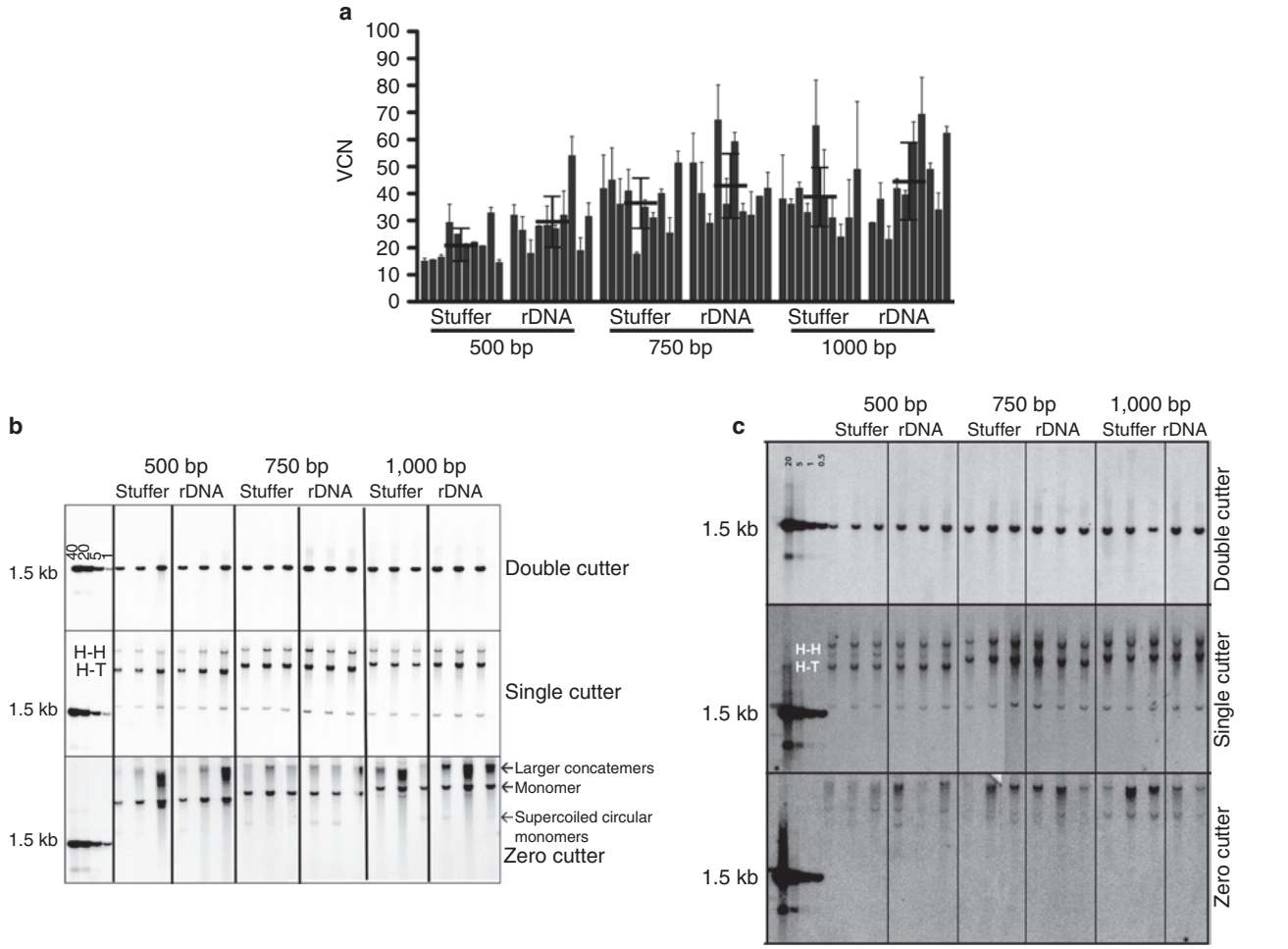


Figure 3 Vector copy number and vector genome structure analysis. (**a**) Vector copy number (VCN) obtained from Southern blot analysis of liver samples harvested during partial hepatectomy (PH) on day 86. Black bars indicate the average value for the given group ± SD $n = 3–5$. (**b**,**c**) Southern blot analysis of vector genome structure in liver samples obtained at the end of the study from (**b**) non–liver-regeneration control or (**c**) post-PH/CCl_4 animals injected with rAAV-rDNA or rAAV-stuffer vectors. Genomic DNA was digested with zero-, single-, or double-cutter endonucleases. Digested plasmid DNA mixed with digested control genomic DNA was used as standards (1,533 bp). Plasmid DNA was diluted to equal 40, 20, 5, or 1 vector copy per diploid genome in (**b**) or 20, 5, 1, 05 vector copies per diploid genome in (**c**). Bands corresponding to head-to-head (H-H) and head-to-tail (H-T) orientations of the rAAV genomes in the concatemers are indicated. Other observed rAAV genome structures are indicated.

Southern blot analysis of post-PH/CCl_4 samples proved to be much more challenging than expected, making it difficult to determine whether there were any differences in provector integration profile between rDNA and control vectors.

The tropism of rAAV vectors is not affected by the presence of rDNA sequences

Although the tropism of the vector is determined by the AAV capsid, and all the vectors used in this study were packaged into AAV8 capsid, we wanted to ensure that the changes in vectors size between different constructs and/or the presence of large rDNA flanks did not affect the biodistribution/processing of the vectors. For this reason we performed TaqMan analysis to determine the average VCN in different tissues extracted at the end of the study (**Supplementary Figure S3**). As expected, there were no detectable differences in vector tropism between the different groups, with the majority of vector being detected in the liver, and small amounts present in kidney and heart.

rAAV-rDNA vectors integrate into rDNA sequences 8–13 times more frequently than corresponding control vectors

The Southern blot studies, while informative, did not provide any specific information about the sites of vector integration, in part because of the low sensitivity of the analyses on post-liver regeneration samples. Thus in order to analyze whether the presence of rDNA flanks of different sizes affected the provector integration profile, as in the pilot study described previously, we performed provector integration analysis by LAM-PCR (**Figure 1e**). In this set of experiments, the size of the minimal amplicon required to include genomic DNA flanking the integration site was different between different vector groups. The 500 bp group would require the shortest PCR amplicon in order to obtain DNA sequence outside of the integrated provector, and the 1,000 bp group would require the longest PCR product in order to be informative. This increased the probability of non-informative PCR clones, caused by premature PCR termination on the GC rich rDNA sequences, obtained from the 750 and 1,000 bp

rDNA vectors. To overcome these limitations, four different DNA polymerases were tested for their efficiency in LAM-PCR (data not shown, see **Supplementary Data**). Although we cannot exclude the possibility that the selected polymerase introduced bias in our LAM-PCR analysis, we believe that should such bias occur, it would apply mostly to the rDNA vector clones and thus would lead to an underestimation of the rAAV-rDNA integration frequency (see also **Supplementary Data**). Furthermore, as shown in **Figure 1e**, due to the fact that AAV often integrates as concatemers (H-H and H-T) it was predicted that fraction of sequenced LAM-PCR clones would contain noninformative DNA sequences derived from amplification of the internal vector sequences.

