A NEWLY DESIGNED AAV-RDNA VECTOR FOR PERSISTENT THERAPY IN GENETIC LIVER DISEASES

By

Zhongya Wang

A DISSERTATION

Presented to the Department of Cell and Developmental Biology

and the Oregon Health & Science University

School of Medicine

in partial fulfillment of

the requirements for the degree of

Doctor of Philosophy

April 6, 2012

School of Medicine

Oregon Health & Science University

CERTIFICATE OF APPROVAL

This is to certify that the PhD dissertation of

Zhongya Wang

has been approved

Mentor/Advisor

Member

Member

Member

Member

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AAV	adeno-associated virus
AFP	α-fetoprotein
AKU	alkaptonuria
DNA	deoxyribonucleic acid
FAA	fumarylacetoacetate
FAH	fumarylacetoacetate hydrolase
F.IX	factor IX
GS	glutamine synthetase
HCC	hepatocellular carcinoma
HGA	homogentisic acid
HPD	homogentisic acid dioxygenase
HT1	hereditary tyrosinemia type 1
ITR	inverted terminal repeats
MAA	maleylacetoacetate
MAI	maleylacetoacetate isomerase
MOI	multiplicity of infection
NTBC	2-(2-nitro-4-trifluoro-methylbenzyol)-1,3 cyclohexanedione
PBS	phosphate buffered saline

PCR	polymerase chain reaction
rAAV	recombinant adeno-associated virus
rDNA	ribosomal deoxyribonucleic acid
RSV	respiratory syncytial virus
SA	succinylacetone
SAA	succinylacetoacetate
TAT	tyrosine amino transferase
WT	wild type
VG	viral genome

ACKNOWLEDGEMENTS

I would like to thank my mentor, Dr. Markus Grompe for his support and training. Markus is an outstanding scientist with extraordinary scientific vision. His profound knowledge and insightful view benefited the graduate students.

I would like to thank my thesis advisory committee members: Dr. Michael Chapman; Dr.Peter Kurre; Dr.Mark Slifka; Dr.Tim Stout. They all have been providing me with useful suggestions and advice in the last several years.

I want to thank the Grompe lab members for their support and help. In the last several years, they have helped me a lot with my experiments. I want to thank Craig and Branden for their help with my thesis. Also I want to thank our collaborators: Dr. Mark Kay from Stanford and Dr. Hiroyuki Nakai of OHSU for their help with my projects.

Finally, I want to thanks my parents in China for their support. I want to thank my wife Xi and my son Zizheng. Their smiles and hugs have been my motive. Without their support and encouragement, this thesis cannot be accomplished.

ABSTRACT

AAV is one of the most promising gene therapy vectors. There are several clinical trials with AAV vectors for different genetic disease. For gene therapy in genetic diseases, the diseases' nature requires persistent expression of the transgene. One of the main disadvantages for AAV vectors is that a vast majority of AAV vectors *in vivo* exist as episomal, as the integration frequency is very low.

In this dissertation, we engineered a new integrating AAV-rDNA vector. We firstly used the HT1 disease model to characterize this vector. At the DNA level, AAV-rDNA-FAH were found to give 10-30 times more integrated viral genome than sizable control AAV-stuffer-FAH at two different doses by real time PCRs. At protein level, we used anti-FAH immunostaining to stain the AAV infected liver and found that AAV-rDNA-FAH gave 18 times more FAH positive nodules than control vectors. At functional level, AAV-rDNA-FAH vectors were able to rescue HT1 mice, which will die without any treatment, at 10-30 times lower dose than comparable control. Importantly, rescued $Fah^{-/-}$ mice all displayed normal liver function test as similar to wild type mice no matter rescued by high dose or low dose. These data provide evidence that AAV-rDNA vectors gave about 10-30 times more viral vector integration with functional transgene expression. These results have been confirmed in two different disease models, HT1 mouse model and hemophilia B mouse model.

In this dissertation, we also characterized that the minimum homologous rDNA length required for efficient integration will be 750 bp. In both disease models, we

detected the existence of site-specific integration. By LAM-PCR, shuttle vector strategy, PCR base next generation sequencing, we identified different sets of integration sites of AAV-rDNA vectors, all of which show significant enrichment of AAV-rDNA integration in rDNA repeats by more than 10 times. This provided consistent results with DNA level, protein level, and functional level data generated in two disease models.

Taken together, we have designed an efficient integrating vector, which gives 10 times more integration with functional transgene expression. When treated for genetic liver diseases or other genetic diseases, which required persistent transgene expression, this vector should be considered and could provide a significantly better therapy than regular AAV vectors.

Chapter 1

Introduction

I. Wild type AAV and AAV vectors

AAV general biology

Adeno-associated virus (AAV) belongs to the *Dependovirus* genus of the *Parvoviridae* family (Berns and Giraud, 1996). The wild type form of AAV is a non-enveloped virus with a 22 nm capsid, which contains a single stranded DNA genome, approximately 4.7 kb in length, with either plus or minus strand orientation. The life cycle of wild type AAV has two phases: a productive phase, in which AAV requires facilitation by a helper virus, such as an adenovirus, in order to replicate and then infect new host cells; and a latent phase, in which the AAV genome can integrate into the host genome and replicate within the host genome (Berns and Giraud, 1996).

AAV genome

The AAV genome is composed of two open reading frames (ORFs) and two inverted terminal repeats (ITRs). The left ORF encodes four replication proteins: Rep 48, Rep 52, Rep 68 and Rep 78, which are responsible for latent integration and regulate the replication of the AAV genome. The right ORF encodes structural proteins that form the capsid of the AAV particle, which are also the key features that distinguish different serotypes. The 147 bp ITRs form a T-shaped secondary structure with a free 3’ hydroxyl group, which is necessary for the initiation of AAV genome replication (Berns and Giraud, 1996; Muzyczka et al., 1984).

AAV serotypes

Since the isolation of the first infectious AAV serotype 2 (AAV2) clone in 1982

(Samulski et al., 1982), 11 AAV serotypes and many AAV variants have been identified from human tissues or from adenovirus stocks, as well as from non-human primate (Gao et al., 2002), cow (Coria and Lehmkuhl, 1978), chicken (Dawson et al., 1982) and goat (Clarke et al., 1979; Qiu et al., 2006; Wu et al., 2006b). Most vectors used in AAV gene therapy are pseudotyped AAV vectors, which carry a AAV2 derived genome with serotype specific capsids, which has been termed a cross-packaging strategy. To date, many different AAV serotypes are available with variable transduction efficacies and tissue tropisms. These features make AAV very useful vectors for tissue-specific gene therapy for many diseases. Out of the 11 serotypes, AAV1 to AAV9 have been extensively studied and are the serotypes most commonly used in gene therapy studies currently (DiPrimio et al., 2008; Frank et al., 2009; Wu et al., 2006b). Among them, AAV2, AAV8 and AAV9 are mostly used in gene therapy directed at the liver and muscle (Gao et al., 2002; Lipiec et al., 2009; Rebuffat et al., 2009; Zincarelli et al., 2008). AAV4 and AAV5 are mostly used in the CNS and eye (Zincarelli et al., 2008). Since the capsids of AAV are responsible for transduction specificity, enormous effort has been put on capsid manipulation, either modifying the key residues on the capsids, or designing hybrid particles (Mitchell et al., 2010; Wu et al., 2006b). For example, a hybrid serotype AAV-DJ, which is a type 2/type 8/type 9 chimera, has been found to have higher transduction efficiency in mammalian cell lines than all other standard serotypes (Grimm et al., 2008).

AAV gene therapy

The successful cloning of the AAV virus has made it possible to engineer recombinant AAV vectors capable of carrying foreign DNA that is expressed in

mammalian cells (Muzyczka et al., 1984). The unique ability of AAV to exist either in a productive phase (episomal) or latent phase (integrated into host chromosome) and its non-pathogenicity are the advantages of AAV for its use as a gene therapy vector (McCarty et al., 2004). Another distinct advantage of AAV is that both wild-type and engineered vectors have a high infectivity for both non-dividing and dividing cells, which outperform another gene therapy vector, the retrovirus. The capability of infecting quiescent cells significantly boosts its potential for application, as in many adult tissues most of the cells remain in a quiescent state (Samulski, 2003; Stilwell and Samulski, 2003).

For AAV-mediated gene therapy, the original studies focused on two things: achieving efficacy in animal models and large scale production and purification of AAV vectors. The first question has been tested and solved in many disease models that don't require integration, although the large-scale production of AAV still remains partially unsolved (Hasbrouck and High, 2008). With the transition from animal models to clinical trials, the immune response to AAV vectors and transgenes has become the central issue (Manno et al., 2006; Mingozzi and High, 2007). This issue can be divided into two problems: 1) pre-existing antibody to AAV2 or other vector types in patients because of prior exposure to wild type virus (Calcedo et al., 2011); 2) CD8 T-cell mediated immune response to AAV vectors or transgenes after virus administration (Gao et al., 2009; Mingozzi et al., 2007). To solve these issues, many studies have been performed which either use the vectors without a pre-existing neutralizing antibody (Nathwani et al., 2011), or re-administration with different serotypes (Vandenberghe et al., 2009).

One of the primary aims in this dissertation was to design a new AAV vector with high frequency integration into the host genome, which can provide persistent transgene expression at therapeutic levels and makes re-administration unnecessary, therefore bypassing the re-administration related immune-response.

II. AAV integrations and homologous recombination

AAV integration

A vector capable of permanently introducing exogenous DNA into the host genome would be superior to episomal delivery of target genes for gene therapy of genetic diseases. Early usage of AAV as a gene therapy vector was driven largely by its ability to integrate into host chromosomes, which allows stable expression of the transgene and transference to daughter cells during cell division (McCarty et al., 2004). However recombinant AAV (rAAV) integration is inefficient with an integration frequency of only $0.2\text{-}1.0\text{x}10^{-3}$ per infectious vector genome in a Rep-independent manner (McCarty et al., 2004). In fact even wild type AAV integrates into chromosome 19 inefficiently at a rate of only 0.1-0.5% per infectious unit (Flotte et al., 1994; McCarty et al., 2004). Comparison of AAVS1, the Ch19 integration site sequence of wild type AAV, with cloned AAV provirus sequence showed that only small (4-5 bp) homologies were found between AAVS1 and the virus sequence (Kotin et al., 1990; Snyder et al., 1990; Young et al., 2000a). In addition, sequencing the junction fragment and adjacent chromosomal region revealed that chromosomal rearrangement and deletion were frequent around the integration sites (Kotin et al., 1990; Samulski et al., 1991; Snyder et al., 1990). Together, these findings suggest that Rep-dependent site-specific integration occurred in non-homologous end joining manner and might also involve other mechanisms. Importantly, very weak chromosome 19 site-specific PCR signals from the liver tissue of two intravenously infected animals suggested that the integration frequency for wt AAV *in vivo* was low (McCarty et al., 2004; Weitzman et al., 2003). More importantly, the Rep-dependent site-specific integration happens only with wild type

AAV, not with the recombinant AAV used for gene therapy (McCarty et al., 2004). Development of hybrid Rep containing vectors has been one of the central interests. However, introduction of the AAV replication gene (*Rep*) raises potential safety concerns for clinical application.

Compared with Rep-mediated site-specific integration in wild type AAV, rAAV mediated random integration is spread throughout the whole host genome, with a preference for actively transcribed genes and chromosome breakage points, which raises concerns about malignant transformation as a result of random integration into promoters or tumor suppressor genes (Miller et al., 2005; Nakai et al., 2003; Nakai et al., 2005b). For example, AAV integration has been found to be associated with hepatocellular carcinoma (HCC) through dysregulation of microRNA in a mouse model of mucopolysaccharidosis (Donsante et al., 2007).

AAV mediated homologous recombination

It has been reported that AAV carrying homologous sequences were able to correct point mutations within the host genome via homologous recombination (Miller et al., 2003). In addition, Miller *et al.* reported an increased efficiency in correction of point mutations in the presence of double-strand breakage sites *in vitro* in cultured cells (Miller et al., 2003, 2004; Miller et al., 2002). Further support for correction of gene mutations by homologous recombination also comes from Miller *et al.* and Paulk *et al.*, who reported successful targeted correction *in vivo* in three different mouse models including *RosaAnZΔ4*, *MPS VII* and HT1(Donsante et al., 2007; Miller et al., 2006; Paulk et al., 2010). They noted a correction efficiency of 10^{-4}-10^{-5} (10^{-3} in the HT1 mouse model), that was significantly below what is required for therapeutic correction and noted that the

efficiency of random integration was higher than that of targeted integration (Miller et al., 2006; Paulk et al., 2010). This low efficacy of correction currently limits the application of rAAV based gene correction for use in clinical gene therapy trials. An alternative strategy in which a novel rAAV is capable of site-specific integration with high efficiency is greatly needed and would be of value for future human gene therapy trials.

III. The tyrosine catabolic pathway and hereditary tyrosinemia type I

Tyrosine catabolic pathway

Tyrosine (abbreviated as Tyr or Y) or 4-hydroxyphenylalanine, is a non-essential amino acid in mammals. Its catabolism consists of five sequential enzymes as shown in Figure 1.1. One of the final products, acetoacetate, is a ketone body. With succinyl-CoA, it can be converted into acetyl-CoA, which is a substrate for both the Krebs cycle and fatty acid synthesis. The complete pathway of tyrosine degradation is only expressed in hepatocytes and renal proximal tubular cells (Mitchell, 2001). Four of the five enzymes in the tyrosine catabolic pathway, all except maleylacetoacetate isomerase (MAI), are associated with human genetic diseases.

The primary disease model in this dissertation is hereditary tyrosinemia type I (HT1), in which fumarylacetoacetate hydrolase (FAH) is deficient. As shown in the Figure 1.1 legend, another three out of the four diseases are directly related to the tyrosine catabolism pathway: Type II tyrosinemia, Type III tyrosinemia and Alkaptonuria. Each of these diseases corresponds to one of the five enzymes involved in the tyrosine catabolic pathway (Greelish et al., 1999).

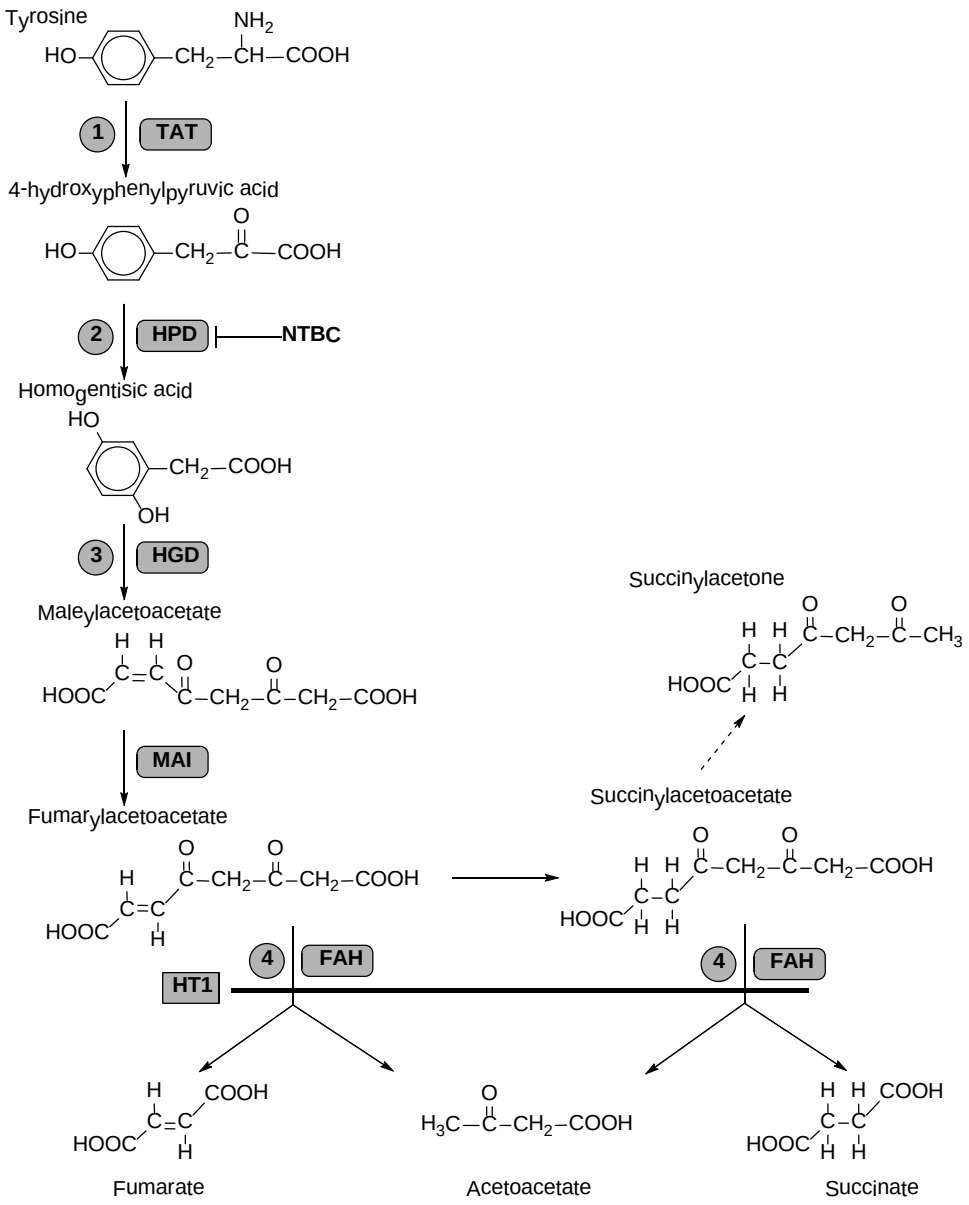


Figure 1.1. The tyrosine catabolism pathway. The diseases related to the enzymes of the pathway are (**1**) Tyrosinemia II, caused by tyrosine aminotransferase (TAT) deficiency; (**2**) Tyrosinemia III, caused by 4-hydroxyphenylpyruvate dioxygenase (HPD) deficiency; (**3**) Alkaptonuria (AKU), caused by homogentisic acid oxidase (HGD) deficiency; (**4**) Tyrosinemia I (HT1), caused by deficiency of fumarylacetoacetate hydrolase (FAH). NTBC, 2-(2-nitro-4-trifluoromethylbenzoyl)- 1,3-cyclohexanedione, is

the pharmacological inhibitor of HPD used in this study.

Hereditary tyrosinemia type I

Type I tyrosinemia or HT I is caused by a deficiency of FAH. HT1 is an autosomal recessive disorder. In humans, HT1 is characterized by severe liver dysfunction, *hepatocellular carcinoma* **(**HCC) and renal tubular damage. The severity of symptoms varies with age of onset. In infants, the phenotype can be very severe and can progress very quickly as an acute form (Grompe, 2001; Mitchell, 2001). In the most chronic form of HT1, symptoms such as hypophosphataemic rickets may not be diagnosed until the patient is a few years old. The onset and severity of the symptoms might be related to diet and genetic background. Importantly, in those who survived the first few years of life, there is a high risk of developing HCC (Weinberg et al., 1976). The molecular causes for liver damage and renal injury, although the exact mechanism remains unclear, are likely through the accumulation of upstream substrates of tyrosine catabolism, such as fumarylacetoacetate (FAA) and/or maleylacetoacetate (MAA), as well as their reduced derivatives succinylacetoacetate (SAA) and succinylacetone (SA), which cause liver and renal damage (Greelish et al., 1999). Further studies show that FAA is primarily responsible for the liver damage in the HT1 patients, because the SA level, but not FAA, is significantly elevated in the urine and plasma of HT1 patients (Greelish et al., 1999). If SA was the primary metabolite causing liver injury, the damage should be more universal; however, this might be also due to hepatocytes being more susceptible to SA, or that SA concentration in Fah deficient hepatocytes is much higher than in urine and blood.

To diagnose HT1, there are several indicators, for example, liver disease, plasma levels of alpha-fetoprotein (AFP) and urine SA(Grompe, 2001), which is also a hallmark

metabolite of HT1. Fah enzyme activity in the liver specimen or skin culture is, of course, an essential diagnostic marker. Other diagnostic markers include: elevated plasma concentrations of tyrosine and methionine, hypoglycemia and abnormal coagulation (Mitchell, 2001).

HT1 is a rare disease with a worldwide incidence of 1/100,000 to 1/120,000 of the population. However, in Quebec, Canada, HT1 is very common. About 1 in 16,786 live births in Quebec is diagnosed as HT1 (De Braekeleer and Larochelle, 1990).

Another tyrosinemia that will be discussed here is hereditary tyrosinema type III (HT III). HT III is caused by a *HPD* mutation, the rarest disorder of tyrosine metabolism. There are very few patients who have been diagnosed with HT III. The symptoms are not clear, except for elevated tyrosine concentration (Mitchell, 2001). These diagnosed HT III patients have no liver dysfunction or renal injury. This might be because accumulation of 4-hydroxyphenylpyruvate is not pathogenic (Cerone et al., 1997).

Treatment for hereditary tyrosinemia type I

HT1 therapy prior to NTBC

Before NTBC therapy was introduced to HT1 patients, dietary restriction of phenylalanine and tyrosine, which are upstream substrates of the tyrosine catabolic pathway, was the only treatment for HT1 patients (Bain et al., 1990). Although the dietary restriction slowed the progression of the disease, the patients still developed HCC (Mitchell, 2001). HT1 remained uncurable until liver transplantation became the standard treatment (Mohan et al., 1999). That HT1 patients can be cured by liver transplantation, but not dietary restriction of tyrosine, also suggests normal FAH function in the liver of

patients will be the necessary for treatment of HT1 patients, which makes gene therapy a candidate for this monogenic disease (Horwich, 1991).

NTBC

NTBC, 2-(2-nitro-4-trifluoromethylbenzoyl)-1, 3-cyclohexanedione, is a triketone and 4-hydroxyphenylpyruvate dioxygenase (HPD) inhibitor (Schulz et al., 1993). NTBC was originally developed as herbicide. Early studies on the toxicity of this compound found that rats treated with NTBC developed tyrosinemia symptoms, such as the high levels of tyrosine seen in human beings. Further investigation showed that NTBC was a potent inhibitor of HPD and could reduce formation of the toxic compounds FAA, MAA, as well as SA, by blocking the catabolic pathway upstream. Lindstedt and co-workers successfully treated five patients with NTBC in 1991 (Lindstedt et al., 1992). Since then, NTBC has become a standard treatment for HT1 patients. In about 95 % of patients (over 300 patients) that received NTBC treatment, liver functions greatly improved. Importantly, NTBC not only restored liver functions, but also greatly improved kidney functions. NTBC has been used clinically for almost 20 years and so far no significant side effects have been reported concerning NTBC at the recommended dosage, typically 1 mg/kg/day (Holme and Lindstedt, 1998; Mitchell, 2001).

So far, most reports have not revealed any significant side effects of NTBC(Mitchell, 2001; Ros et al., 1999). However, NTBC itself is a competent inhibitor of HPD in the tyrosine catabolic pathway. The underlying mechanism of NTBC treatment is, in essence, to convert the severe tyrosinemia type I into the more survivable tyrosinemia type III (HPD deficient) (Holme and Lindstedt, 1998). As NTBC blocks the

tyrosine catabolic pathway, the dietary restriction of tyrosine and phenylalanine is still necessary for these patients (Holme and Lindstedt, 1998). In a report on French patients, overall, although 97.5 % of patients survived, 17 of 45 patients developed persistent abnormal heterogeneous liver and six patients had cirrhosis, as well as 15 patients having consistent elevated levels of alpha-fetoprotein, which indicates the risk of HCC (Masurel-Paulet et al., 2008). Another outcome from the treated patients with NTBC is neurological impacts (Masurel-Paulet et al., 2008).

Fah mouse model

To better understand HT1, many efforts have been made to generate mouse models. So far there are two popular strains of HT1 mouse models: the c^{14CoS} lethal albino mutant mouse and the $Fah^{\Delta exon\ 5}$ knockout mouse. The albino lethal c^{14CoS} mouse is neonatal lethal. The c^{14CoS} mouse was generated by x-ray–induced large deletions of chromosome 7. The homozygous offspring died several hours after birth (Gluecksohn-Waelsch, 1979; Russell et al., 1979). The lethal phenotype is due to abnormal expression of hepatocyte-specific genes in the liver and impaired liver function during the perinatal period (Kelsey et al., 1993; Kelsey and Schutz, 1993; Klebig et al., 1992). The genetic mapping and functional rescue proved that the phenotypes of this mutant mouse can be almost completely attributed to the deficiency of the *Fah* locus in this region (Niswander et al., 1991) (Kelsey et al., 1993).

Another HT1 mouse model, the mouse model used in this dissertation, is the $Fah^{\Delta exon\ 5}$ knockout mouse, which was generated by homologous recombination mediated gene disruption in embryonic stem cells. $Fah^{\Delta exon\ 5}$ knockout mice have exactly the same phenotype as c^{14CoS} mice (Grompe et al., 1993). $Fah^{\Delta exon\ 5}$ knockout mice are lethal unless

treated with NTBC, with which these mice can survive free of liver disease for more than a year (Grompe et al., 1995). A very important feature of NTBC treatment is that it is reversible, which means that it can be withdrawn at any desired time, which is critical in the study of this dissertation.

Advantages of HT1 mice model for the study of AAV integration

It is critical to find a disease model to study AAV integration *in vivo*. A good mouse model needs the following features: 1) positive selection of the integrated AAV genome *in vivo* while eliminating episomal AAV genomes; 2) easy methods to distinguish cells containing AAV genomes from cells without AAV genomes; 3) quantitative measurement of the AAV genomes and the therapeutic effect *in vivo*; 4) survival long enough to see the long term effects of AAV integration. It is very important to evaluate the safety of AAV integration.

The HT1 mouse model mimics the symptoms of HT1 patients with FAH deficiency. In HT1 mice without treatment, the hepatocytes and renal tubular cells die due to the accumulation of FAA and its derivative SA; although the exact mechanisms remain unclear (Kubo et al., 1998; Lindblad et al., 1977; Naylor et al., 1996; Sun et al., 2000). HT1 mice receiving NTBC in drinking water will survive without liver damage caused by FAA and SA accumulation (Grompe et al., 1995).

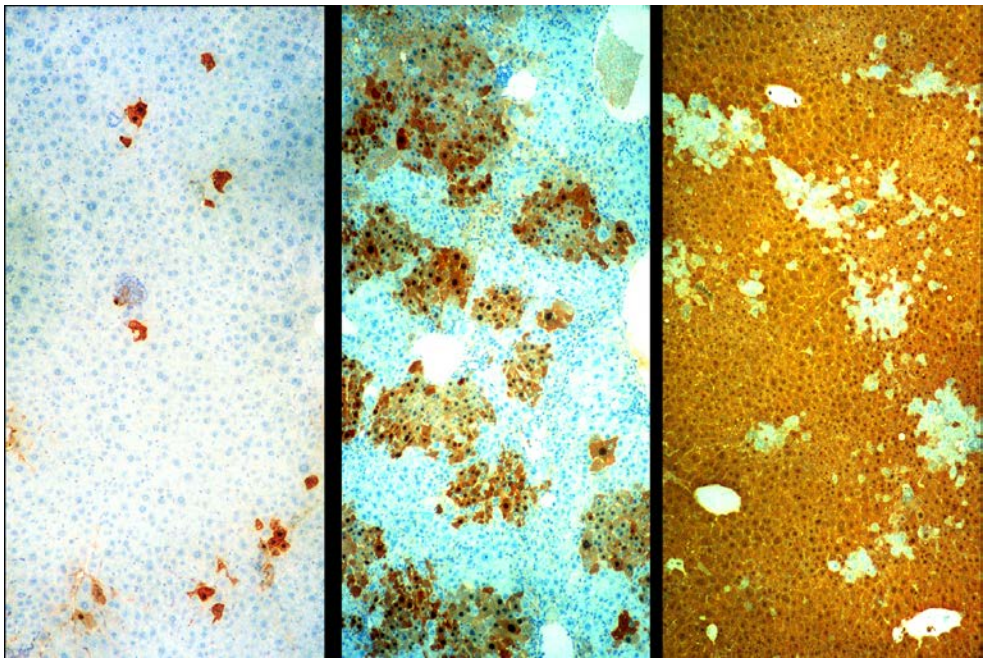


Figure 1.2. Time course of FAH positive hepatocyte mediated repopulation in *Fah*$^{\Delta exon\ 5}$ mouse liver(Overturf et al., 1996). Anti-FAH immunostaining of FAH positive hepatocytes transplanted in a FAH deficient liver. The red/brown colored cells are FAH positive hepatocytes. Left panel: three days after transplantation; middle panel: 21days off NTBC; Right panel: two months off NTBC.

Somatic reversion in HT1

The somatic genetic reversion phenomenon was originally discovered by finding that FAH immunostaining of liver specimens from two HT1 patients were mosaic with FAH positive staining. Also the FAH enzyme activity and FAH Western blotting from skin fibroblasts, but not from liver specimens were negative (Kvittingen et al., 1993). This has been confirmed by PCR and sequencing, which clearly showed wild type and

mutant alleles of FAH gene in the liver samples from these patients, while the skin samples only showed the mutant allele. These finding suggested a selective advantage for FAH positive hepatocytes over the negative hepatocytes since the FAH positive nodules were clonal (Kvittingen et al., 1994; Kvittingen et al., 1993; Mitchell, 2001). The selective advantage of FAH positive hepatocytes was further confirmed in the HT1 mouse model, in which hepatocytes with retrovirus mediated FAH expression repopulated the FAH deficient livers. This hypothesis was also tested by liver repopulation in the HT1 mouse model transplanted with FAH+ hepatocytes (Overturf et al., 1996). The selective process in HT1 mice is shown in Figure 1.2. We can see that FAH positive nodules grow with time over the FAH negative hepatocytes and finally repopulate the whole liver. During the liver repopulation, each transplanted FAH+ hepatocyte will divide several times. For example, 10,000 *Rosa 26* hepatocytes were transplanted into *Fah*$^{-/-}$ mice. After fully repopulating the liver, each FAH positive hepatocyte will have divided at least 10 times (Overturf et al., 1997a). Partial hepatectomy is a routine way to induce liver regeneration and hepatocyte division, in which hepatocytes only divide about once or twice (Kay and Fausto, 1997). Compared with partial hepatectomy in wild type mouse liver, the selective pressure and hepatocyte division potency in HT1 mouse model is significantly higher. Besides this, convenient FAH immunostaining and FAH enzyme assay, serial transplantation and reversible NTBC treatment are other advantages for studying AAV integration in the liver in the HT1 mouse model. Overall the HT1 mouse model is an excellent disease model for AAV integration study.

IV. Hemophilia B and current AAV gene therapy

Hemophilia B

Hemophilia is a group of hereditary genetic disorders with impaired ability to control coagulation (Hedner et al., 2000; Lozier and High, 1990). Hemophilia A (factor VIII deficiency) is the most common form of hemophilia. Its frequency is about 1 in 5,000–10,000 male births (Lozier and High, 1990). Hemophilia B (factor IX deficiency) is less common, in around 1 in about 20,000–34,000 male births (Lozier JN, 2005). Hemophilia B patients are classified by circulating levels of F.IX. Mild, moderate and severe hemophilia respectively refer to plasma concentrations of F.IX of 5–25 %, 1–5 % and <1 % of the normal level (Hasbrouck and High, 2008; Lozier JN, 2005).

Hemophilia B gene therapy

Nowadays, hemophilia B is treated with intravenous infusion of F.IX. Plasma-derived or recombinant F.IX is readily available at considerable cost in many developed countries (Ponder, 2006). For gene therapy in hemophilia B disease, gene addition is usually the method used to achieve therapeutic effect, which involves AAV, retrovirus or lentivirus mediated F.IX expression in the cells in the body and supplied secreted functional proteins into the blood. However, the short half-life of F.IX in the circulation and the costs, as well as the inconvenience of intravenous infusion, have driven gene therapy for hemophilia B (High, 2002; High, 2003).

It has been almost 20 years since the first gene therapy for hemophilia B in dogs was reported (Kay et al., 1993). In the last two decades, many vectors have been applied to mediate F.IX expression in the host to maintain normal coagulation function. In

clinical or preclinical studies, retroviral, lentiviral, adenoviral as well as AAV vectors are commonly used viral vectors (Ishiwata et al., 2009). Retroviral vectors and lentiviral vectors are capable of host chromosomal integration, which can provide sustained expression of functional F.IX, however their integration ability and pathogenicity also raise the possibility of insertional mutagenesis (Hacein-Bey-Abina et al., 2003a). Retroviral vectors only infect dividing cell, which limits their application (Ponder, 2006). Toxicity and strong host immune responses to adenoviral vectors at high doses impair their clinical applications (Hartman et al., 2007; Urosevic et al., 2007). AAV is the most promising viral vector currently used in hemophilia B gene therapy.

The primary and most effective target tissues for AAV gene therapy in Hemophilia B are liver and muscle. AAV-mediated F.IX expression targeting both tissues has been in clinical trials (Manno et al., 2003; Manno et al., 2006). However, the muscle-directed AAV vector-mediated gene therapy trial for hemophilia B did not yield F.IX at a therapeutic level and enrollment has been discontinued (Manno et al., 2003). The liver directed AAV2 mediated F.IX expression encountered two issues: neutralizing antibodies existing in patients and transient expression of F.IX (Manno et al., 2006). Both issues have been partially solved in another clinical trial using AAV8-mediated F.IX gene expression, in which severe hemophilia symptoms have been converted into mild symptoms (Nathwani et al., 2009; Nathwani et al., 2011). These exciting results support development of long lasting AAV vectors, which will provide a better therapeutic effect.

V. Thesis rationale

An integrating AAV with persistent transgene expression has been of central interest ever since AAV was initially applied to gene therapy (McCarty et al., 2004). For genetic liver disease, an integrating AAV vector will provide long lasting transgene expression; while regular AAV mediated transgene expression is transient. Therefore this dissertation will focus on engineering and characterizing a new integrating AAV vector.

Successful homologous sequence mediated gene targeting *in vivo* makes AAV mediated site-specific integration/targeting realistic (Miller et al., 2006; Paulk et al., 2010). One of the known hotspots of rAAV integration is within the ribosomal DNA repeats, which could serve as a potential target for directing site-specific integration. This locus is actively transcribed with more than 200 copies present in a mouse haploid genome (Nakai et al., 2003; Nakai et al., 2005b). These make ribosomal DNA repeats ideal as site-specific integration targets. In addition there is an endonuclease, I-*Ppo*I, which specifically recognizes a 15 bp sequence inside the rDNA gene and cleaves ribosomal DNA to generate a double-strand break (Lin and Vogt, 1998; Monnat et al., 1999). The double-strand break will significantly increase the homologous recombination frequency (Donoho et al., 1998; Smih et al., 1995).