The summary of LAM-PCR data is shown in **Table 1**. Analysis of LAM-PCR sequences provided additional information on the mechanism of provector integration (**Supplementary Figure S4**) and genomic location of most of the informative clones (**Supplementary Figure S5a–b** and **Supplementary Table S1**). In the 500 bp stuffer group, out of 44 informative clones, only 3 (7%) were found to be integrated into rDNA loci while 41 (93%) integrated into non-rDNA regions of the genome. One of these clones (2.4%) was integrated *via* homologous recombination into *Fah* gene. In contrast, from 54 informative rAAV-rDNA-500 clones, 19 (35%) were integrated into loci different than rDNA, whereas 35 (65%) were found to be located within the rDNA region. Of the 35 clones, 4 (11%) integrated *via* homologous recombination into 28S rDNA and 31 (89%) integrated through other mechanisms. Out of 45 informative clones isolated from animals treated with 750 bp stuffer vector, 5 (11%) were integrated into rDNA and 40 (89%) into non-rDNA loci. Again, 1 clone (2.5%) was found to be integrated into *Fah* gene. In the rDNA-750 bp group, the majority of informative clones (57 out of 68, 84%) were integrated into rDNA, with 8 (14%) of those occurring *via* homologous recombination. None of the 34 informative 1,000 bp stuffer clones were found in the Fah gene and only 2 (6%) clones were in rDNA loci. Similarly to what was observed in 500 and 750 bp rDNA groups, in the 1,000 bp rDNA group, the majority of informative clones (42 out of 54, 78%) were found to be integrated into rDNA, with 7 (17%) of those occurring *via* homologous recombination.

There were no significant differences in integration profile within rDNA region between rAAV-rDNA vectors that integrated into rDNA loci (**Table 1** and **Supplementary Figure S5c**). From the 115 nonhomologous rDNA integrations, 7 (6%) were actually integrated at the site expected from homologous recombination, but because they contained deletions and/or rearrangements at the junction site they were counted as nonhomologous events. From the remaining 108 clones, 35 clones (30%) were found in

Table 1 LAM-PCR analysis

						Informative amplicons					
						rDNA			non-rDNA		
	Animal	Number of amplicons	Totalno of amplicons	Noninformative amplicons (H-H/H-T/ET) (% total)	Informative amplicons (% total)	Total (% informative amplicons)	Homologous (% rDNA total)	Nonhomologous (% rDNA total)	Total (% informative amplicons)	Homologous (% non-rDNA total)	Nonhomologous (% non-rDNA total)
	1	32									
500 stuffer	2	27									
	3	36	95	51 (54%)	44 (46%)	3 (7%)	NA	3 (100%)	41 (93%)	1 (2.4%)	40 (97.6%)
	1	35									
500 rDNA	2	31									
	3	28	94	40 (43%)	54 (57%)	35 (65%)	4 (11%)	31 (89%)	19 (35%)	NA	55 (100%)
	1	28									
750 stuffer	2	33									
	3	29	90	45 (50%)	45 (50%)	5 (11%)	NA	5 (100%)	40 (89%)	1 (2.5%)	39 (97.5%)
	1	38									
750 rDNA	2	29									
	3	25	92	24 (26%)	68 (74%)	57 (84%)	8 (14%)	49 (86%)	11 (16%)	NA	11 (100%)
	1	29									
1,000 stuffer	2	26									
	3	28	83	49 (59%)	34 (41%)	2 (6%)	NA	2 (100%)	32 (94%)	0 (0%)	32 (100%)
	1	31									
1,000 rDNA	2	32									
	3	33	96	42 (44%)	54 (56%)	42 (78%)	7 (17%)	35 (83%)	12 (22%)	NA	12 (100%)

Abbreviations: LAM-PCR, linear amplification-mediated-PCR; NA, not applicable; rDNA, ribosomal DNA; PH, partial hepatectomy.
Tabular representation of LAM-PCR data obtained from post-PH liver samples harvested at the end of the study. Informative sequences indicate all sequences for which genomic localization was obtained. Noninformative sequences include internal-amplicons (see **Figure 1e**), early terminations (ET) resulting from polymerase terminating within the vector encoded rDNA/stuffer flanks and thus not containing any endogenous genomic sequences, as well as sequences that could not be annotated. Numbers indicate absolute number of clones in each groups, while percentages are given in parenthesis. Where applicable, informative clones were divided into homologous and nonhomologous integration events. Head-to-head (H-H) and head-to-tail (H-T) refers to orientations of AAV genomes in higher concatemers.

28S region, 23 (20%) in 18S and 3 (3%) in 5.8S. In addition, 5 (4%) clones integrated into 5′ETS, 9 (8%) into ITS1 and ITS2, 27 (24%) into IGS and 6 (5%) clones were found in 3′ETS.