Positive selection in the HT1 mouse model makes selection of hepatocytes with integrated AAV vectors, but not episomal AAV vectors, feasible (Nakai et al., 2003). Based on these, design and study of new AAV-rDNA vectors with integration capabilities in the HT1 mouse model desirable, reasonable and achievable.

Chapter 2

AAV vectors containing rDNA homology display increased chromosomal integration and transgene persistence

Zhongya Wang1, Lescek Lisowski2, Milton J. Finegold3, Hiroyuki Nakai4, Mark A. Kay2 and Markus Grompe1*

As of May 14, 2012, under review at Molecular Therapy

I **Abstract**

Although rAAV vectors are promising tools for gene therapy of genetic disorders, they remain mostly episomal and hence are lost during cell replication. For this reason it is desirable to construct AAV vectors capable of chromosomal integration. Ribosomal DNA repeat sequences are overrepresented during random integration of rAAV. We therefore sought to enhance AAV integration frequency by including 28S rDNA homology arms into our vector design. A vector containing ~1kb of homology on each side of a cDNA expression cassette for human fumarylacetoacetate hydrolase (FAH) was constructed. rAAV of serotypes 2 and 8 were injected into Fah-deficient mice, a model for human tyrosinemia type 1. Integrated FAH transgenes are positively selected in this model. We found that rDNA-containing AAV vectors had a ~30x higher integration frequency than controls. Site-specific integration into the 28S rDNA vector was seen in multiple tissues. Furthermore, rDNA-containing AAV vectors for human factor IX demonstrated increased transgene persistence after liver regeneration. We conclude that rDNA containing AAV vectors may be superior to conventional vector design for the treatment of genetic diseases.

II Introduction

Recombinant adeno-associated viral (rAAV) vector currently is one of the most promising gene therapy vectors. It has been used with success in many animal models, but recently also in human clinical trials (Mingozzi and High, 2011). Treatment of Leber's congenital amaurosis by intraocular injection of serotype 2 rAAV expressing the RPE65 gene resulted in clinical improvement in multiple studies (Maguire et al., 2008; Simonelli et al., 2009). More recently, hepatic gene transfer with AAV serotype 8 was successfully used in hemophilia B (Factor IX deficiency) (Nathwani et al., 2011). Generally the safety profile of these vectors has been excellent (High and Aubourg, 2011).

For genetic liver diseases, long-term AAV-mediated transgene expression at therapeutic levels would be highly desirable. However, in animal models the vast majority of rAAV genomes in liver remain episomal and less than 5-10% of the expression persists after partial hepatectomy (Nakai et al., 2001). In healthy humans, hepatocytes are known to divide every 200-300 days (Macdonald, 1961) and this rate is predicted to be significantly higher with liver injury such as viral hepatitis. Many common infections are associated with mild elevations in serum transaminases (Reisler et al., 1967) suggesting that the therapeutic benefit of non-integrated rAAV will be transient.

AAV2 vectors can be pseudotyped with capsids from diverse AAV-types and many different serotypes have been developed (Gao et al., 2005; Simonelli et al., 2009; Vandenberghe et al., 2009). Much effort has been invested into the modification and evolution of AAV capsids to increase transduction efficiency of different tissues (Girod

et al., 1999; Wu et al., 2000; Wu et al., 2006a). To date, however, limited progress has been made in terms of developing rAAV vectors capable of more efficient chromosomal integration.

Large-scale analysis of AAV integration sites in hepatocytes has been performed to understand the mechanisms and patterns of integration (Miller et al., 2005; Nakai et al., 2005b). Studies in the mouse model of HT1 showed that about 60% of rAAV vector integrations occurred in active genes (Nakai et al., 2003). In addition, integration 'hot spots' were found and one of the surprising finding was that 3 to 8% of integrated vectors resided in ribosomal DNA repeats. Even though this sequence has ~200 copies/ diploid genome this frequency corresponds to a strong overrepresentation in terms of integration specificity. Given the preference of unmodified rAAV vectors for integration into rDNA and knowing the uncoating of the virus occurs in the nucleolus (Johnson and Samulski, 2009; Miller et al., 2005; Nakai et al., 2005b) we wished to ask if introduction of rDNA homology into rAAV vectors would increase integration frequency, especially into rDNA repeats, by homologous recombination or non homologous end joining.

Here we report the application of an rDNA homology-containing vector for *in vivo* gene therapy focusing on a model system which permits selection of hepatic transgenes as well as hemophilia B disease model.

III **Materials and Methods**

AAV vector construction, virus preparation and injection

A 2,038 bp fragment of the 28 S ribosomal DNA locus of the mouse was amplified from wild-type129s4 strain mouse genomic DNA with the following primers: forward: 5'-*GCGGCCGC*AAAGGGAGTCGGGTTCAGAT-3', reverse: 5'-*GCGGCCGC*TCCTC AGCCAAGCACATACA-3'. This fragment was cloned into the blunted *Not*I sites of a pAAV2 vector (Nakai et al., 2003) generating the plasmid pAAV-rDNA. A human *FAH* expression cassette containing a Rous Sarcoma Virus (RSV) promoter, *FAH* cDNA and bovine growth hormone polyadenylation sequence A was amplified from plasmid pRSV-FAH (Overturf et al., 1997b) (forward primer: CTTAAGGCTGCTTCGCGATG TACG, reverse: CTTAAGCCATAGAGCCCACC GCATC) and cloned into a unique *Afl* II site centered in the middle of pAAV-rDNA, thus generating pAAV-rDNA-FAH. The overall insert length was 2,250 bp. To make a control vector of comparable size, pAAV-stuffer-FAH was constructed. A 1,968 bp fragment of mouse *Fah* gene intron 3 was subcloned to pAAV plasmid *Not* I sites after amplification of with the following primers, forward:5'-TTGCGGCCGCATTCCCAC CCACTCTCCTCT-3', reverse: GCGGCCGCGGGGTTGGATTCAATGTTTG, generating pAAV-stuffer. Next, the RSV-FAH-PolyA expression cassette was subcloned into the *Afl* II site of pAAV-stuffer. To construct pAAV-rDNA-hF.IX, a 2,280 fragment of a human Factor IX (*FIX*) expression cassette was amplified from pNEB193-SynEnh-TTR-hF.IX10-90-spa (courtesy of Mark Kay at Stanford University) and sub-cloned into the *Afl*II site of pAAV-rDNA. These plasmids were used to generate rAAV vectors of capsid serotypes 2 and 8 (Gao et al., 2002; Nakai et al., 2003) by transient transfection of HEK293 using

published technology (Choi et al., 2007a, b). In adult animals all virus injections were done intravenously by retro-orbital administration in 200 µl of sterile saline at the doses indicated. For neonatal animals facial vein injections were done with 30 µl.

Animal husbandry and HT1 mouse model

All animal experiments were performed according to the guidelines for animal care at Oregon Health & Science University. Fah deficient mice were of the $Fah^{\Delta exon5}$ strain on the 129s4 background as previously described (Grompe et al., 1993). $Fah^{-/-}$ mice were maintained with drinking water containing 2-(2-nitro-4-trifluoro-methylbenzoyl) -1,3-cyclohexanedione (NTBC; Swedish Orphan AB) at a concentration of 4 mg/l (Held et al., 2005). *Rosa*26 transgenic mice were developed originally by Dr. Philippe Soriano and maintained on 129S4 background (Friedrich and Soriano, 1991).

Hepatocyte transplantation and in vivo selection

For *in vivo* selection of FAH expressing hepatocytes with integrated rAAV2 vector genomes, we withdrew NTBC from the drinking water as previously described (Held et al., 2005).

To further select for hepatocytes containing integrated vector genomes and to dilute non-integrated vector genomes, serial transplantation was performed (Overturf et al., 1997a). Primary recipients of AAV injections were subjected to NTBC withdrawal and positive selection of Fah-expressing hepatocytes. After their weight and health had stabilized, the livers of primary animals were perfused and hepatocytes were isolated for serial transplantation according to a previously described protocol (Overturf et al., 1997a). For transplantation, 300,000 hepatocytes from AAV infused primary liver were

intrasplenically injected into secondary *Fah*$^{-/-}$ recipient mice. NTBC was discontinued one week after transplant to permit selection of donor hepatocytes.

Histology and immunohistology

AAV infected or transplanted *Fah*$^{-/-}$ mice were sacrificed and specimens were fixed in *10*% buffered *formalin*, sectioned, and immuno-stained by custom rabbit polyclonal FAH antibody diluted 1:1000 and detected with 10 ng/ml goat anti-rabbit secondary antibody (Santa Cruz Biotechnology, Santa Cruz, CA.) as previously described (Held et al., 2005).

Southern blot analysis

Genomic DNA was isolated from 350 mg of liver tissue from AAV infected primary or secondary mice using the Qiagen genomic DNA isolation kit (Qiagen, Valencia, CA, USA). 10 micrograms of genomic DNA was digested with the restriction enzyme *EcoR* I (NEB, Ipswich, MA) and subjected to electrophoresis in a 0.9% agarose gel overnight. Capillary transfer and hybridizations were performed according to standard protocols (Grompe et al., 1991). A 1.5 kb fragment containing FAH cDNA from plasmid pRSV-FAH (Overturf et al., 1997b) was digested with *Eco*RI and gel purified with a Qiagen gel purification kit (Qiagen, Valencia, CA, USA). This fragment was radioactively labeled with the Random Primed DNA Labeling Kit (Roche Applied Science, Indianapolis, IN) with α-32p dCTP (PerkinElmer, Waltham, MA) and used as the probe. Hybridizations were conducted in 0.5 M sodium phosphate pH 7.2, 1 mM EDTA, 7% SDS containing 2 x 10^6 cpm probe/ml and 100 µg of herring sperm DNA/ml at 65°C for 36 h. Washing was performed with 40 mM sodium phosphate pH 7.2, 1%

SDS, once at room temperature for 15 min and twice at 60°C for 40 min. Signal was visualized and quantitated by a PhosphorImager (GE Healthcare Life Sciences, Piscataway, NJ).

Quantitative real-time PCR

Quantitative real-time PCR was used to measure copy numbers of the human *FAH* and human Factor IX transgenes and the donor marker *Rosa26*. This was done with the Bio-Rad MyiQ single-color real-time PCR detection system using SYBR green dye. The single copy mouse gene *Maai* **(**maleylacetoacetate isomerase**)** was used for normalization**.** The following primer pairs were used*:*

FAH:5'CCCGTATATGGTGCCTGCAA3' and 5'TCGTGCACTCCAGTCGTTCA3', *Rosa26*:5'CATCAGCCGCTACAGTCAACAG3'and 5'CAGCCATGTGCCTTCTTCC GC3'; *Maai*:5'GTACCATTGAGGTGGGCCTA3' and 5'GCTGGTTCGTCCACTCTT TC3', HF.IX:5'AAGATGCCAAACCAGGTCAATT3' and5' GATAGAGCCTCCACA GAATGCA3'. The real-time PCR program parameters were: 95^{o}C x5'; (95^{o}C x15'', 62^{o}C x20'', 72^{o}C x20'') x45; 45^{o}C x10"x100.

PCR for the detection of site-specific 28S rDNA integration

Junction PCR program (touch-down PCR):

95^{o}Cx5'; (95^{o}C x15'', 65^{o}C >55^{o}C x20'', 72^{o}C x2'); (95^{o}C x15'', 55^{o}C x20'', 72^{o}C x2') x35; 72^{o}C x20'.

The following primer pairs were used: In each case one primer was located within the *FAH* expression cassette and one was located within 28 S rDNA sequence, but not within the homology sequence contained within the rDNA vector (see Fig.2.1A).

For J-PCR1, forward primer: 5'-CCGCTTTTCGCCTAAACACAC-3', reverse primer: 5'-CCCAGAAGCAGCAGGTCGTCTCTACG-3'; For J-PCR2, forward primer: 5'-GCCGTATCGTTCCGCCTG-3',reverse primer: 5'-GTATATCTGGCCCGTACAT-3'; For J-PCR3, forward primer: 5'-GTATATCTGGCCCGTACAT-3', reverse primer: 5'-GCTCCTCAGCCAAGCACA-3'; For J-PCR4, forward primer: 5'-CACCCCGTT TCCCAAGACG-3', reverse primer: 5'-AACACACCCT AGTCCCCT-3';

To increase the sensitivity of detecting junction fragments a nested PCR assay was developed. The nested PCR was same as the junction PCR described above, but a first round PCR was performed with outside primers, forward: 5'-CCGCTTTTCGCCTA AACACAC-3', reverse primer: 5'-CCCAGAAGCAGCAGGTCGTCTCTACG-3'. 1 µl from the first 50 µl PCR reaction was used as a template for the second round PCR. The second round was performed as in J-PCR4.

Human Factor IX analysis

Mouse blood was collected by *retro-orbital* bleeding into heparin tubes and an ELISA assay was employed to analyze for total human factor IX antigen as described previously (Nakai et al., 2001). The specific activity of hF.IX in a mouse plasma background may be slightly varied because of unknown interactions of the human protein with mouse plasma proteins.

Quantitation of site-specific integration by serial dilution junction PCR

Liver genomic DNA from serially transplanted *Fah*$^{-/-}$ mice was diluted sequentially from 1µg to 0.01 ng. 30 separate nested junction PCRs were performed for each DNA amount to identify a concentration at which only a percentage of all reactions was

positive. To calculate the integration frequency we used a Poisson distribution analysis (Brisco et al., 2010; Fuller et al., 2001). $P0 = e^{-x}$ where P0 is the percentage of samples (PCR reactions) which were positive at a given DNA concentration and x represents the frequency of the event analyzed (site-specific integration).

This formula rearranges to: $x = \ln (1/ P0)$, in our study, the percentage of positive site-specific PCRs was 16 out of 32 using 0.3ng genomic liver DNA as template per reaction. $x=\ln (1/(16/32))=\ln 2=0.693$. Therefore, on average 0.693 integrations were present in 0.3 ng genomic DNA. Each diploid mouse genome contains approximately 5.7 pg genomic DNA, which means 0.3 ng genomic DNA has 52.6 diploid genomes. Hence, the site-specific integration frequency in our sample can be estimated at: 0.693/52.6=1.32%

Integration sites rescue

Shuttle vector AAV-rDNA-I-SceI-KO was constructed based pAAV-rDNA and pNKan as described in supplementary methods. AAV8-rDNA-I-SceI-KO was produced and purified as described before (Nakai et al., 2003). We injected 12 adult male wild type mice via the retro orbital route with AAV8-rDNA-I-SceI-KO at a dose of 3.0×10^{11} vector genomes per mouse. After one month, these mice were sacrificed and livers were immediately frozen in liquid nitrogen and genomic DNA was made from these frozen livers as described in above. AAV8-rDNA-I-SceI-KO integration library was constructed as described before (Nakai et al., 2003). Library screening was based on the strategy in supplementary Fig. S3. The screened clones were sequenced and the detail of bioinformatics process was done as described before (Nakai et al., 2005b).

Measurement of absolute integration frequency

The integration frequency of Fah-expressing AAV vectors was determined by counting FAH+ nodules stained by immunohistochemistry (Wang et al., 2010). NTBC was withdrawn for only 25 days in order to prevent repopulation nodules from becoming too large and confluent for accurate counting. The nodule size seen in 2D sections was used to calculate a correction factor for the estimation of the number of clonal nodules as previously described (Wang et al., 2002). The total number of hepatocytes in a section was determined by precisely measuring the surface area of each section. 1 mm^2 of a mouse liver section contains about 1,860 hepatocytes (Wang et al., 2002). Surface area was measured by scanning the slides along with a size standard using a Canon scanner at a resolution of 300 dpi. The pixels of each liver section were counted using Adobe PhotoShop 7 software and converted into mm^2 based on the size standard.

IV Results

AAV-rDNA-FAH results in significantly higher integration frequency

In order to increase the integration frequency of rAAV into the genome, a novel vector genome containing DNA from the 28S ribosomal repeat DNA gene was designed with the purpose of achieving site-specific recombination. This specific sequence was chosen because rAAV has a propensity to integrate into rDNA in hepatocytes (Nakai et al., 2003; Nakai et al., 2005b) and because it is repeated in hundreds of very similar copies throughout the genome. Furthermore, the region chosen contains the recognition site for a rare endonuclease, I-*Ppo*I, providing the possibility of enhancing recombination by creating site-specific double-strand breaks (Argast et al., 1998; Monnat et al., 1999). A FAH expression cassette was cloned into the middle of a ~2.2 kb fragment of 28S rDNA generating homology arms of ~ 1kb on either side (Fig.1A) and then inserted into an AAV2 genome.

This vector (rDNA-FAH AAV) was packaged into AAV8 and AAV2 capsids and injected into *Fah*$^{-/-}$ mice of different ages. As in our previous experiments using conventional vector designs (Nakai et al., 2003), we used *Fah* knockout mice, because this model permits *in vivo* selection of Fah-expressing hepatocytes. Episomal transgenes are lost because of injury induced hepatocyte turnover and only cells with integrated transgenes can survive (Nakai et al., 2003; Nakai et al., 2001). In order to exclude simple size differences as the reason for different integration frequencies, a control vector was generated in which the identical FAH-expression cassette was flanked by single-copy intronic sequence of the same size. Thus both the rDNA-FAH and the control stuffer-FAH vectors were about 4.3 kb in total length.