Pair-wise comparison of rAAV-rDNA and corresponding stuffer controls showed that in the 500 bp group the frequency of rDNA specific integration increased nine times between rAAV-stuffer and rAAV-rDNA (7% rDNA integration in the control group versus 65% in the rDNA group) when only the informative sequences were taken into account. In the 750 bp group, the rDNA flanking sequences increased the specificity of rDNA integration among informative clones by 7.6-fold (11% for control versus 84% for rDNA vector) and in the 1,000 bp group, the difference was the largest—13 fold (6% versus 78%, stuffer versus rDNA). Using a plasmid rescue approach, we have previously established that ~3% of rAAV vector genomes integrations fall into rDNA loci.[21] Based on that, we can roughly estimate that 23–39% of rDNA vector integrations were targeted into the rDNA locus.

rAAV-rDNA vectors perform better than traditional rAAV vectors in constitutively dividing tissue culture cells

We next evaluated the AAV-rDNA vectors in rapidly dividing cells in culture. To do so, we modified the rAAV-rDNA-1,000 bp constructs and replaced the TTR-hFIX-pA cassette with hPGK-eGFP-pA cassette, which allowed for quick and efficient transgene expression analysis by fluorescence-activated cell sorting (FACS). As the new transgene cassette was smaller than the hFIX cassette, we introduced 400 bp of hFIX cDNA as a DNA spacer between pA and 3′rDNA/stuffer to keep the vector size the same as in the *in vivo* studies (**Figure 4a**). This "spacer" also allowed us to use the same Southern blot and TaqMan probes used in the *in vivo* study. As a control, we cloned an identical vector harboring 1,000 bp "stuffer" sequences in place of rDNA flanking sequences (**Figure 4a**).

Both vectors were packaged into AAVDJ[35] capsid because of its robust *in vitro* transduction properties. Based on data from the *in vivo* studies, we predicted that even when transduced at the same multiplicity of infection (MOI), there might be a difference in initial levels of green fluorescent protein (GFP) expression between rDNA and stuffer control vectors (**Figure 2b–d**). Thus cells were transduced at two densities (1.2×10^6 and 1.5×10^6/well) and at various MOIs in order to generate groups that can be compared to one another without bias. To avoid complex issues resulting from saturation, all groups that expressed GFP at levels higher than 90% were excluded from the study. **Figure 4b** shows number of transduced cell and MOI used as well as GFP levels on day 2 for 3 selected groups transduced with rAAV-stuffer/

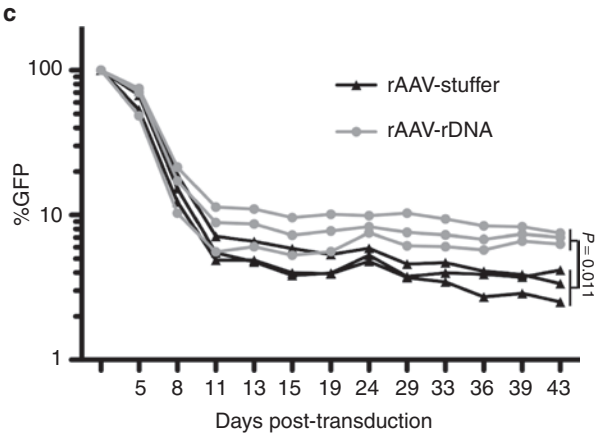
a

b

	AAV-stuffer		
Group	Cell number	MOI	% GFP day 2
1	1.20E+06	8000	65
2	1.50E+06	9000	77
3	1.50E+06	5000	56

	AAV-rDNA		
Group	Cell number	MOI	% GFP day 2
1	1.20E+06	9000	61
2	1.50E+06	17000	81
3	1.50E+06	6000	46

c



d

Group	AAV-stuffer (%GFP)	AAV-rDNA (%GFP)	Flod difference
1	3.36	7.49	2.2
2	2.52	6.32	2.5
3	4.16	7.03	1.7

e

	Group	Number of amplicons	Total number of amplicons	Non-informative amplicons (H-H/H-T/ET) (% total)	Informative ampilcons (% total)	rDNA			Non-rDNA		
						Total (% informative amplicons)	Homologous (% rDNA total)	Non-homologous (% rDNA total)	Total (% informative amplicons)	Homologous (% rDNA total)	Non-homologous (% rDNA total)
1,000 stuffer	1	43	138	77 (56%)	61 (44%)	7 (11%)	NA	7 (100%)	54 (89%)	1 (2%)	53 (98%)
	2	55									
	3	40									
1,000 rDNA	1	37	138	77 (57%)	59 (43%)	32 (54%)	6 (19%)	26 (81%)	27 (46%)	NA	27 (100%)
	2	47									
	3	54									

Figure 4 Recombinant adeno-associated virus (rAAV)-ribosomal DNA (rDNA) versus rAAV-stuffer comparison in NIH3T3 cells. (**a**) Graphic representation of vectors used in this part of the study. Vectors encoded hPGK-GFP-pA cassette flanked by 1,000 bp rDNA or "stuffer" control DNA. Fragment of human factor IX (hFIX) cDNA was cloned downstream of pA to keep the overall size of the vectors identical to the 1,000 bp rAAV-rDNA/stuffer vectors used in the *in vivo* studies (see **Figure 2a**). The black arrow indicates the biotinylated oligo used during primer extension step of linear amplification-mediated-PCR (LAM–PCR). (**b**) Comparison of different conditions used to transduce the NIH3T3 cells. Cells were transduced at two different cell numbers (as indicated) and different multiplicities of infection (MOIs). The resulting green fluorescent protein (GFP) levels at the first time point (day 2 post transduction) are indicated. (**c**) GFP levels in cells transduced with recombinant adeno-associated virus (rAAV)-rDNA and rAAV-stuffer vectors throughout the course of the study. (**d**) Quantification of GFP levels at the last time point (day 43) of the study. (**e**) LAM-PCR data. DNA samples extracted from cells on day 43 were analyzed by LAM-PCR. See legend for **Table 1** for detailed explanation.

rDNA-GFP vectors. Similarly to what was observed in the *in vivo* studies, cells transduced with the rDNA vectors expressed lower levels of GFP even when slightly higher (Groups 1 and 3) or two times (Group 2) greater MOIs were used. The differences were qualitatively similar although not as great as those observed in the *in vivo* setting (compare to **Figure 2b–d**).

As expected, the GFP levels decreased rapidly within the first 11 days after transduction, and from day 11 until day 43 (the end of the experiment) the GFP levels were more stable in each group (**Figure 4c**). By TaqMan PCR analyses, the vector copies paralleled the GFP levels (data not shown). As shown in **Figure 4c** and **d**, on day 43 the percentage of GFP^+ cells in groups treated with rDNA vectors were significantly higher (1.7–2.5 times higher, two-way ANOVA $P = 0.0109$) than in the groups transduced with rAAV-stuffer vector.

While the higher GFP levels in the rAAV-rDNA groups paralleled our findings *in vivo*, we wanted to determine whether the observed differences were due to locus specific integration of the rDNA vectors. Thus, similarly to the *in vivo* experiments, we performed extensive LAM-PCR analysis on the DNA samples extracted on the last day of the experiment.

Figure 4e shows the summary of LAM-PCR results. Out of 61 informative rAAV-stuffer vector integration sites 7 (11%) were located in rDNA loci, whereas the remaining 54 clones (89%) were in non-rDNA loci. Out of those 54 clones, 1 (2%) was found to be integrated into *Fah* gene *via* homologous recombination. In contrast, out of 59 rAAV-rDNA vector integration sites 32 (54%) were found to be in rDNA, or a 4.9-fold increase in the rDNA specific integration when compared to the control vector. Six (19%) out of the 32 clones found in rDNA locus integrated *via* homologous recombination mechanism and the remaining 26 (81%) integrated into different rDNA loci. While the differences were not as great as the *in vivo* results, enhanced rDNA integration was demonstrated in cultured cells.

DISCUSSION

The choice of rDNA as a locus for rAAV targeting was not a random one. Our group and others have shown in the past that rAAV vectors integrate more frequently into rDNA loci,[20,21,28] an observation that might be related to the fact that rAAV vectors seem to uncoat and locate near the nucleoli.[29] The nucleoli form in discrete locations determined by the position of nucleolus organizer regions,[36] a chromosomal loci where multiple copies of the rRNA genes are located in tandemly repeated H-T units and span millions of base pairs.[37,38] This not only further increases the chance of targeted integration but, due to the redundant nature of the rDNA loci, also possibly lowers the risk of unintended toxicity related to disruption of some of the *rRNA* genes.[32,33] Data from *S. cerevisiae* show that the number of *rRNA* genes changes over time in each cell, with copies of the genes being lost through recombination events and then added back through gene duplication mechanisms.[39] To make the situation even more complicated, the ribosomal chromatin can exist in three different states: inactive (similar to heterochromatin) or one of the two euchromatin-like states: transcriptionally active or non-active state.[40] Due to these specific and unique characteristic of the rDNA loci, it was hard to predict what effect this highly active locus would have on transgene expression. Our persistent hFIX expression post-liver regeneration and the stable expression of a transgene from NIH3T3 cells transduced with rAAV-rDNA suggest that the integration events into this locus is stable.

During the course of our studies it was noticed that prior to liver regeneration rAAV-rDNA vectors expressed at lower levels than rAAV-stuffer control vectors (**Figure 1c** and **2b–d**). Similar behavior was independently observed by Wang *et al.*[34] (see accompanying paper) further substantiating this phenomenon. This drop in expression correlated with the length of rDNA flanks, with the vectors containing 1,000 bp rDNA expressing at the lowest level. Southern blot and TaqMan analyses of samples obtained during PH and samples harvested at the end of the study confirmed similar gene transfer between rAAV-rDNA and rAAV-stuffer groups (**Figure 3a** and **Supplementary Figure S3**). The fact that the differences in transgene expression levels between rDNA and stuffer groups were observed only before liver regeneration, when majority of AAV genomes exists in episomal forms, indicated that the observed difference applied mostly to expression from episomal rDNA vectors and not integrated provectors. This can thus explain the observed differences in hFIX persistence after PH/CCl_4 reported in **Figures 1d** and **2e**. Both groups, treated with rDNA and stuffer vectors, lost similar amounts of episomal vectors, but the rAAV-stuffer episomes expressed significantly higher level of hFIX than rAAV-rDNA episomes, thus leading to a more significant drop in hFIX levels in stuffer control groups.

Our data does not provide any explanation as to why transgene expression was significantly lower from episomal rAAV vectors containing long-stretches of rDNA sequences when compared to control vectors. This effect was specific to rDNA sequences, however, and not the size of flanking sequences, as no decrease of transgene expression was observed between 500, 750, and 1,000 bp stuffer vectors. rDNA sequences are very GC rich and, based on our experience, are difficult templates for DNA polymerases. One possible explanation could be that rDNA sequences affected single-stranded DNA (ssDNA) to double-stranded DNA (dsDNA) conversion, step required for transgene expression from episomal vectors, but not for provector integration. A delay in ssDNA to dsDNA conversion could also account for the increased frequency of homologous recombination of the rDNA vectors, as it has been shown that ssDNA participate in homologous recombination.[41–43] However, if rDNA flanking sequences did affect rAAV genome replication and the dsDNA formation step, it was relatively minor. This is because the rDNA sequences did not affect rAAV vector production (**Supplementary Figure S2**), and Southern blot analysis on transduced tissues (**Figure 3b–c** and **Supplementary Figure S1**) confirmed that rAAV-rDNA vector genomes were present in the expected episomal and proviral conformations. This further suggests that rDNA sequences had no apparent effect on vector uncoating and downstream steps leading to the episomal rAAV forms and/or vector integration.

The study addressed another very important point related to use of rAAV-rDNA vectors. As discussed earlier, the hFIX and eGFP data proved that rDNA integration of a transgene allows for stable long-term expression. However, it was important to determine if a transgene integrated into rDNA locus would express at a similar level to a transgene integrated at other non-rDNA

genomic locus. The observed similar levels of hFIX expression after liver regeneration (**Figure 2b–d**) together with the similar average VCN in liver (**Supplementary Figure S3**), indicates that unlike episomal provector genomes, integration of these vectors into the rDNA genomic locus did not have any measureable effect on the level of transgene expression.

Based on above discussion, one can imagine a potential alternative use for rDNA vectors. The fact that rAAV-rDNA vectors show significantly decreased transgene expression from episomal vectors but not from integrated provectors, without any detectable effect on episomal structure or vector titer, opens a possibility of generating a rAAV-rDNA based expression system which would allow for transgene expression only after vector integration into genomic loci.