To determine whether the rDNA-FAH vector yielded a higher integration frequency, we first established the minimal vector dose required to rescue the lethal liver failure phenotype of Fah knockout mice. One of the important features of the Fah-knockout model is that complete liver repopulation and phenotypic rescue of treated animals can be achieved even if the frequency of Fah+ hepatocytes is very low (Overturf et al., 1996). To obtain clinical correction multiple cycles of 2-(2-nitro-4-trifluoro-methylbenzoyl)-1,3-cyclohexanedione (NTBC) withdrawal and re-administration are used. The animals initially lose weight but recover from liver failure while back on NTBC. In each withdrawal cycle the Fah+ nodules grow in size until a sufficient mass of healthy hepatocytes is reached and liver function is fully restored. The number of NTBC cycles required for animals to achieve normal liver function can be used to estimate the frequency of functional hepatocyte nodules. AAV8-rDNA-FAH and AAV8-stuffer-FAH were injected into five two months old *Fah*$^{-/-}$ mice each at 4 different doses (20 mice total for each vector), ranging from $3x10^{10}$ to $1x10^{9}$ vg/mouse. After injection, NTBC was stopped to induce liver failure and to permit selection of Fah^{+}-hepatocytes. If drug withdrawal produced weight loss exceeding 20% of the starting weight NTBC therapy was reinstituted to restore health as schematically shown in Fig.2.1B.

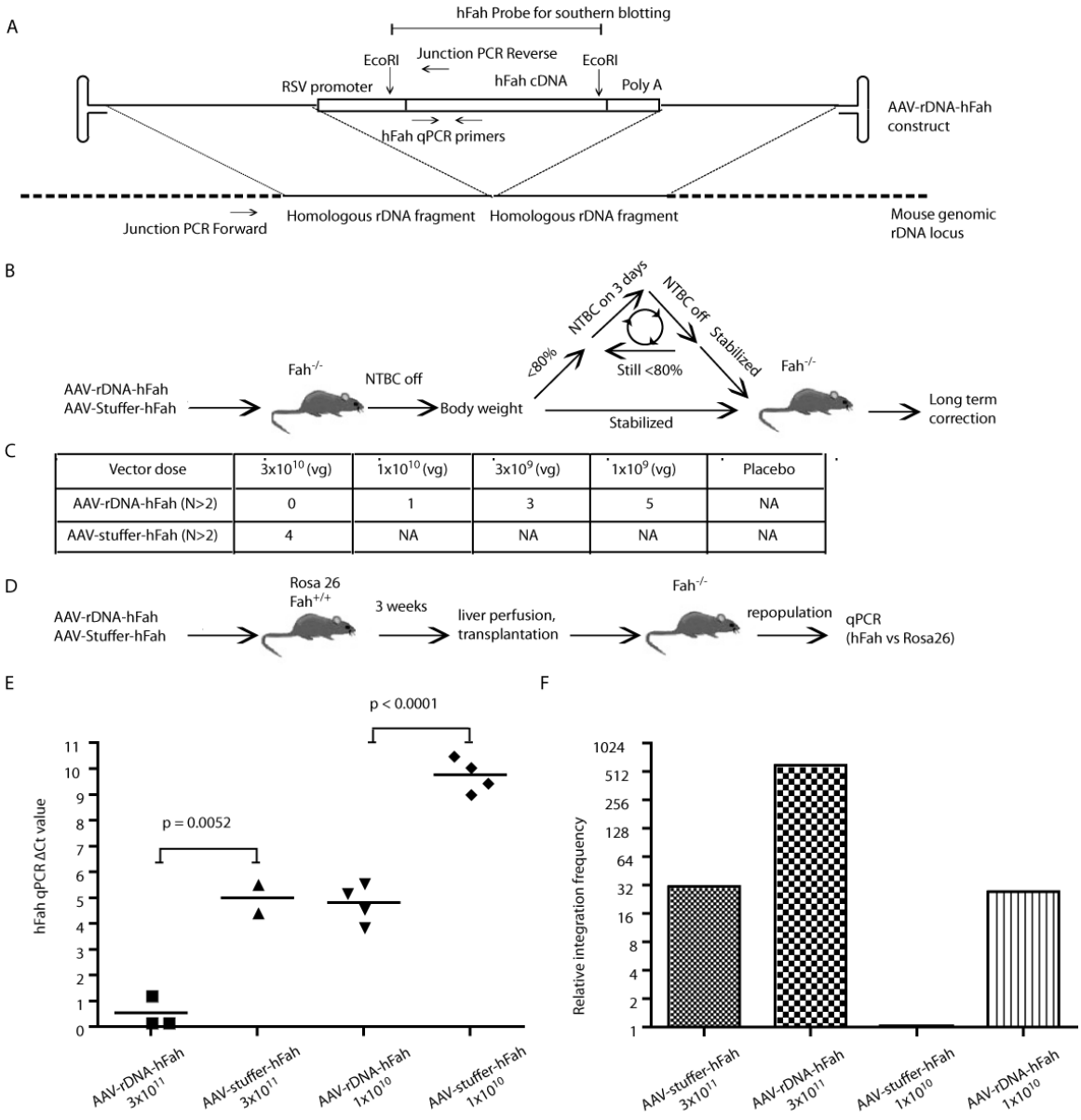


Figure 2.1. AAV-rDNA-FAH yields about 30 times higher integration frequency.

(A) AAV-rDNA-FAH structure and homologous 28S rDNA integration. **(B)** Schematic of Fah-/- liver function rescue experiment. **(C)** Comparison of the dose responses of AAV8-rDNA-FAH and AAV-stuffer-FAH. The NTBC cycle number needed to rescue the Fah-/- mice is given. At each virus genome (vg) dose AAV-rDNA-FAH needed fewer NTBC cycles to rescue liver function. **(D)** Experimental design for the comparison of AAV-rDNA-FAH and AAV-stuffer-FAH without transgene selection. **(E)** Copy number determination by qPCR. The ΔCt for *FAH* vs. *Rosa26*, (donor hepatocyte marker) is shown for AAV-rDNA-FAH and AAV-stuffer-FAH. All *FAH* and *Rosa26* PCR Ct

values were normalized to the single copy gene *Maai*. **(F)** Relative total integration frequencies at the indicated dose. All integration frequencies were normalized to that of AAV-stuffer-FAH at $1x10^{10}$ (3rd bar), which was arbitrarily set to 1.

Interestingly, the rDNA-FAH and stuffer-FAH cohorts displayed markedly different clinical dose responses. As shown in Fig.2.1C, AAV-rDNA-FAH directly rescued *Fah*$^{-/-}$ mice at a dose of $3x10^{10}$ vg in the very first cycle. None of the treated animals lost more than 20% of body weight and NTBC did not have to be restarted. In stark contrast, animals injected with the same dose of AAV-stuffer-FAH vector required 4 cycles of NTBC withdrawal before they stabilized. At lower doses ($1x10^{10}$ or lower), the stuffer vector was not able to rescue *Fah*$^{-/-}$ mice at all, whereas the AAV-rDNA-FAH vector produced permanently corrected mice at all doses tested. Based on the number of NTBC cycles required it could be estimated that the clinical rescue dose for the rDNA vector was about 10-30-times lower than for the control vector, lacking rDNA homology. Similar data were obtained with AAV2 (data not shown).

Next, we tested whether the more effective dose response of the rDNA vector was due to a higher integration frequency. Since the improved functional rescue obtained with the rDNA-vector could also be the result of better Fah-expression rather than higher integration, we sought to uncouple the measurement of integration frequency from the levels of Fah-expression. This was achieved by injecting the AAV-vectors into *Rosa26* mice, which have normal Fah-activity, followed by serial transplantation into *Fah*$^{-/-}$ mice (Fig.2.1D). In this experiment, every hepatocyte from the AAV-injected primary animal expresses Fah and can therefore be selected in the recipient regardless of whether AAV has integrated. During the expansion of the donor hepatocytes episomal AAV-vectors are

lost, but integrated AAV-genomes are retained, regardless of whether they result in Fah-expression or not. Two doses ($3x10^{11}$ and $1x10^{10}$ vg) of rAAV8-rDNA-FAH and rAAV8-stuffer-FAH were injected into 6-week-old *Rosa26* mice (n=4/each dose). Injected mice were kept for 3 weeks to allow for AAV vector integration and then their hepatocytes were isolated and transplanted into secondary *Fah*$^{-/-}$ recipients (n=4/each dose). After repopulation of the secondary recipients by donor hepatocytes these mice were sacrificed and liver genomic DNA was extracted. Quantitative real time PCR was performed to measure the copy numbers of human *FAH* cDNA (AAV vector), the *Rosa26* gene (marker for donor hepatocytes) and mouse *Maai* gene (a single copy gene, common to both donor and recipient as internal control). If the AAV-rDNA-FAH had a higher integration frequency than AAV-stuffer-FAH, the copy number of human *FAH* cDNA would be expected to be higher when normalized to the donor input (*Rosa26* gene). Indeed, as seen in Fig.2.1E, the copy number of *FAH* was ~ 21x higher for AAV-rDNA-FAH than the control vector at the lower dose and ~31x higher at the higher dose. These figures fit the clinical dose response very well and suggested that the superior performance.

AAV-rDNA-FAH was due to a higher integration frequency

To further confirm this finding, we also measured integration by determining the frequency of Fah^{+} nodules, similar to experiments done previously for sleeping beauty transposons (Held et al., 2005; Montini et al., 2002). AAV-rDNA-FAH and control AAV-stuffer-FAH were injected into 2-month-old *Fah*$^{-/-}$ mice at a dose of $3x10^{10}$ vg/mouse (Fig.2.2A). NTBC was stopped to permit positive selection of Fah^{+}-hepatocyte nodules and mice were sacrificed 3 weeks after injection. Fah

immunohistochemistry was done on multiple liver sections and nodule size and number were ascertained. The average nodule frequency for AAV-rDNA-FAH was 1/1,080 (0.092%) and 1/20,000 (0.005%) for the control (Fig.2.2C), once again confirming the significantly higher integration frequency of the rDNA vector.

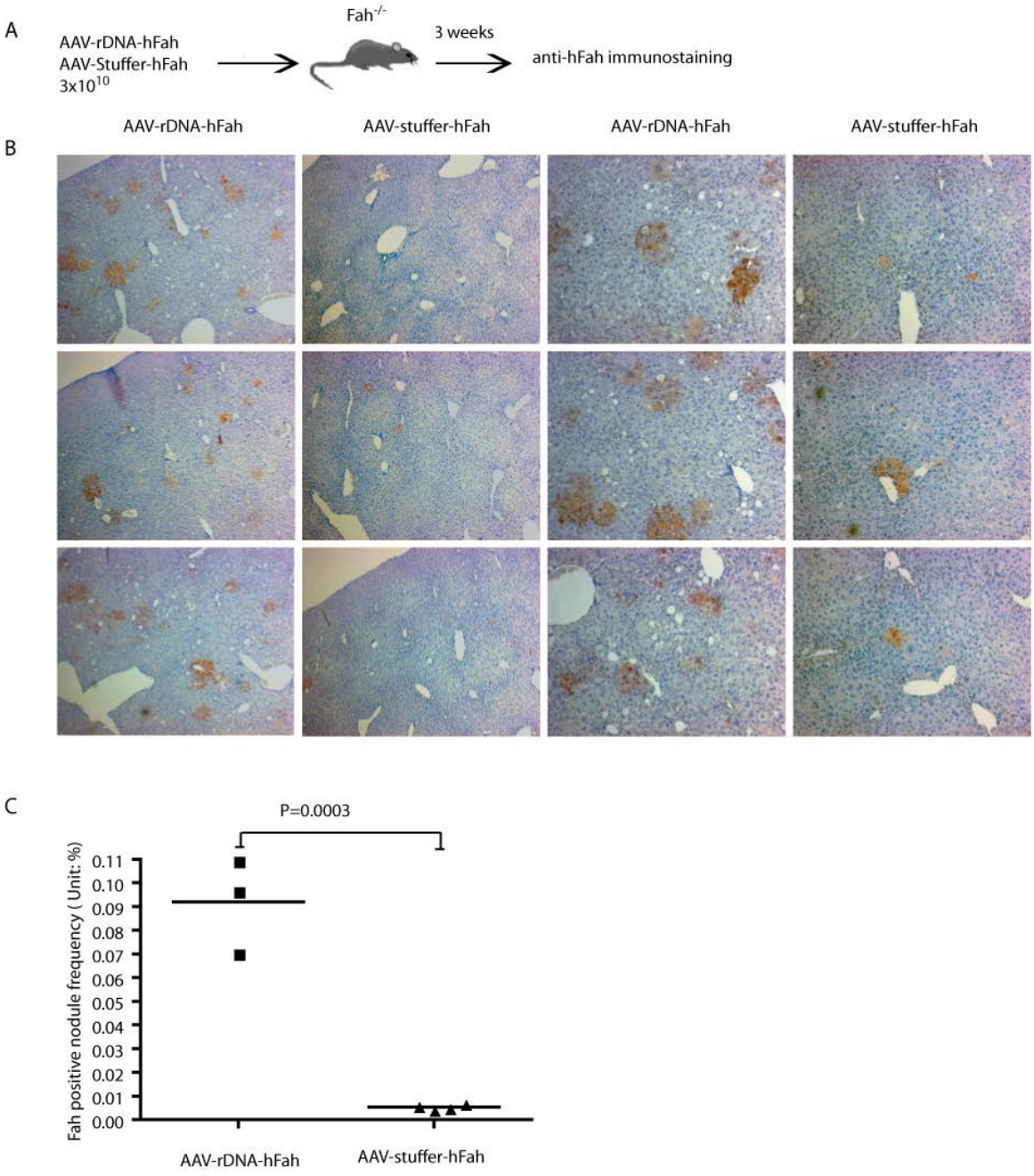


Figure 2.2. Assessment of integration frequency by nodule analysis.

(A) Experimental design. **(B)** Representative FAH staining (brown color) of liver specimens after AAV infusion after 25 days of selection. Left two panels: 50x. Right two panels: 200X. The increased nodule frequency obtained with the rDNA vector is clear. **(C)** Quantitative analysis of nodule frequency.

AAV-rDNA-FAH can integrate specifically into rDNA

It has been shown previously that ribosomal DNA is one of the hot spots for all rAAV vector integration (Miller et al., 2005; Nakai et al., 2005b). We therefore wanted to determine whether the inclusion of rDNA homology into the vector genome would enhance integration into this region by site-specific homologous recombination.

The presence of 28S rDNA site-specific integration events was ascertained by performing PCR with primers specifically designed to amplify junction fragments unique to such events. Three independent primer sets were developed in order to ensure maximum sensitivity and specificity of the assay. As a control for template switching during PCR, we mixed circular vector plasmid with mouse genomic DNA and demonstrated that no junction-specific PCR products could be amplified from such samples (Fig. 2.4). $1x10^{10}$vg of AAV8-rDNA-FAH were injected into 4 neonatal *Fah*$^{-/-}$ mice by facial vein injection and NTBC was stopped immediately. Interestingly, all injected mice showed complete phenotypic rescue and became NTBC independent. Given the high turnover of liver cells in neonates and the associated loss of episomes, this result again supports a high frequency of integration by the rDNA vector. The junction PCR was positive in the liver of all injected animals, indicating that site-specific integration indeed occurred.

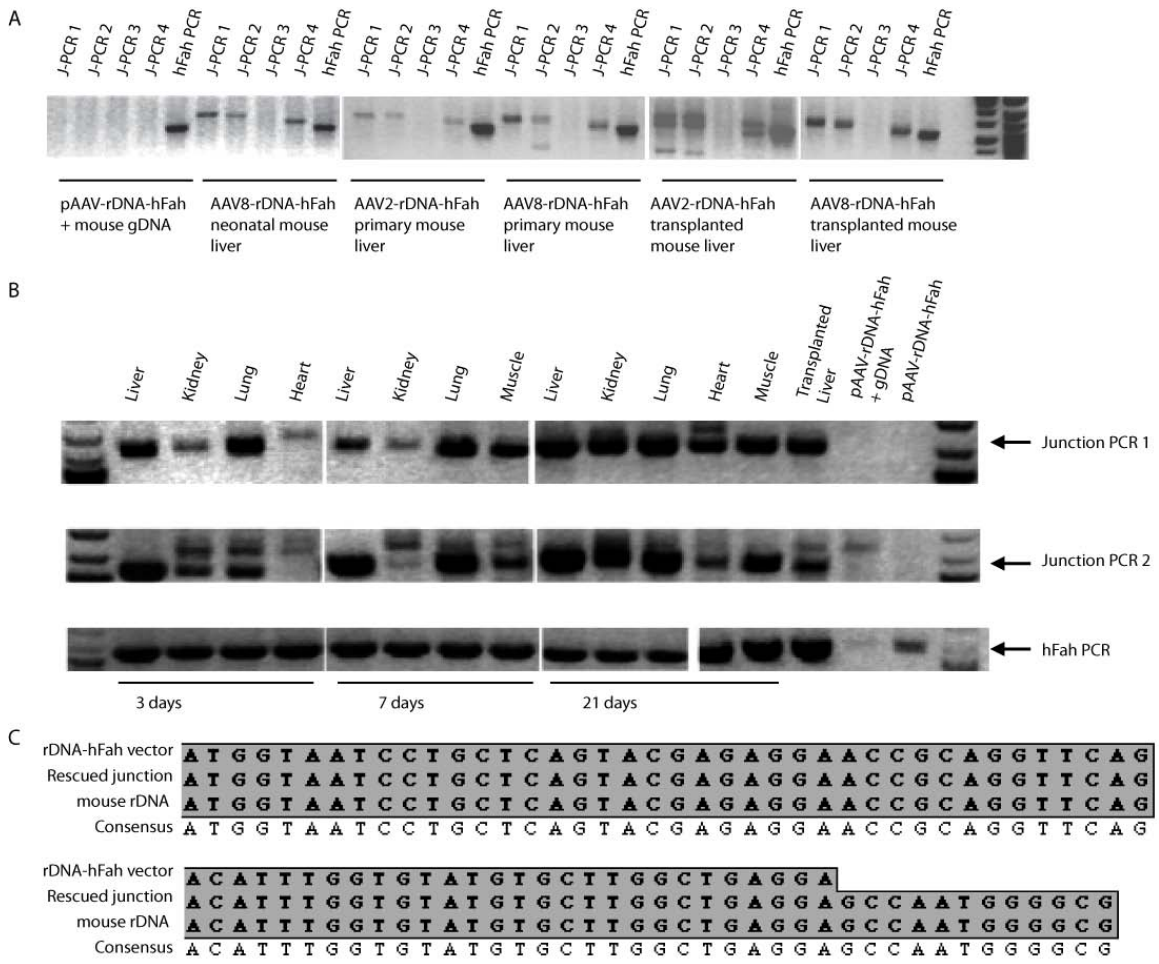


Figure 2.3: Site-specific integration of AAV-rDNA-FAH.