Despite the fact that episomal expression from rAAV-rDNA vectors was significantly lower than from control stuffer episomal vectors, the most important finding of the study was the fact that each tested vector pair, based on LAM-PCR results, showed a significantly higher percentage of locus specific integration of the rAAV-rDNA vectors when compared to the rAAV-stuffer controls (**Table 1**). When evaluating the efficiency of rAAV-rDNA vector integration into rDNA loci (**Table 1** and **Figure 4e**), it is important to keep in mind that the numbers represent percentages of sequenced clones with the expected homologous versus nonhomologous integration and not a percentage of cells with targeted integration. In fact, the correct way to quantify the LAM-PCR result is to look at the relative increase in rDNA versus non-rDNA integrations for a given vector pair.

rAAV-stuffer vector integration site analysis revealed that 6–11% of amplified clones integrated into rDNA, compared to 65–84% of rDNA integration among rAAV-rDNA vectors integration sites, an 8–13-fold increase in rDNA specific integration between rAAV-stuffer and rAAV-rDNA vectors *in vivo*. Using our previous data showing that ~3% of rAAV vectors integrate randomly into rDNA,[21] we estimated that the rAAV-rDNA vectors integrate into rDNA at frequency of 23–39%. This is in agreement with data reported by Wang *et al.*[34] (see accompanying paper), who used rAAV-rDNA-hFah vectors in Fah-deficient mice and showed that rAAV-rDNA vectors integrate at rate 10–30 times higher than control vectors. Using a plasmid rescue strategy, although the recovery yield was small, they found integrations into the rDNA locus at a frequency of 33% consistent with our estimation by LAM-PCR of 23 to 39%.

Analysis of the LAM-PCR amplicon sequences provided detailed information not only on the exact genomic location of each informative clone (**Supplementary Figure S5** and **Supplementary Table S1**) but also allowed us to gain some insight into specific rearrangements of the endogenous and vector encoded sequences that occur at the integration site (**Supplementary Figure S4**). Integrations were often associated with deletions and truncations of AAV ITR, as well as deletions of the vector encoded flanking sequences and/or endogenous sequences adjacent to the integration site.

Among all analyzed integrations of the rAAV-stuffer vectors, 2 out of 113 (1.8%) were integrated into *Fah* gene (1 in 500 bp and 1 in 750 bp group) *via* homologous recombination. No nonhomologous integrations into *Fah* gene were identified in the study. Out of all rDNA integrations of the rAAV-rDNA vector tested in the study, 11–17% occurred through homologous recombination, and 83–89% occurred *via* nonhomologous integration mechanisms. Due to the high number of rDNA genes per genome, we do not consider this substantial fraction of nonhomologous off-target integrations as a potential risk factor. The much higher ratio of homologous recombination for the rDNA vectors compared to stuffer vectors (11–17% for rDNA vectors versus 1.8% for stuffer vectors) was most probably a result of the specific cellular localization of AAV vectors uncoating and the fact that rDNA loci exist in hundreds of copies throughout the genome, compared to only two copies of the Fah gene. The ratio of specific rDNA to non-rDNA integrations was the highest for rAAV-rDNA-750 and 1,000 vectors with 84% and 78% of the vectors integrated into rDNA, respectively. For rAAV-rDNA-1,000 vector this represented a 13-fold increase when compared to the rAAV-stuffer-1,000 control (78% versus 6%) (**Table 1**). It is interesting that studies with rAAV vector integration into AAVS1 site showed that even under optimal conditions of Rep68/78 expression 40–70% of all integrations occurred in the AAVS1 site in tissue culture cells.[44] This shows that the numbers observed in our study are within the same range as those observed for a more cumbersome system requiring both rAAV and Rep proteins *in trans*.

The fact that Southern blot analysis was not sensitive enough to provide any specific information on differences in provector integration profile between rDNA and control vectors (**Supplementary Figure S1**) was not a surprising one. A high percentage of similar AAV provectors integrations at a single precise locus would be required for a discrete band to be visible on Southern blot. Since rAAV form higher concatemers (**Figure 3c**) and LAM-PCR data demonstrated that majority (83–89%) of rAAV-rDNA vectors integrated into rDNA loci *via* mechanisms other than homologous recombination, we would not expect to be able to detect homologous rDNA integration using Southern blot analysis. This is in agreement with observation made by Wang *et al.*[34] (see accompanying paper) who had to turn to a more sensitive PCR approach in order to detect site-specific 28S integration.

When comparing our results from *in vivo* and *in vitro* studies (compare **Table 1** to table in **Figure 4e**) it becomes apparent that the effects the rDNA flanking sequences had on locus specific integration were more apparent in the *in vivo* studies. This could be due to the different target cells (hepatocytes versus fibroblasts) or presence of different transgene cassettes in the *in vivo* and *in vitro* studies.[45] Furthermore, the amount of time from transduction to cell division was significantly different in those two studies. *In vivo*, liver regeneration was induced by PH 86 days after transduction, while *in vitro* despite the fact that cells were transduced at high density, the cell numbers were constantly increasing even after the cells reached confluency. The last major difference was the AAV serotype used to pseudotype the vectors. AAV8 was used in the *in vivo* studies, while AAVDJ[35] was used *in vitro* due to its enhanced transduction profile on tissue culture cells. Extensive work from the Russell group suggests that only the single-stranded genomes participate in homologous recombination.[41–43] Although this observation was based on the parvovirus minute virus of mice and not rAAV, it is highly probable that the same principle could apply to rAAV. The use of different AAV serotypes could lead to differences in kinetics of AAV

uncoating and conversion of the ssDNA genome to dsDNA[46] and thus affect the frequency of homologous recombinations.

The exact mechanism of rAAV-rDNA vectors integration still needs to be investigated. A significant fraction of rAAV-rDNA vectors integrated into rDNA locus *via* homologous recombination (11–17%) indicating that the rDNA flanks present in the vector are directly responsible for some of the retargeting of the rDNA vectors. However, 83–89% of rAAV-rDNA vectors integrated into rDNA loci *via* unknown nonhomologous integration mechanisms that still requires further studies to be fully understood. It also remains to be tested if the rAAV-rDNA vectors reported here would show the same behavior if other rDNA sequences were used to flank the transgene, and if other AAV serotypes were used to deliver the flanked transgene. Despite all the unanswered questions and the need to further optimize transduction conditions in rapidly dividing cells, our data clearly show that the rAAV-rDNA vector breach the gap between rAAV and lentiviral vectors and open a door for the use of rAAV vectors in context of dividing cells *in vivo* and *in vitro/ex vivo*.

MATERIALS AND METHODS

AAV vector construction, vector production, and vector titration. All vectors used in the *in vivo* studies carried the 2.1 kb hFIX cassette composed of TTR promoter/synthetic enhancer, hFIX cDNA + minimal Intron 1,[47] and synthetic pA, flanked by murine 28S rDNA sequences. The vector used in the pilot study contained 1,157 bp rDNA downstream and 879 bp upstream of the hFIX cassette (from position 2,266–3,432 and 3,427–4,306 of the murine 28S ribosomal RNA, NCBI Reference Sequence NR_003279.1). The rAAV-rDNA vectors used in the second part of the study contained rDNA flanks as follow: 1,000 bp rDNA vector 5′ flank (2,425–3,433) 3′flank (3,427–4,426), 750 bp vector 5′flank (2,681–3,433) 3′flank (3427–4178) and 500 bp vector 5′flank (2,933–3,433) 3′flank (3,427–3,928) (all positions based on 28S sequence NCBI Reference Sequence NR_003279.1). The 5′ and 3′ flanks overlapped by 6 nucleotide in the middle of the I-*Ppo*I endonuclease recognition sequence (CTCTCTTAAGGTAGC) at the *Afl*II endonuclease recognition site. Control vectors used in the study contained the same hFIX cassette flanked by stuffer DNA sequences from murine fumarylacetoacetase (*FAH*) gene. The 3′ flank in the control vector used in the pilot study contained sequences between position 647 and 1,009 (NCBI Reference Sequence: NT_039433.7), whereas the 5′ flank spanned sequences from 1,004–2,590. The control vectors used in the second study contained FAH DNA sequences (5′ followed by 3′): 1,010–2,014 and 2,003–3,003 for the 1,000 bp vector, 1,251–2,014 and 2,003–27,258 for the 750bp vector and 1,501–2,014 and 2,003–2,502 for the 500 bp control vector.

The vectors used in the NIH3T3 studies contained identical 1,000 bp rDNA or FAH stuffer flanks as described for FIX vector. The hFIX cassette was replaced with hPGK-eGFP-bGHpA cassette to allow for convenient monitoring of transgene expression by FACS. The 880-bp fragment of the *hFIX* gene was included downstream of the pA to bring the total size of the vector to 4.7kb and to enable the use of an identical Southern blot probe and TaqMan primers/probe for all vectors used in the study.

rAAV vectors were produced by triple transfection into 293 cells as previously described,[48] viral DNA was extracted by the sodium iodide method, using a DNA Extractor kit (Wako Chemicals USA, Richmond, VA) and titrated by a quantitative dot blot assay.

Animal work and transgene expression in vivo. Animal work was performed in accordance to the guidelines for animal care at Stanford University. Six–eight-week-old female C57BL/6 mice were purchased from the Jackson Laboratory (Bar Harbor, ME). rAAV8-rDNA/Stuffer-hFIX vectors were administered *via* retro-orbital injection in 200 µl of 1xPBS. A two-thirds PH was performed as described.[19] Blood was collected at the indicated time points *via* retro-orbital bleeding, and plasma hFIX levels were determined *via* an enzyme-linked immunosorbent assay as described previously.[49]

Analysis of VCNs in liver and other tissues. Total DNA was extracted from liver and other tissues as described by Sambrook *et al*. and used for Southern blot analysis[50] using a [^{32}P]dCTP-labeled hFIX specific probe and/or TaqMan. Twenty microgram of genomic DNA was digested with *BamH*I (double cutter), *Spe*I (single cutter), or *Sca*I (zero cutter). Genomic DNA from untreated mouse spiked with plasmid DNA served as a copy number standards. Multiplex real-time TaqMan PCR was performed and analyzed on ABI 7500 PCR System (Applied Biosystems, Foster City, CA) using hFIX specific probe (5′-GTGCCATTTCCATGTGGAAGAGTTTCTGT-3′) and primers (Forward; 5′- AACCAGAAGTCCTGTGAACCAGCA-3′, reverse; 5′- GCACGGGTGAGCTTAGAAGTTTGT-3′), and mouse β-actin probe (5′-GCTGTGTTCTTGCACTCCTTGCATGT-3′) and primers (forward; 5′-TGAGACTCCCAGCACACTGAACTT-3′, and reverse; 5′-ACACTCAGGGCAGGTGAAACTGTA-3′).

LAM-PCR. Provector integration site analysis was performed on genomic DNA extracted from liver samples at the end of an experiment using LAM-PCR as previously described[51] with minor modifications. In brief, 100–200 ng genomic DNA was used for 2 × 50 cycles of linear primer extension using LAM-SE1-Bio or LAM-FIX1-Bio biotinylated oligos, with fresh enzyme and dNTPs added after the first 50 cycles, using Phusion High-Fidelity DNA Polymerase (NEB, Ipswich, MA) and the following program: 98 °C 3 minutes, 50 cycles of 98 °C 10 seconds, 72 °C 2.5 minutes, followed by final extension at 72 °C for 10 minutes. After magnetic capture and second strand synthesis the dsDNA products were digested with EcoRV and ligated to dsDNA linker formed by annealing of Linker1 and Linker 2 oligos. The product of the ligation served as a template for PCR amplification using LAM-SE2 or LAM-FIX2 and LAM-LK1 oligos and PCR program: 98 °C 30 seconds, 34 cycles of 98 °C 10 seconds, 65 °C 20 seconds, 72 °C 2.5 minutes, final extension 72 °C 10 minutes. 1/50th of the 1st PCR was used for nested PCR using LAM-SE3 or LAM-FIX3 and LAM-LK2 oligos: 98 °C 30 seconds, 34 cycles of 98 °C 10 seconds, 58 °C 20 seconds, 72 °C 2.5 minutes, final extension 72 °C 10 minutes. Fraction of the PCR product was analyzed on agarose gel and the remaining PCR reaction was used for TOPO cloning (Invitrogen, Carlsbad, CA) and subsequent sequencing.

LINKER1	TTCAGACCCGGGAGATCTGAATTCAGTGGCACA GCAGTTAGG
LINKER2	CCTAACTGCTGTGCCACTGAATTCAGATCTCCC GGGTC
LAM-LK1	CTGAATTCAGTGGCACAGCAG
LAM-LK2	TCAGTGGCACAGCAGTTAGG
LAM-SE1-Bio	AATCGCGGCCTCGAATCAATATTCGCGAG
LAM-SE2	ATAATTACCAGCGCGCGCTG
LAM-SE3	CCGCGCTGGTAATTATTAACTC
LAM-FIX1-Bio	CGTGTTGTTGGTGGAGAAGATGC
LAM-FIX2	CTTGGCAGGTTGTTTTGAATGG
LAM-FIX3	TGGTAAAGTTGATGCATTCTGTGG

NIH3T3 transduction and transgene expression. NIH3T3 cells were maintained under standard tissue culture conditions. For transduction, cells were split and seeded at two densities into 6-well plates. Approximately 24 hours later, cells were counted and transduced with rAAVDJ-stuffer/rDNA-hPGK-eGFP vectors at multiple MOIs (see **Figure 4b**). Transduction was allowed to proceed for 16 hours. Forty eight hours later eGFP levels were analyzed by FACS. For the remaining of the study, cells were split and FACSed on average every 3–4 days and cells were never allowed to reach confluency, to ensure constant cell divisions.