(A) AAV-rDNA-FAH mediated site-specific rDNA integration was detected by different sets of junction PCRs (J-PCR1, J-PCR2, J-PCR3, J-PCR4) in primary infected and secondary transplanted liver as well as neonatal mouse liver. One primer was located within the FAH transgene and one within rDNA adjacent to the homology (see Fig.2.1A). A PCR for the human FAH transgene (FAH PCR) served as positive control for presence of the vector. J-PCRs 1, 2 and 4 were positive in multiple samples of AAV-rDNA-FAH injected liver. The far left panel shows that junction PCRs were negative, when genomic DNA was mixed with unintegrated pAAV-rDNA-FAH plasmid. **(B)** Site-specific

integration was detected in liver, kidney, lung, heart and muscle, as early as 3 days post injection. **(C)** Alignment of rescued junction fragment sequences with genomic rDNA sequence.

To better understand the kinetics of rDNA mediated integration, two 2-month-old AAV-rDNA-FAH mice injected with different doses of vector were harvested at 3, 7 and 21 days after injection and multiple tissues were analyzed by a junction PCR assay. Site-specific rDNA locus integration was detected as early as day 3 in liver, kidney, and lung, but not in heart (Fig.2.3B). At 7 days site-specific integration was seen in skeletal muscle and by 21 days the heart became positive. These results demonstrate that rDNA vectors can integrate site-specifically in multiple different tissues, but have different temporal kinetics with muscle tissues being delayed in comparison to solid organs. To confirm the specificity of the PCR reaction, junction PCR products were cloned and sequenced (Fig.2.3C and Figure 2.4B), showing perfect alignment with genomic 28 S rDNA sequences.

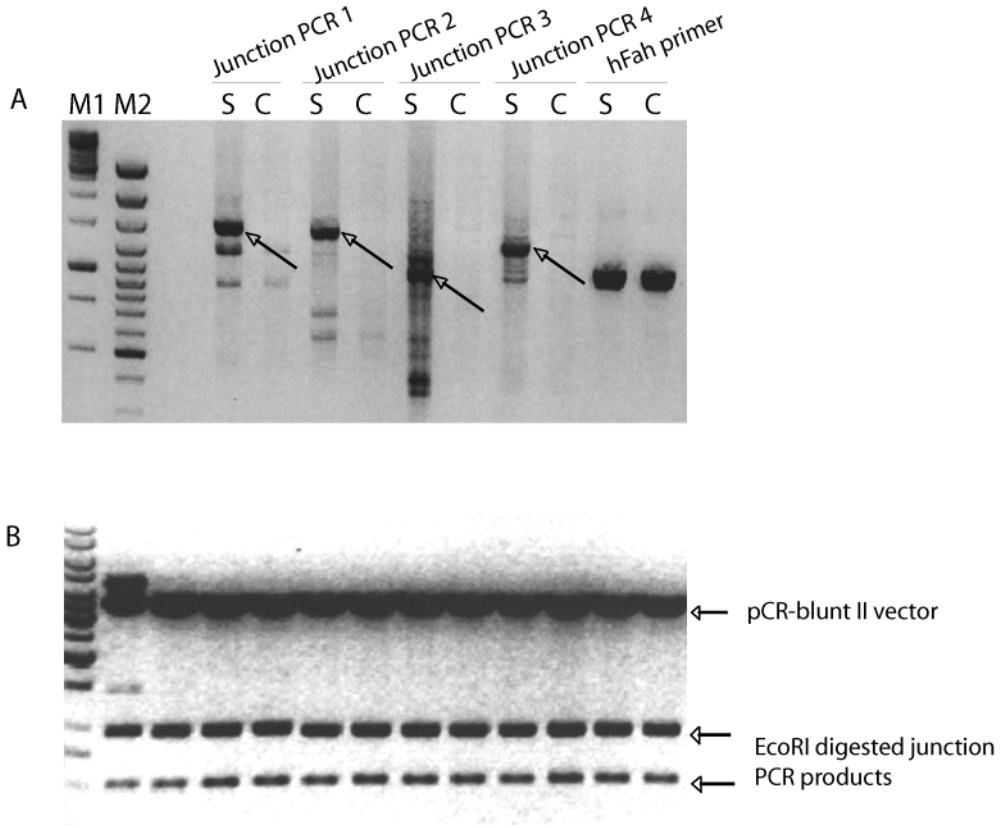


Figure 2.4. Site-specific junction PCR test and Junction enzyme digestion.

(A) Site-specific junction PCR test. All 4 junction PCRs showed specific bands only in liver samples from AAV-rDNA-FAH mice (S) but not in controls (C) where the linearized plasmid pAAV-rDNA-FAH was mixed with genomic DNA. Primers for the human FAH cDNA (last 2 lanes) were positive in both samples and controls. This experiment ruled out template switching in the detection of site-specific integration. Arrows point out the site-specific junction PCR products from different primer pairs. **(B)** EcoR I digested Topo vector cloned junction PCR products from different mice amplified with junction PCR2. All digested junction products showed the same pattern and lined up perfectly with rDNA.

AAV-rDNA-FAH copy number and site-specific integration ratio

The data described above clearly indicated that site-specific 28S rDNA integration did indeed occur reproducibly, but the precise frequency of such events remained unknown. Given the fact that the AAV-rDNA-FAH vector integrates about 30x more frequently than the stuffer control, one possibility was that site-specific integration accounted for the majority of this increase. To address this possibility, we performed Southern blot analysis of liver DNA from *Fah*$^{-/-}$ mice which had been rescued by AAV-rDNA-FAH injection, i.e. undergone selection for vector integration. DNA was digested with an enzyme that cuts once within the vector genome and we reasoned that a junction-specific band would be visible, if at least 10% of the integration events had occurred site-specifically in the 28S rDNA locus. While the integrated AAV genomes could be readily detected by using a human FAH probe, no junction fragments were seen (Fig.2.5A and Fig.2.6). The band sizes seen with single cutters were indicative of concatemeric genomes (Fig.2.6). This result indicated that the frequency of site-specific integration was below the detection limit of a southern blot and hence represented <10% of all integration events. The Southern blot was informative in regards to the number of integrated AAV copies per hepatocyte (Fig.2.5A). With serotype 2 each selected hepatocyte had 1-3 copies of *FAH* per diploid genome equivalent. AAV8 produced higher copy numbers of ~6-13 vg/cell. Thus, the majority of AAV integration events found were concatemeric with multiple copies.

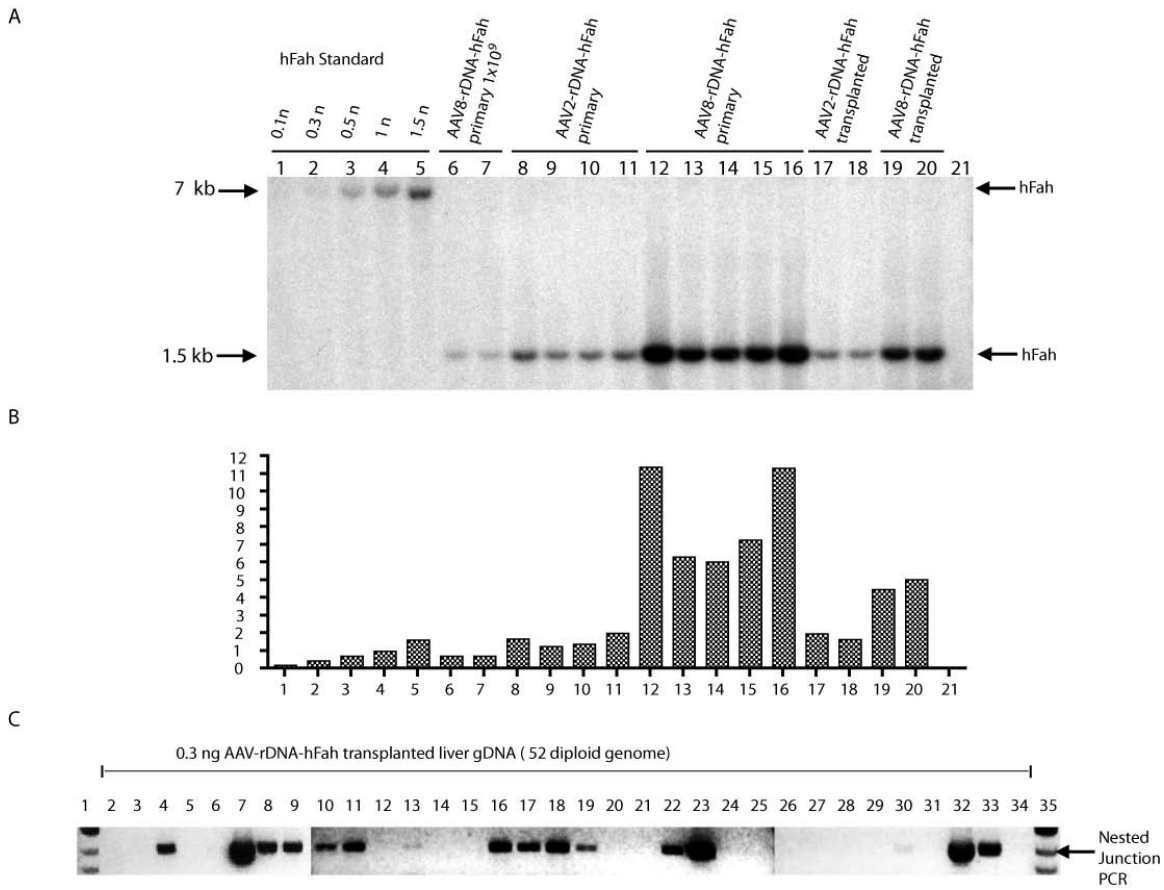


Figure 2.5. Analysis of integrated AAV-rDNA-FAH.

(A) FAH southern blot of liver DNA after selection for AAV-rDNA-FAH integration and full repopulation (>90%). Each lane is from one mouse from either primary repopulation after virus injection (lanes 6-16) or from serial transplant recipients of virally rescued livers (lanes 17-20). Both AAV2 and AAV8 were used. A dilution series of linear 7.0 kb plasmid pAAV-rDNA-FAH (0.1n-1.5n) was used for quantitation (left lanes).Location of probe and EcoR1 restriction sites shown in Fig.2.1A. **(B)** Quantification of integrated vector copy numbers by Phosphoimager readings. **(C)** Nested junction PCR on 0.3 ng of genomic DNA/sample from liver fully repopulated with AAV-rDNA-FAH transduced hepatocytes. Lanes 2 to lane 33: individual 0.3ng samples.Lane 34: water control. Lanes 1 and 35: size markers.

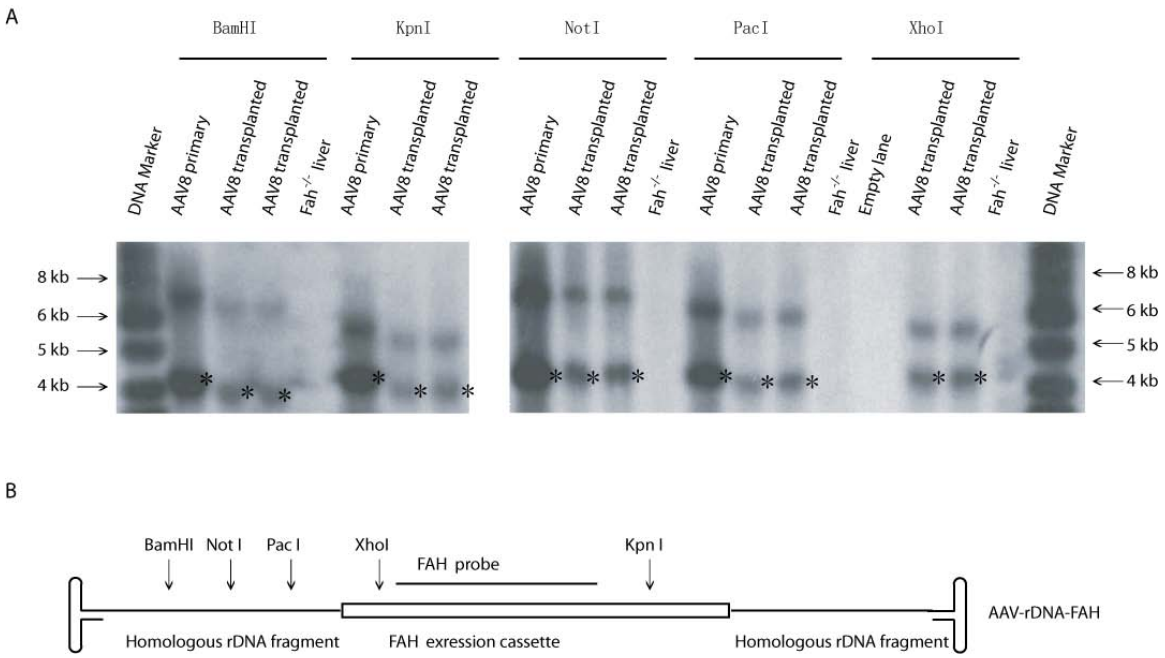
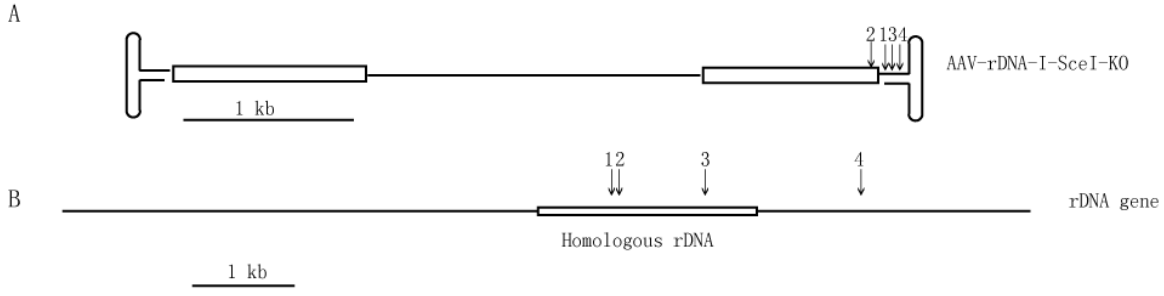


Figure 2.6. Southern blotting on AAV-rDNA-FAH infected liver concatemeric integrations.

(A) Different one-cutter enzyme digested genomic DNAs showed head to tail concatemers of AAV-rDNA-FAH (4.3 kb), lower band in each lane, marked by an *, head to head (*Kpn*I) concatemers or tail to tail concatemers (*Bam*HI: 7.9 kb, *Not* I: 7 kb, *Pac* I: 6.5kb, *Xho*I: 5.7kb) **(B)** Scheme of the vector with the position of one cutter enzymes and FAH probe.

Since the frequency of 28S site-specific integration was below the detection limit for Southern blot analysis, we devised a PCR based method to estimate the rate. The DNA from hepatocytes, which had undergone selection for FAH vector integration, was diluted such that each PCR tube contained the equivalent of 1, 5, 20, 50 and 100 hepatocytes. Next, 30-50 samples of each DNA amount was subjected to a highly sensitive nested junction PCR (Fig.2.5C). The number of samples positive for the junction PCR in each group was determined and the Poisson distribution was used to

infer the number of cells that had at least one site-specific integration event (Fig.2.5C). When 0.3ng of AAV8 injected adult hepatocyte repopulated liver genomic DNA per tube were used (= ~52 diploid genomes), 16/32 samples were PCR positive. Using Poisson distribution mathematics (see methods), this indicated that 1.32% of hepatocytes had a site-specific integration event. This Fig.2.5C represents a minimal estimate, as some events could have been missed due to limitations of the sensitivity of the PCR. In addition, not all liver DNA comes from hepatocytes and the majority of hepatocytes are polyploidy (Duncan et al., 2009; Duncan et al., 2010). Taken these factors into account, we estimated that ~2-3% of all selected hepatocytes had a 28S rDNA site-specific integration event. This number is compatible with the lack of a detectable junction fragment on Southern blot.