Statistical analyses. All statistical analyses were done by the two-way ANOVA test using GraphPad Prism 5 software.

SUPPLEMENTARY MATERIAL

Figure S1. Southern blot analysis of vector genome structure in liver samples from animals post-PH/CCl4 and control animals without liver regeneration followed for the same period of time.
Figure S2. Titer comparison of vectors used in the study.
Figure S3. AAV copy number and vector tropism comparison.
Figure S4. Representative rescued junction sequences.
Figure S5. Genomic integration profiles for rAAV-rDNA and rAAV-stuffer vectors.
Table S1. Detailed analysis of non-rDNA integrations sites identified in the study.
Supplementary Data.

ACKNOWLEDGMENTS

We thank Dr Mathur for her assistance. This work was supported by NIH R01 HL064274 and HL092096 to M.A.K. Leszek Lisowski was supported in part by the Berry Fellowship Foundation.

REFERENCES

1. Blömer, U, Naldini, L, Kafri, T, Trono, D, Verma, IM and Gage, FH (1997). Highly efficient and sustained gene transfer in adult neurons with a lentivirus vector. *J Virol* **71**: 6641–6649.
2. Baum, C, Düllmann, J, Li, Z, Fehse, B, Meyer, J, Williams, DA *et al.* (2003). Side effects of retroviral gene transfer into hematopoietic stem cells. *Blood* **101**: 2099–2114.
3. Hacein-Bey-Abina, S, von Kalle, C, Schmidt, M, Le Deist, F, Wulffraat, N, McIntyre, E *et al.* (2003). A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. *N Engl J Med* **348**: 255–256.
4. Rosenber, N and Jolicoeur, P (1997). Retroviral pathogenesis. In: Coffin, AM, Hughes SH and Varmus HE eds. *Retroviruses*. Cold Spring Harbor Laboratory Press: New York. pp 475–585.
5. Maguire, AM, High, KA, Auricchio, A, Wright, JF, Pierce, EA, Testa, F *et al.* (2009). Age-dependent effects of RPE65 gene therapy for Leber's congenital amaurosis: a phase 1 dose-escalation trial. *Lancet* **374**: 1597–1605.
6. Nathwani, AC, Tuddenham, EG, Rangarajan, S, Rosales, C, McIntosh, J, Linch, DC *et al.* (2011). Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. *N Engl J Med* **365**: 2357–2365.
7. Gao, G, Vandenberghe, LH and Wilson, JM (2005). New recombinant serotypes of AAV vectors. *Curr Gene Ther* **5**: 285–297.
8. Chiorini, JA, Kim, F, Yang, L and Kotin, RM (1999). Cloning and characterization of adeno-associated virus type 5. *J Virol* **73**: 1309–1319.
9. Chiorini, JA, Yang, L, Liu, Y, Safer, B and Kotin, RM (1997). Cloning of adeno-associated virus type 4 (AAV4) and generation of recombinant AAV4 particles. *J Virol* **71**: 6823–6833.
10. Gao, G, Vandenberghe, LH, Alvira, MR, Lu, Y, Calcedo, R, Zhou, X *et al.* (2004). Clades of Adeno-associated viruses are widely disseminated in human tissues. *J Virol* **78**: 6381–6388.
11. Gao, GP, Alvira, MR, Wang, L, Calcedo, R, Johnston, J and Wilson, JM (2002). Novel adeno-associated viruses from rhesus monkeys as vectors for human gene therapy. *Proc Natl Acad Sci USA* **99**: 11854–11859.
12. Mori, S, Wang, L, Takeuchi, T and Kanda, T (2004). Two novel adeno-associated viruses from cynomolgus monkey: pseudotyping characterization of capsid protein. *Virology* **330**: 375–383.
13. Muramatsu, S, Mizukami, H, Young, NS and Brown, KE (1996). Nucleotide sequencing and generation of an infectious clone of adeno-associated virus 3. *Virology* **221**: 208–217.
14. Rutledge, EA, Halbert, CL and Russell, DW (1998). Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2. *J Virol* **72**: 309–319.
15. Schmidt, M, Voutetakis, A, Afione, S, Zheng, C, Mandikian, D and Chiorini, JA (2008). Adeno-associated virus type 12 (AAV12): a novel AAV serotype with sialic acid- and heparan sulfate proteoglycan-independent transduction activity. *J Virol* **82**: 1399–1406.
16. Xiao, W, Chirmule, N, Berta, SC, McCullough, B, Gao, G and Wilson, JM (1999). Gene therapy vectors based on adeno-associated virus type 1. *J Virol* **73**: 3994–4003.
17. Michelfelder, S and Trepel, M (2009). Adeno-associated viral vectors and their redirection to cell-type specific receptors. *Adv Genet* **67**: 29–60.
18. Büning, H, Perabo, L, Coutelle, O, Quadt-Humme, S and Hallek, M (2008). Recent developments in adeno-associated virus vector technology. *J Gene Med* **10**: 717–733.
19. Nakai, H, Yant, SR, Storm, TA, Fuess, S, Meuse, L and Kay, MA (2001). Extrachromosomal recombinant adeno-associated virus vector genomes are primarily responsible for stable liver transduction in vivo. *J Virol* **75**: 6969–6976.
20. Nakai, H, Montini, E, Fuess, S, Storm, TA, Grompe, M and Kay, MA (2003). AAV serotype 2 vectors preferentially integrate into active genes in mice. *Nat Genet* **34**: 297–302.
21. Nakai, H, Wu, X, Fuess, S, Storm, TA, Munroe, D, Montini, E *et al.* (2005). Large-scale molecular characterization of adeno-associated virus vector integration in mouse liver. *J Virol* **79**: 3606–3614.
22. Donsante, A, Miller, DG, Li, Y, Vogler, C, Brunt, EM, Russell, DW *et al.* (2007). AAV vector integration sites in mouse hepatocellular carcinoma. *Science* **317**: 477.
23. Dyall, J, Szabo, P and Berns, KI (1999). Adeno-associated virus (AAV) site-specific integration: formation of AAV-AAVS1 junctions in an *in vitro* system. *Proc Natl Acad Sci USA* **96**: 12849–12854.