C

1	Sequence and break point position on AAV: 4120, Sequence and break point position on rDNA: 16702, Sequence of junction:	CAGGAACCCCTA
2	Sequence and break point position on AAV: 4050, Sequence and break point position on rDNA: 16752 Sequence of junction:	AGGAACCGCAGGTTC
3	Sequence and break point position on AAV: 4170, Sequence and break point position on rDNA: 17554, Sequence of junction:	TCGCTCACTGAGGC
4	Sequence and break point position on AAV: 4174, Sequence and break point position on rDNA: 19145 Sequence of junction:	ACTGAGGCCGCC

Figure 2.7. Analysis of AAV-rDNA integration sites in the rDNA region.

(A) Schematic of 4 integration breakpoints (from Table 2.1) on the AAV-rDNA vector. The AAV-rDNA vector genome is 4,354 nucleotides in length. The terminal 145 nucleotides (positions 1-145 and 4110-4254) represent inverted terminal repeats (ITRs). **(B)** Location of integration breakpoints on rDNA relative to the homologous sequence (position on rDNA: 10387-12425) present within the AAV-rDNA vector. Position is based on rDNA sequence GenBank: X82564.1. **(C)** Junction sites sequence alignments with the AAV-rDNA vector and rDNA locus.

Since site-specific homologous recombination could not explain the clearly documented increase in integration frequency, we used a plasmid rescue strategy (Miller et al., 2005; Nakai et al., 2003; Nakai et al., 2005b) to determine integration sites. The presence of GC-rich repeat sequences made this approach very difficult, but we were able to isolate 12 genomic integration sites. Of these 4 (30%) were located within the ribosomal DNA locus (Table 2.1), approximately 100x more frequent that expected since rDNA represents only 0.3% of the genome. The junctions were not perfect and contained deletions, indicating that integration did not occur by homologous recombination (Fig.2.7).

AAV-rDNA integration sites	Chromosome[a]	DNA strand	Location	Refseq gene	Gene function	Tumor gene[b]	Details
1	6	-	121806565	Mug1	Endopeptidase inhibitor activity	No	
2	8	+	42212855	Mtus1	Molecular function unknown	No[c]	
3	9	+	59322717	Arih1	Ubiquitin-protei n ligase activity	No	
4	11	-	7113263	Igfbp3	Insulin-like growth factor binding	No	
5	13	-	50529615	Intergenic	NA	NA	
6	15	-	3438032	Ghr	Growth factor binding	No	
7	x	+	139531170	Intergenic	NA	NA	
8	x	-	113127425	Intergenic	NA	NA	LINE
9	NA	-	16702[d]	rDNA	NA	NA	
10	NA	-	16752[d]	rDNA	NA	NA	
11	NA	-	17554[d]	rDNA	NA	NA	
12	NA	-	19145[d]	rDNA	NA	NA	

Table 2.1 integration sites analysis.

[a]: rDNA repeats are spread on several chromosomes and identification of the exact rDNA gene where integration occurred was not attempted. Therefore these are labeled as NA. [b]: For tumor genes, we used Tumor Gene Family Data Base (http://www.tumor-gene.org/TGDB/tgdb.html) and Retrovirus Tagged Cancer Gene Database (http://rtcgd.ncifcrf.gov/). [c]: Mtus1, also the name mitochondrial tumor suppressor 1, is not listed in Tumor Gene Family Data Base and Retrovirus Tagged Cancer Gene Database. Here we listed as "not tumor gene". [d]: The position on rDNA is based on *M.musculus* 45S pre rRNA gene (GenBank: X82564.1)

Improved persistence of factor IX levels after gene therapy with an rDNA AAV vector

As a preliminary test whether the rDNA vector design could also work for another disease model lacking the feature of transgene selection, we generated an AAV-rDNA-hF.IX as shown in Fig.2.8A. This vector contained rDNA homology arms similar to those of the FAH vector.

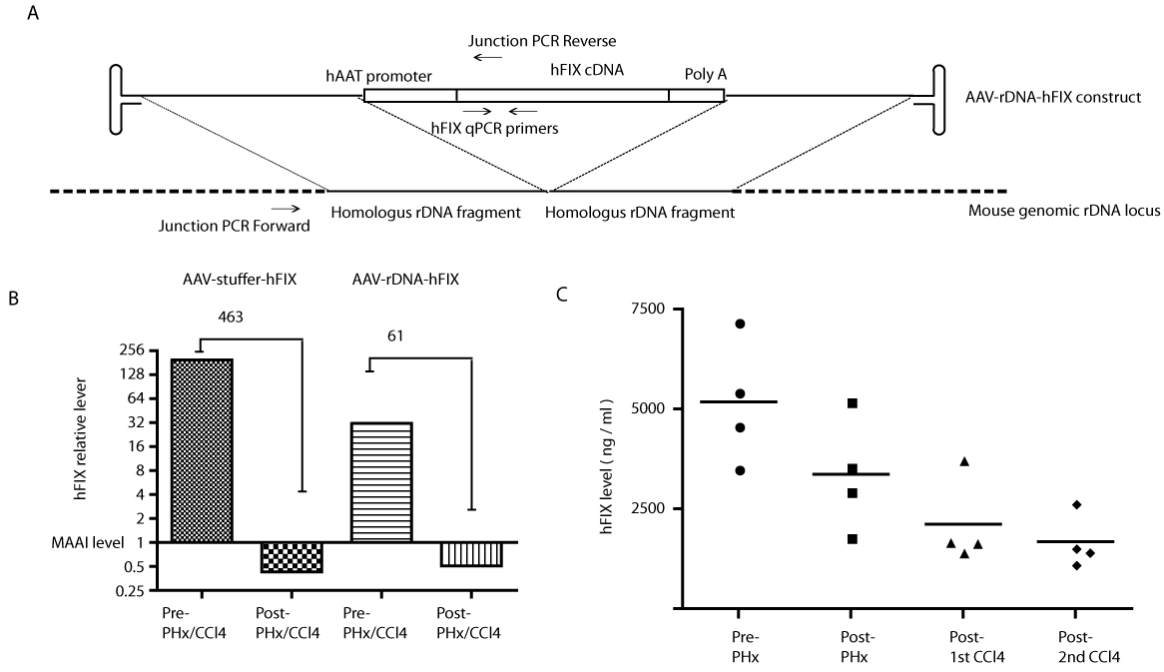


Figure 2.8. AAV-rDNA-hF.IX demonstrates improved transgene persistence.

(A) Structure of the AAV-rDNA-hF.IX vector. **(B)** Copy number changes of the hF.IX transgene after liver regeneration by qPCR. The *Maai* copy number served as an internal control. **(C)** hF.IX serum levels after partial hepatectomy and two sub-lethal CCl4 treatments. About 30% of the expression persisted.

Four 8-week old wild-type mice were injected with AAV serotype 8 at a dose of $3x10^{11}$ vg/mouse. 14 days after injection the baseline levels of hF.IX were measured and then, the mice were sequentially subjected to partial hepatectomy and two sub-lethal CCl_4

treatments. These manipulations induce high levels of hepatocyte turnover and hence eliminate most episomal AAV genomes. The serum levels of human factor IX were measured after each manipulation. Importantly, about 40% of the initial hF.IX level was still present despite massive hepatocyte replication in mice injected with the rDNA vector (Fig.2.8C). In contrast, as previously shown, only 5-10% of the human factor IX levels were present in those mice injected with the AAV-hF.IX vector (Nakai et al., 2001). In order to determine whether the improved persistence of FIX levels was due to a higher frequency of transgene integration, qPCR was used to measure hF.IX DNA copy number before and after liver replication. This was done by comparing the DNA of the liver lobe removed by partial hepatectomy to liver DNA obtained at final harvest. On average the copy number of hF.IX had dropped 61-fold in AAV-rDNA-hF.IX injected mice while the difference was 480-fold when the stuffer was used (Fig.2.8B). Most of this drop of course reflects loss of episomes, but the data show that a higher percentage of the vector genomes are resistant to cell turnover when an rDNA vector was used.

V Discussion

rAAV mediated gene therapy is rapidly emerging as the currently best technology for the treatment of some monogenic disorders, particularly those affecting the retina (Maguire et al., 2008) and liver (Manno et al., 2006), but also skeletal (Fisher et al., 1997) and cardiac muscle (Wang et al., 2005) and the central nervous system (Chen et al., 2009). Although cell turnover rates in these tissues are thought to be low or absent in the absence of injury, cell division clearly occurs in hepatocytes even during normal homeostasis (Grompe, 2006). Mild liver damage and hepatocellular death are a very common feature of many disorders, including viral infections (Iannacone et al., 2005) as well as nutritional (Mitchell and Herlong, 1986) and drug induced (Trautwein et al., 1998a; Trautwein et al., 1998b) insults. For this reason it can be predicted that therapeutic gene expression will not be permanent after gene transfer into hepatocytes if the transgene does not integrate into the chromosome. rAAV vectors are well known to remain mostly episomal and be lost quickly lost even after one round of cell replication (Nakai et al., 2000; Nakai et al., 2001). Permanent, life-long gene expression is, of course, desirable in the treatment of inherited disorders and therefore transgene persistence is an important issue. There are multiple strategies that may overcome this problem. One way to achieve transgene persistence is to re-administer another round of rAAV therapy. The immune responses to re-administration (Manno et al., 2006) could be overcome by switching serotypes (Pien et al., 2009; Vandenberghe et al., 2006). Alternatively, chromosomal integration of the rAAV vector would also achieve the desired goal. Here, two basic strategies can be considered. The first is gene repair, in which the disease causing mutation is removed by homologous recombination (Miller et al., 2006; Paulk et

al., 2010). This approach has the advantage of reproducing the natural levels of gene regulation and expression. Indeed, AAV vectors are particularly capable of producing gene repair (Miller et al., 2003; Miller et al., 2006) even *in vivo* (Miller et al., 2006; Paulk et al., 2010). Furthermore, the induction of DNA double-strand breaks near the mutation, using for example zinc finger nucleases, can enhance the efficiency of this approach. A major disadvantage of the gene repair strategy, however, is the fact that rAAV to be used is mutation specific. In the case of larger genes, different rAAV vectors would be required for each and every different mutation, causing practical and regulatory hurdles.

The second strategy is the one used by us. Chromosomal integration is targeted to a safe harbor (McCarty et al., 2004), which could be applicable for transgene expression in many different disorders. We targeted 28S rDNA in our experiments, but other loci could also be considered in the future. Overall, we observed a significant increase (~ 10-30x) of transgene integration and persistence when rAAV vectors contained 28S rDNA homology were used. It is important to note that the integration frequency of ~ 1/1,000 in Fig.2.2 was achieved at low doses for the purpose of retaining a countable number of integration nodules. At higher doses, the rDNA-FAH AAV vector produced integration in 2-5% of all hepatocytes. This increased frequency of stable integration lies considerably below frequencies that have been reported for retroviral vectors (Kalpana, 1999). Nonetheless, the increased integration frequency had functional significance. The rescue dose for the treatment of tyrosinemia was ~30-fold lower than for conventional vectors. In addition, a higher percentage of Factor IX expression was retained after liver regeneration.

In the case of the *FAH* transgene, we found only 2-4% of all integration events to be site-specific in 28S rDNA. Thus, the 30-fold increase in overall integration frequency cannot be explained by homologous integration. Nonetheless, our own limited integration site analyses suggest a strong enrichment for non-homologous integrations into rDNA. The exact mechanism by which the presence of rDNA homology can change the pattern of genomic integration is unclear.

Nonetheless, the increased integration frequency of such vectors is desirable for the treatment of genetic diseases. RDNA-containing AAV vectors should be considered when therapeutic transgenes are small enough to accommodate the inclusion of rDNA homology arms.

Chapter 3

Non-random integrating rAAV-rDNA vectors allow for stable transgene expression from rDNA loci

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

(Contribution statement: Zhongya Wang, under the supervision of Markus Grompe, conceived and initiated the AAV-rDNA-hF.IX project to test AAV-rDNA vectors in the hemophilia B disease model. Zhongya Wang constructed AAV-rDNA-hF.IX and AAV-stuffer-hF.IX vectors and did the preliminary AAV-rDNA-hF.IX in vivo experiments.
Leszek Lisowski, Ashley Lau, Feijie Zhang, under the supervision of Mark Kay, analyzed hF.IX levels, constructed short form AAV-rDNA vectors and did the LAM-PCR to identify integration sites of AAV-rDNA-hF.IX.)

As of May 14, 2012, under review at Molecular Therapy

I Abstract

While recombinant AAV vectors are proving to be efficacious in clinical trials, the episomal character of the delivered transgene restricts their effectiveness for 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 sequences capable of homologous recombination into genomic ribosomal DNA. We show that after *in vivo* delivery the rAAV-rDNA vectors integrated into the genomic rDNA locus 8-18 times more frequently than control vectors, providing an estimate that 23-54% of the integrations were specific to the rDNA locus. Moreover, a rAAV-rDNA vector containing a human factor IX expression cassette resulted in sustained therapeutic levels of serum hF.IX even after repeated manipulations to induce liver regeneration. Because of the relative safety of integration in the ribosomal DNA locus, these vectors expand the usage of rAAV for therapeutics requiring long-term gene transfer into dividing cells.

II 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(Blomer et al., 1997), a property allowing for persistent transgene expression from dividing cells but also increasing the risk of insertional mutagenesis.(Baum et al., 2003; Hacein-Bey-Abina et al., 2003b; Rosenberg and Jolicoeur, 1997)

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(Maguire et al., 2009) and hemophilia B(Mattar et al., 2011). The large number of natural AAV variants(Chiorini et al., 1999; Chiorini et al., 1997; Gao et al., 2002; Jennings et al., 2005; Kruger et al., 2008; Lee et al., 2005; Muramatsu et al., 1996; Rutledge et al., 1998; Taymans et al., 2007; Young et al., 2000b), as well as different strategies for the generation of new variants (reviewed in (Buning et al., 2008; Michelfelder and Trepel, 2009; Van Vliet et al., 2008)), provide diverse tissue tropism.

Nakai, *et. al.* have 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(Inagaki et al., 2007; Nakai et al., 2003; Nakai et al., 2001). 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 (Donsante et al., 2007).

The non-integrating 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 co-expression of other protein factors, such as AAV Rep protein which can be toxic (Dyall et al., 1999; Kogure et al., 2001; Recchia et al., 1999; Rinaudo et al., 2000; Rizzuto et al., 1999; Tsunoda et al., 2000).

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 moving 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 novel rAAV vector capable of safe and site-specific integration into the host genome. As a target sequence we selected ribosomal DNA (rDNA) operon for two reasons: multiple large-scale analyses have shown that rAAV integrations occur at the rDNA locus at a higher frequency than random(Inagaki et al., 2007; Miller et al., 2005; Nakai et al., 2003), possibly because rAAV is processed at the nucleolus(Moghrabi et al., 2009) 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 (Bengtsson et al., 1989; Sakai et al., 1995), 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 *D. melanogaster* non-LTR retrotransposons disrupt approximately 50% of the rDNA transcription units(Johansen et al., 2007), while studies in *S. cerevisiae* show that the number of rDNA genes fluctuates through natural mechanisms of rDNA deletions and additions(Kobayashi, 2006), without any effect on cell viability. Experiments in tyrosinemia type 1, a liver disease model, which selects for integrated rAAV-FAH genomes, indicated that rDNA AAV vectors indeed have higher integration frequencies. In order to determine whether this novel vector design has general utility even without transgene selection, we developed FIX rDNA-AAV vectors. Here we demonstrate that rDNA might be an ideal safe locus for transgene integration.

III Materials and Methods

AAV vector construction, vector production and vector titration

All vectors used in the *in vivo* studies carried 2.1kb hF.IX cassette composed of TTR promoter/synthetic enhancer, hF.IX cDNA+minimal Intron 1(Nakai et al., 1998), and synthetic poly A, flanked by marine 28S rDNA sequences. Vector used in the pilot study contained 1157bp rDNA downstream and 879bp upstream of the hF.IX cassette (from position 2266-3432 and 3427-4306 of the murine 28S ribosomal RNA, NCBI Ref. Seq. NR_003279.1). The rAAV-rDNA vectors used in the second part of the study contained rDNA flanks as follow: 1000bp rDNA vector 5' flank (2425-3433) 3'flank (3427-4426), 750bp vector 5'flank (2681-3433) 3'flank (3427-4178) and 500bp vector 5'flank (2933-3433) 3'flank (3427-3928) (all positions based on 28S sequence NCBI Ref. Seq. NR_003279.1). The 5' and 3' flanks overlapped by 6nt 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 hF.IX 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 1009 (NCBI Reference Sequence: NT_039433.7), while the 5' flank spanned sequences from 1004-2590. The control vectors used in the second study contained FAH DNA sequences (5' followed by 3'): 1010-2014 and 2003-3003 for the 1000bp vector, 1251-2014 and 2003-27258 for the 750bp vector and 1501-2014 and 2003-2502 for the 500bp control vector.

Vectors used in the NIH3T3 studies contained identical 1000bp rDNA or FAH stuffer flanks as described for FIX vector. The hF.IX cassette was replaced with

hPGK-eGFP-bGHpA cassette to allow for convenient monitoring of transgene expression by FACS. 880bp fragment of hF.IX gene was included downstream of the polyA to bring the total size of the vector to 4.7kb and to enable the use of 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(Grimm, 2002), 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. 6 to 8 weeks old female C57BL/6 mice were purchased from the Jackson Laboratory (Bar Harbor, ME). rAAV8-rDNA/Stuffer-hF.IX vectors were administered *via* retro-orbital (r.o.) injection in 200ul of 1xPBS. Two-thirds partial hepatectomies were performed as earlier described.(Nakai et al., 2001) Blood was collected at the indicated time points *via* r.o. bleeding, and plasma hF.IX levels were determined *via* an enzyme-linked immunosorbent assay (ELISA) as described previously (Walter et al., 1996).

Analysis of vector copy numbers 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(Sambrook, 1989) using [^{32}P]dCTP-labelled hF.IX specific probe and/or TaqMan. 20ug of genomic DNA was digested with *Bam*HI (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 standard. Multiplex real-time TaqMan PCR was performed and analyzed on ABI 7500 PCR System (Applied Biosystems, Foster City, CA) using hF.IX specific probe (5'-GTGCCATTTCCAT GTGGAAGAGTTTCTGT-3') and primers (Forward; 5'- AACCAGAAGTCCT GTGAACCAGCA-3', reverse; 5'- GCACG GGTGAGCTTAGAA GTTTGT-3'), and mouse β-actin probe (5'-GCTGTGTTC TTGCACTCCTTGCATGT-3') and primers (forward; 5'-TGAGACTCC CAGC ACACTGAACTT-3', and reverse; 5'-ACACTCAG GGCAGGTGAAACTGTA -3') .

LAM-PCR

Provector integration site analysis was performed on genomic DNA extracted from liver samples at the end of the study using linear amplification-mediated (LAM) PCR as previously described(Raake et al., 2008) with minor modifications. In brief, 100-200ng genomic DNA was used for 2x50cycles 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 3min, 50cycles of 98°C 10sec, 72°C 2.5min, followed by final extension at 72°C for 10min. After magnetic capture and second strand synthesis the dsDNA products were digested with *Eco*RV 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 30sec, 34 cycles of 98°C 10sec, 65°C 20sec, 72°C 2.5min, final extension

72°C 10min. 1/50th of the 1st PCR was used for nested PCR using LAM-SE3 or LAM-FIX3 and LAM-LK2 oligos: 98°C 30sec, 34 cycles of 98°C 10sec, 58°C 20sec, 72°C 2.5min, final extension 72°C 10min. PCR product was analyzed on agarose gel and used for TOPO cloning (Invitrogen, Carlsbad, CA) and subsequent sequencing.

LINKER1	TTCAGACCCGGGAGATCTGAATTCAGTGGCACAGCAGTTAGG
LINKER2	CCTAACTGCTGTGCCACTGAATTCAGATCTCCCGGGTC
LAM-LK1	CTGAATTCAGTGGCACAGCAG
LAM-LK2	TCAGTGGCACAGCAGTTAGG
LAM-SE1-Bio	AATCGCGGCCTCGAATCAATATTCGCGAG
LAM-SE2	ATAATTACCAGCGCGCGCTG
LAM-SE3	CCGCGCTGGTAATTATTAACTC
LAM-F.IX1-Bio	CGTGTTGTTGGTGGAGAAGATGC
LAM-F.IX2	CTTGGCAGGTTGTTTTGAATGG
LAM-F.IX3	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 transduced with rAAVDJ-stuffer/rDNA-hPGK-eGFP vectors at multiple MOIs (see Figure 3.4b). Transduction was allowed to proceed for 16hrs. 48hrs 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 unpaired t-test using GraphPad Prism 5 software.

IV Results

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

Experiments performed in Fah knockout mice indicated an increased transgene persistence with rDNA-containing vectors when compared to historical controls(Nakai et al., 2001). In order to determine if rDNA vectors truly were superior for integration into the 28S rDNA locus (Figure 3.1a), we compared two rAAV vectors: rAAV-rDNA-hF.IX and a size matched rAAV-stuffer-hF.IX control (Figure 3.1b). The sequences chosen as rDNA flanks were the immediate upstream and downstream sequences around I-*Ppo*I endonuclease recognition site. In this first experiment we intravenously injected six C57BL/6 female mice with 5E10vg/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.(Nakai et al., 2001) To further ensure that all the episomal rAAV sequences were lost, five weeks later, animals received two sublethal injections of CCl_4. Serum hF.IX levels in animals administered the rAAV-rDNA-hF.IX and rAAV-stuffer-hF.IX vectors (with and without PH/CCl_4) are shown in Figure 3.1c. After liver regeneration, the serum hF.IX levels in rAAV-rDNA-hF.IX-treated mice were 23% of what was observed in the first few weeks post-treatment (~1300ng hF.IX/ml before PH and ~300ng hF.IX/ml after PH and CCl_4) or 41% of non-liver regeneration controls followed for the same period of time at week 29 (Figure 3.1c and d). In contrast, rAAV-stuffer-hF.IX-treated mice had only 3.8% of their original serum hF.IX after liver regeneration or 9% of non-liver regeneration

controls followed for the same period of time (Figure 3.1c and d). This suggested that the rAAV-rDNA-hF.IX injected mice had enhanced vector integration compared to animals that received the control, non-rDNA containing vector. Of note, the rAAV-rDNA-hF.IX mice maintained a therapeutic level of serum hF.IX even after the induction of robust liver regeneration.

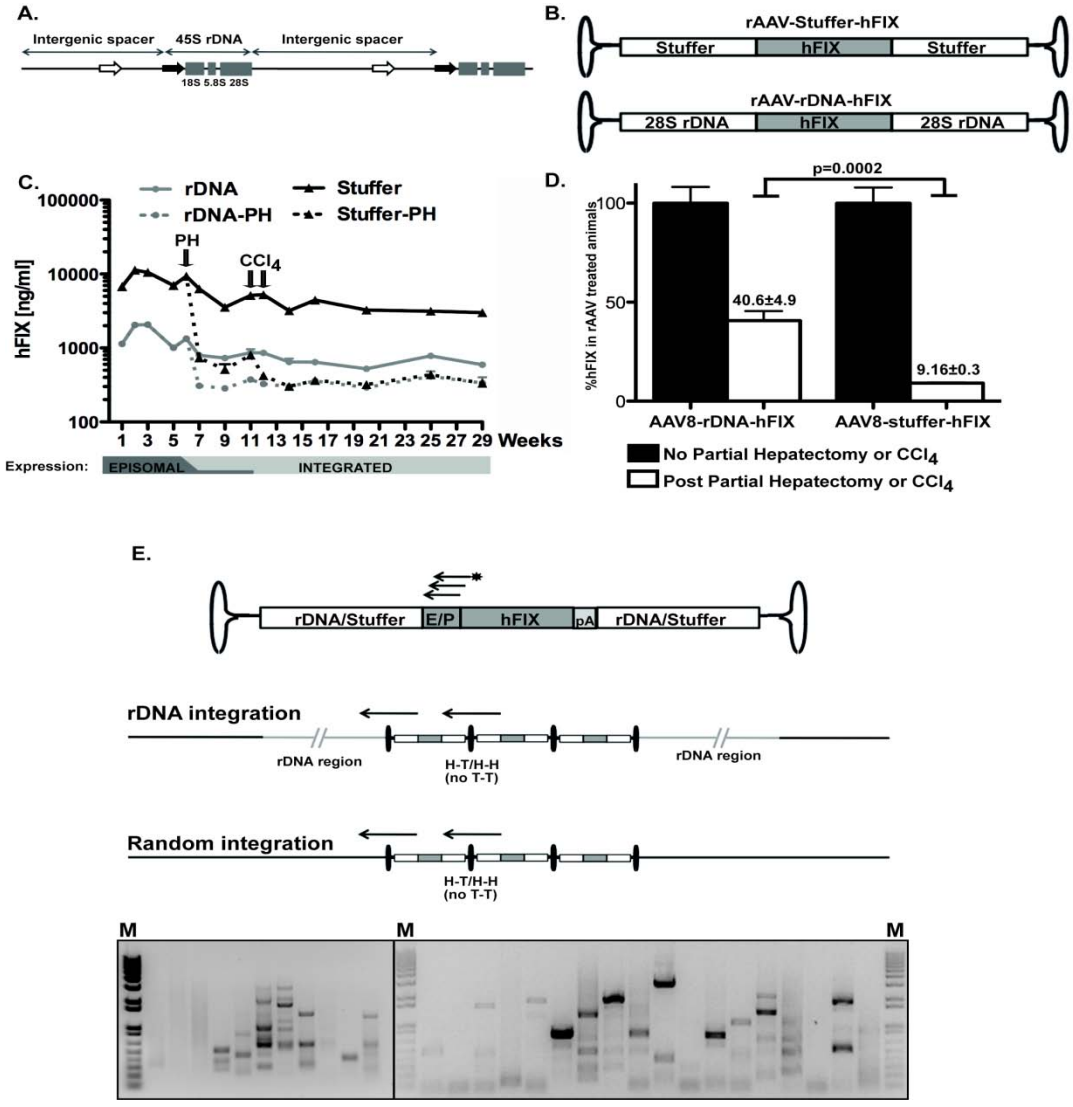


Figure 3.1. (A) Schematic representation of the organization of the rRNA genes in eukaryotes (Adopted from (Preuss and Pikaard, 2007)). (**B**) rAAV-rDNA-hF.IX and rAAV-stuffer-hF.IX vector design. The white boxes represent the 28S rDNA and the

corresponding “stuffer” sequences. The gray boxes represent hF.IX expression cassette (TTR promoter, hF.IX cDNA+1st intron, and bovine growth hormone pA). (**C)** hF.IX 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. (**D)** Quantification of hF.IX levels at the end of the study (week 29). (**E)** LAM-PCR design and sample gel (M indicates DNA marker). The biotinylated primer used for primer extension step is shown as arrow with a star at the end. Long arrows show two possible amplicons – internal non-informative amplicon and external informative amplicon containing genomic DNA.

Detailed Southern blot analysis indicates differences in provector integration profiles between rAAV-rDNA and the rAAV-stuffer vectors

To study the provector integration profile, we performed Southern blot analyses on genomic DNA from the livers of animals that had received rAAV-rDNA-hF.IX or rAAV-control-hF.IX with or without induction of liver regeneration (Figure 3.2). As we have previously shown, prior to PH/CCl_4 the majority of rAAV proviral genomes persist as monomeric or concatemeric episomes, while after liver regeneration only the integrated provector remains.(Nakai et al., 2001) Using a restriction enzyme that cuts twice in the vector sequence, we determined and found as expected(Nakai et al., 2001) that the average total vector copy number (VCN) was significantly higher in animals that did not undergo liver regeneration compared to those that did (Figure 3.2a). Episomal and

integrated vectors formed random concatemers (due to digest performed and probe used only head-to-head (H-H) and head-to-tail (H-T) orientations are visible in Figure 3.2b). The monomeric and concatemeric forms of the episomal vector in non-liver regeneration animals were also detected using a zero-cut restriction enzyme (Figure 3.2c); while the lack of detectable bands in the PH/CCl_4 animals treated with rAAV-rDNA or rAAV-stuffer vectors indicated that the remaining rAAV genomes were likely integrated into the genomic DNA. The absence of any strong bands indicated that the vectors did not integrate into a single site or that the provector integrated as a large concatemer and thus was not transferred efficiently from the gel onto the membrane.

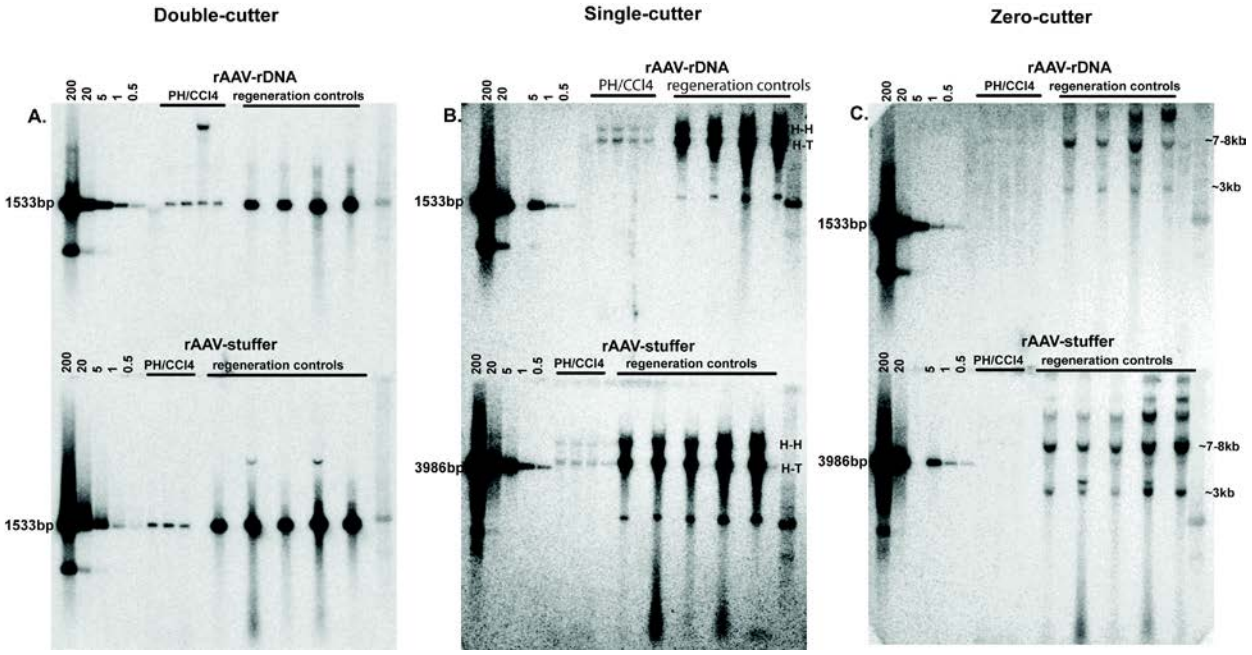


Figure 3.2. Southern blot analysis of vector genome structure in liver samples from animals postPH/CCl4 and control animals without liver regeneration followed for the same period of time. Liver DNA samples from animals treated with rAAV-stuffer

and rAAV-rDNA vectors (as indicated in the figure) were analyzed by Southern blot using double-(**A**), single-(**B**), or zero-cutter(**C**) endonucleases. For rAAV-rDNA group, 4 animals that underwent PH/CCl_4 treatment and 4 control animals without liver regeneration followed for the same period of time were analyzed. For rAAV-stuffer treatment group, samples from 3 PH/CCl4 group and 5 control animals were analyzed. Standards corresponding to 200, 20, 5, 1, and 0.5 copy/genome were run on the same gel mixed with digested genomic DNA from an untreated control animal. Two DNA fragments (1533bp and 3986bp) were used as standards. For rAAV-rDNA vector the expected bands were: single episome - 4.1kb; Head-to-head (H-H) orientation of the rAAV genomes in concatemers – 6.0kb, and Head-to-tail (H-T) – 4.1kb. For rAAV-stuffer vector: single episome – 4.1kb; Head-to-head orientation of episomes in a concatemer – 7.4kb; and head-to-tail episomes – 4.1kb.

Despite the fact that the average VCN was higher in the postPH/CCl_4 rAAV-stuffer treated animals than in the samples from animals treated with rAAV-rDNA (Figure 3.2a), when the same samples were analyzed using a single-cut restriction enzyme (Figure 3.2b), the signal was stronger for the rAAV-rDNA postPH/CCl_4 animals than corresponding rAAV-stuffer controls. Furthermore, the PH/CCl_4 treatment groups analyzed using a zero-cutter showed more of a smear-like signal in the rAAV-rDNA treatment groups, but not for the rAAV-stuffer controls. Together this supports the hypothesis that the rAAV-stuffer vector integrated into larger number of random locations in the genome than the rAAV-rDNA vector.

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

In order to investigate if 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 (LAM) PCR analyses. The design of LAM-PCR is shown in Figure 3.1e. Multiple biotinylated oligos, endonucleases and polymerases were tested and compared side-by-side in order to derive the most optimal combination.

LAM-PCR analysis of liver DNA samples from 3 separate animals in each group confirmed that 84% of analyzed informative clones in the rAAV-rDNA-hF.IX group were integrated in the rDNA genomic loci, while the remaining 16% were integrated into random DNA loci or could not be analyzed as the specific sequence returned no matches to the mouse genome. In the rAAV-stuffer-hF.IX group, over 60% of analyzed informative integration sites were found to be in random genomic locations, 2% were found in the rDNA loci, while the remaining 38% were integrated in genomic regions 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, this 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 500-1000bp

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 3.3a), in which the hF.IX

cassette (identical to the one used in the first study, see above) was flanked with 500bp, 750bp or 1000bp of 28S rDNA. Control mammalian stuffer DNA replaced the rDNA sequences in control vectors (Figure 3.3a).