24. Kogure, K, Urabe, M, Mizukami, H, Kume, A, Sato, Y, Monahan, J *et al.* (2001). Targeted integration of foreign DNA into a defined locus on chromosome 19 in K562 cells using AAV-derived components. *Int J Hematol* **73**: 469–475.
25. Recchia, A, Parks, RJ, Lamartina, S, Toniatti, C, Pieroni, L, Palombo, F *et al.* (1999). Site-specific integration mediated by a hybrid adenovirus/adeno-associated virus vector. *Proc Natl Acad Sci USA* **96**: 2615–2620.
26. Rinaudo, D, Lamartina, S, Roscilli, G, Ciliberto, G and Toniatti, C (2000). Conditional site-specific integration into human chromosome 19 by using a ligand-dependent chimeric adeno-associated virus/Rep protein. *J Virol* **74**: 281–294.
27. Rizzuto, G, Gorgoni, B, Cappelletti, M, Lazzaro, D, Gloaguen, I, Poli, V *et al.* (1999). Development of animal models for adeno-associated virus site-specific integration. *J Virol* **73**: 2517–2526.
28. Miller, DG, Trobridge, GD, Petek, LM, Jacobs, MA, Kaul, R and Russell, DW (2005). Large-scale analysis of adeno-associated virus vector integration sites in normal human cells. *J Virol* **79**: 11434–11442.
29. Johnson, JS and Samulski, RJ (2009). Enhancement of adeno-associated virus infection by mobilizing capsids into and out of the nucleolus. *J Virol* **83**: 2632–2644.
30. Sakai, K, Ohta, T, Minoshima, S, Kudoh, J, Wang, Y, de Jong, PJ *et al.* (1995). Human ribosomal RNA gene cluster: identification of the proximal end containing a novel tandem repeat sequence. *Genomics* **26**: 521–526.
31. Long, EO and Dawid, IB (1980). Repeated genes in eukaryotes. *Annu Rev Biochem* **49**: 727–764.
32. Johansen, SD, Haugen, P and Nielsen, H (2007). Expression of protein-coding genes embedded in ribosomal DNA. *Biol Chem* **388**: 679–686.
33. Kobayashi, T (2006). Strategies to maintain the stability of the ribosomal RNA gene repeats–collaboration of recombination, cohesion, and condensation. *Genes Genet Syst* **81**: 155–161.
34. Wang, Z, Lisowski, L, Finegold, MJ, Nakai, H, Kay, MA and Grompe, M (2012). AAV vectors containing rDNA homology display increased chromosomal integration and transgene persistence. *Mol Ther* (in this issue).
35. Grimm, D, Lee, JS, Wang, L, Desai, T, Akache, B, Storm, TA *et al.* (2008). *In vitro* and *in vivo* gene therapy vector evolution via multispecies interbreeding and retargeting of adeno-associated viruses. *J Virol* **82**: 5887–5911.
36. Prieto, JL and McStay, B (2005). Nucleolar biogenesis: the first small steps. *Biochem Soc Trans* **33**(Pt 6): 1441–1443.
37. Ritossa, FM and Spiegelman, S (1965). Localization of dna complementary to ribosomal rna in the nucleolus organizer region of *Drosophila melanogaster. Proc Natl Acad Sci USA* **53**: 737–745.
38. Wallace, H and Birnstiel, ML (1966). Ribosomal cistrons and the nucleolar organizer. *Biochim Biophys Acta* **114**: 296–310.
39. Kobayashi, T, Heck, DJ, Nomura, M and Horiuchi, T (1998). Expansion and contraction of ribosomal DNA repeats in Saccharomyces cerevisiae: requirement of replication fork blocking (Fob1) protein and the role of RNA polymerase I. *Genes Dev* **12**: 3821–3830.
40. Huang, S, Rothblum, LI and Chen, D (2006). Ribosomal chromatin organization. *Biochem Cell Biol* **84**: 444–449.
41. Russell, DW, Miller, AD and Alexander, IE (1994). Adeno-associated virus vectors preferentially transduce cells in S phase. *Proc Natl Acad Sci USA* **91**: 8915–8919.
42. Hirata, RK and Russell, DW (2000). Design and packaging of adeno-associated virus gene targeting vectors. *J Virol* **74**: 4612–4620.
43. Hendrie, PC, Hirata, RK and Russell, DW (2003). Chromosomal integration and homologous gene targeting by replication-incompetent vectors based on the autonomous parvovirus minute virus of mice. *J Virol* **77**: 13136–13145.
44. Daya, S and Berns, KI (2008). Gene therapy using adeno-associated virus vectors. *Clin Microbiol Rev* **21**: 583–593.
45. Hirata, R, Chamberlain, J, Dong, R and Russell, DW (2002). Targeted transgene insertion into human chromosomes by adeno-associated virus vectors. *Nat Biotechnol* **20**: 735–738.
46. Thomas, CE, Storm, TA, Huang, Z and Kay, MA (2004). Rapid uncoating of vector genomes is the key to efficient liver transduction with pseudotyped adeno-associated virus vectors. *J Virol* **78**: 3110–3122.
47. Nakai, H, Herzog, RW, Hagstrom, JN, Walter, J, Kung, SH, Yang, EY *et al.* (1998). Adeno-associated viral vector-mediated gene transfer of human blood coagulation factor IX into mouse liver. *Blood* **91**: 4600–4607.
48. Grimm, D (2002). Production methods for gene transfer vectors based on adeno-associated virus serotypes. *Methods* **28**: 146–157.
49. Walter, J, You, Q, Hagstrom, JN, Sands, M and High, KA (1996). Successful expression of human factor IX following repeat administration of adenoviral vector in mice. *Proc Natl Acad Sci USA* **93**: 3056–3061.
50. Sambrook, J, Fritsch, EF, and Maniatis, T (1989). *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory: Cold Spring Harbor, NY.
51. Schmidt, M, Schwarzwaelder, K, Bartholomae, C, Zaoui, K, Ball, C, Pilz, I *et al.* (2007). High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). *Nat Methods* **4**: 1051–1057.