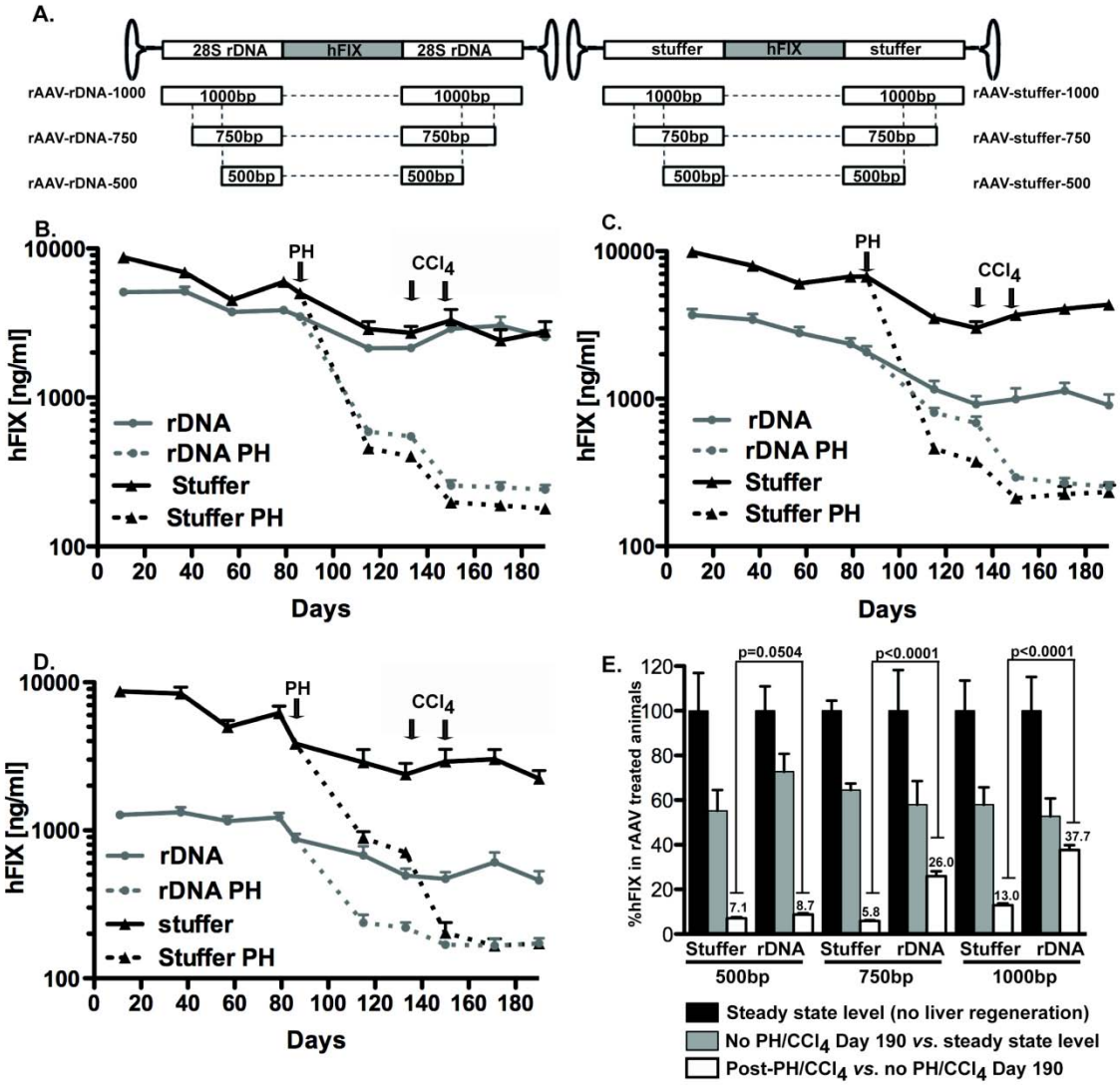


Figure 3.3. ***In vivo*** **comparison of hF.IX levels in animals treated with 500bp, 750bp, or 1000bp rDNA or stuffer vectors. (A)** Graphic representation of the six vectors used in the study. The gray box represents hF.IX expression cassette (TTR promoter, hF.IX cDNA+1st intron, and bovine growth hormone polyA). The size of each flank is indicated. **(B-D)** hF.IX levels throughout the study in animals treated with

rAAV-rDNA/stuffer-hF.IX 500bp (**B**), 750bp (**C**), or 1000bp (**D**). 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 PH and CCl_4 treatments. (**E)** Quantification of data shown in (B-D). hF.IX levels at the last time point before PH were used for normalization and thus were assigned a value of 100%. Shaded bars represent hF.IX levels on day 190 in animals that did not undergo PH/CCl_4. White bars represent hF.IX 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.

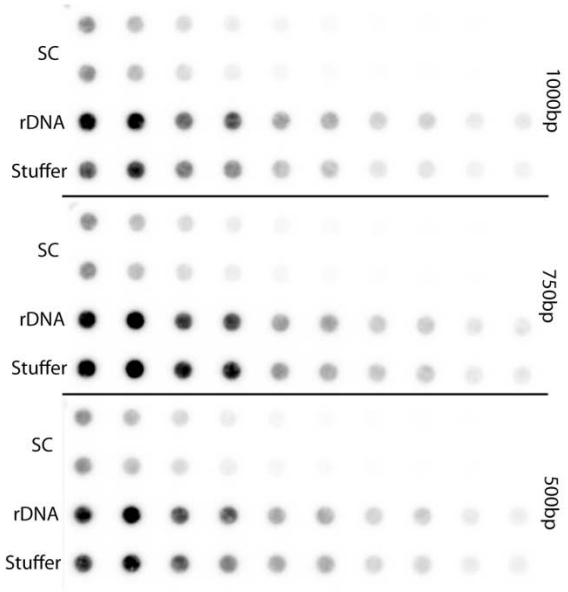
As the size of the flanking sequences increased so did the overall size of the vector making them 3.6, 4.1kb, and 4.6kb in length, respectively. We decided not to introduce additional stuffer sequences in the shorter vectors so as to not complicate the comparison between rAAV-rDNA and stuffer vectors. Despite the significant differences in vector sizes, as shown in Figure 3.4, all 6 vectors were packaged into AAV8 capsid at very similar efficiencies resulting in rAAV preparations with titers between 1.3-2.6E13vg/ml.

rDNA flanks of 750bp 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 $5x10^{10}$vg/animal of each construct into fifteen C57BL6 age matched female mice and monitored the serum hF.IX levels at various time points throughout the experiment (Figure 3.3b-d). On day 86, ten of the animals from each treatment group underwent a 2/3 partial hepatectomy, followed by CCl_4 treatments on days 133 and 150. The detailed

comparison of serum hF.IX levels at the last time point of the experiment (day 190) is shown in Figure 3.3e.

A.



B.

Construct		Titer [vg/ml]
rDNA	1000bp	1.90E+13
	750bp	1.95E+13
	500bp	1.36E+13
Stuffer	1000bp	1.33E+13
	750bp	2.00E+13
	500bp	2.55E+13

Figure 3.4. Titer comparison of vectors used in the study. (A) Dot Blot titration of the vectors shown in Figure 3.3A. As each paid of vectors differed from the other vectors in size, three independent dot blots were performed, each with an appropriate standard curve (SC). (**B)** Quantification of data from (A).

The relative difference in the serum hF.IX levels in animals with or without liver regeneration was similar between the 500bp rAAV-rDNA vector (4.8% prior to PH/CCl_4,

or 8.7% of non-liver regeneration controls) and the AAV stuffer controls (5.1% prior to liver regeneration, or 7.1% of non-liver regeneration controls) (Figure 3.3 b and e). In contrast, the difference between the AAV rDNA and stuffer control vectors was more substantial in the 750bp and 1000bp groups. In the 750bp rAAV-rDNA group after PH/CCl_4 treatment, serum hF.IX levels were 11.2% of the levels observed before the induction of liver degeneration or 26.0% of the levels observed in control animals that received the same vectors but were not subjected to liver regeneration while relative hF.IX expression in the 750bp AAV-stuffer control group were 3.8% and 5.8%, respectively. Thus the relative level of hF.IX in rAAV- rDNA treated animals were about 4-times higher than in corresponding stuffer control animals. In the 1000bp group the difference in average level of hF.IX after liver regeneration between rAAV-rDNA and rAAV-stuffer treatment group was about 3-fold (38% in rDNA *vs*. 13% in the control group).

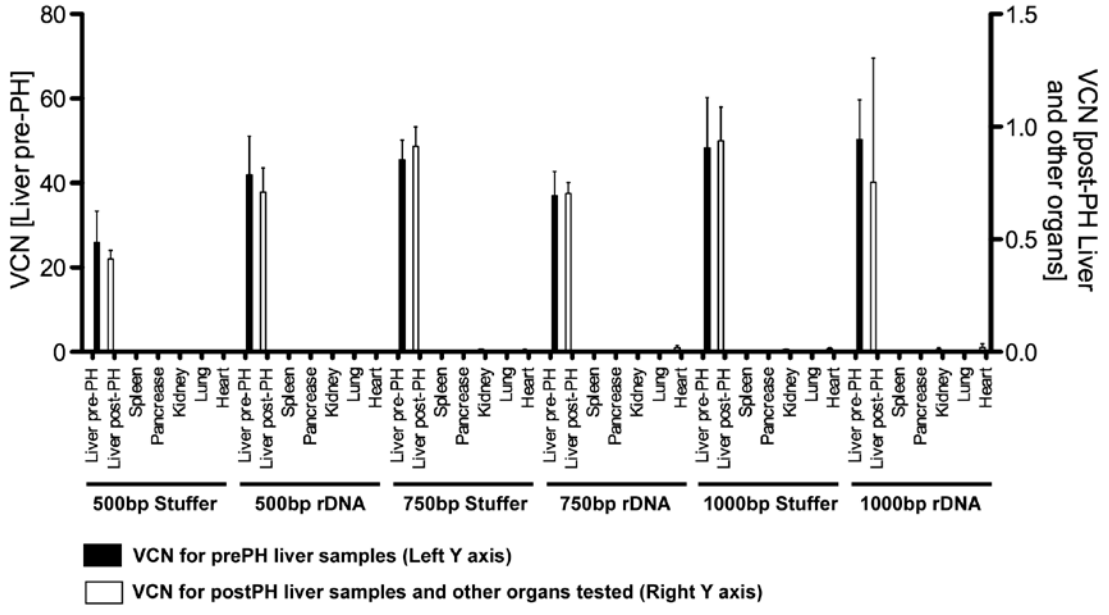


Figure 3.5. VCN and vector tropism comparison. DNA samples from prePH liver samples and samples obtained at the end of the study were analyzed by TaqMan for VCN.

VCN in prePH liver samples is shown in black (plotted against left Y axis). VCN data obtained from postPH liver samples and other organs harvested at the end of the study are shown in white (plotted against right Y axis).

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 VCN in different tissues extracted at the end of the study (Figure 3.5). 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.

rDNA sequences might affect transgene expression from episomal rAAV vectors

In the course of the study it was noticed that despite the fact that all animals were injected with the same total number of vector particles, the expression of serum hF.IX differed between rAAV-rDNA and rAAV-stuffer treated animals in each group (Figure 3.3 b-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). There was a trend that the inclusion of rDNA sequences inhibited episomal transgene expression. The longer the insert the greater the difference between AAV-rDNA and AAV-stuffer mediated transgene expression. The

biggest difference was observed in the 1000bp group at the earliest time point with the rDNA group expressing at about 15% of the "stuffer" control group (1300ng/ml *vs.* 8700ng/ml).

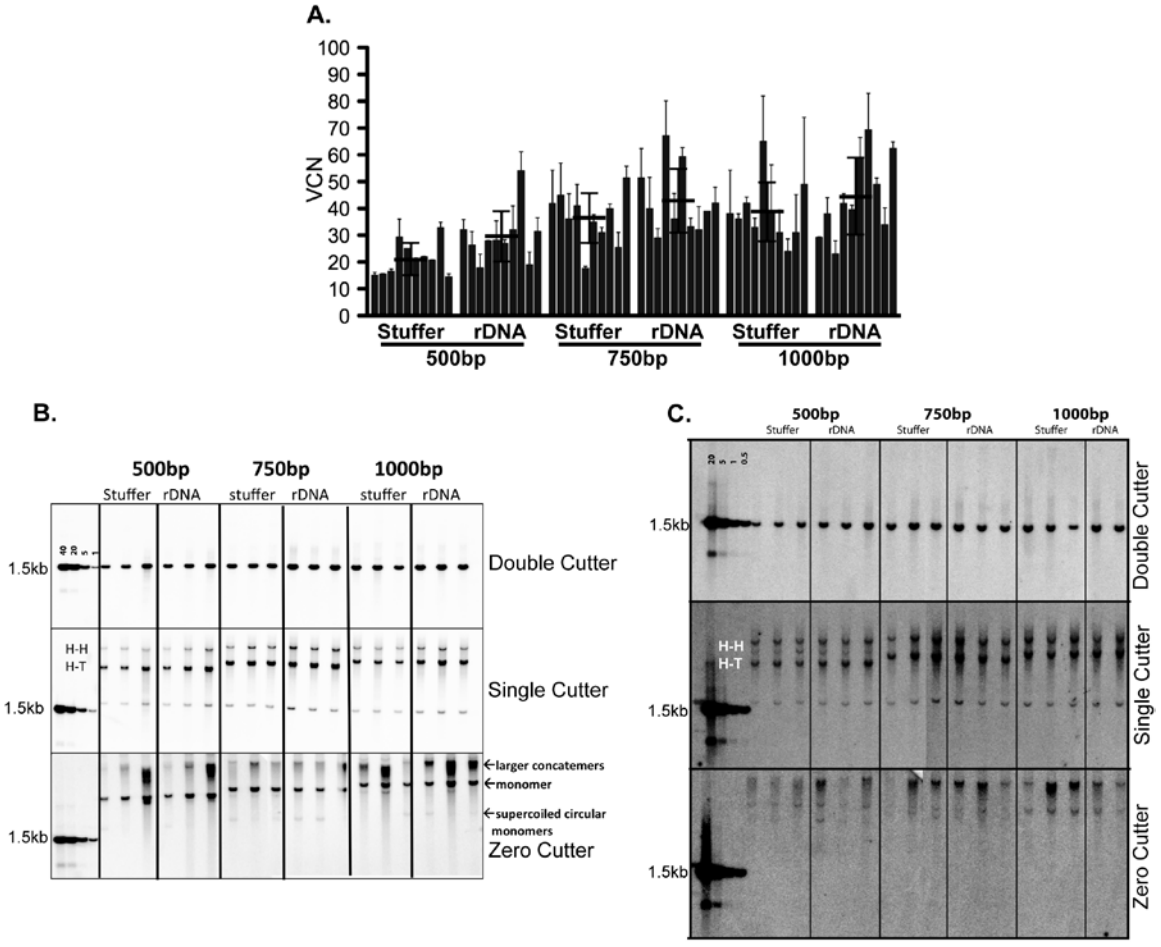


Figure 3.6. VCN and vector genome structure analysis. (A) Vector copy number (VCN) obtained from Southern blot analysis of liver samples harvested during partial hepatectomy (PH) on day 86. Black bars indicate the average value for the given group ±standard deviation. (**B) and (C)** Southern blot analysis of vector genome structure in

liver samples obtained at the end of the study from non-liver-regeneration control (**B**) or postPH/CCl_4 (**C**) 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 (1533bp). Plasmid DNA was diluted to equal 40, 20, 5, or 1 vector copy per haploid genome in (**B**) or 20, 5, 1, 05 vector copies per haploid 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.

Based on the fact that the average VCN in liver samples taken during PH on Day 86 were not significantly different between the experimental groups (Figure 3.3a), we conclude that the inclusion of rDNA sequences did not affect vector DNA transfer efficiency but rather interfered with expression from the episomal vectors.

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

In order to establish if the inclusion of rDNA had an effect on AAV proviral DNA structure, which perhaps could explain the observed differences in transgene expression, Southern blot analysis was performed (Figure 3.6 b and c). Analysis of samples from animals that did not undergo liver regeneration revealed all expected episomal proviral forms (Figure 3.6b). Single-cutter revealed bands corresponding to head-to-head (H-H) and head-to-tail (H-T) orientations of monomeric vectors in concatemers, while zero-cutter revealed bands corresponding to supercoiled circular monomers, monomers

and larger concatemers. 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 3.6c), the single-cutter similarly produced bands corresponding to H-H and H-T orientation originating from concatemers, while the zero-cutter gave rise to high molecular weight bands in all treatment groups. Similarly to results discussed earlier (Figure 3.2), due to low VCN, the Southern blot analysis of postPH/CCl_4 samples proved to be much more challenging than expected and making it difficult to determine if there were any differences in provector integration between rDNA and control vectors.

rAAV-rDNA vectors integrate into rDNA sequences 8-18 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 if 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 3.1e). 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 500bp group would require the shortest PCR amplicon in order to obtain DNA sequence outside of the integrated provector, and the 1000bp group would require the longest PCR product in order to be informative. This increases the chances of non-informative PCR clones caused by premature PCR termination on the GC rich rDNA

sequences obtained from the 750bp and 1000bp rDNA vectors. Furthermore, as shown in Figure 3.1e, due to the fact that AAV often integrates as concatemers (Head-Head and Head-Tail) it was predicted that fraction of sequenced LAM-PCR clones would contain non-informative DNA sequences derived from the internal vector sequences.

A. Unique Sequences Only

	Informative		Non-informative
	%rDNA	%Random	%(H-H/H-T/ET)
500 stuffer	3.3	43.3	53.3
	7.1	92.9	
500 rDNA	37.5	20	42.5
	65.2	34.8	
750 stuffer	5.6	44.4	50.0
	11.1	88.9	
750 rDNA	61.9	11.8	26.2
	83.9	16.1	
1000 stuffer	2.4	39.0	58.6
	5.9	94.1	
1000 rDNA	43.8	12.5	43.7
	77.8	22.2	

B. All Sequences

	Informative		Non-informative
	%rDNA	%Random	%(H-H/H-T/ET)
500 stuffer	1.3	40.0	58.8
	3.0	97.0	
500 rDNA	45.9	11.8	42.4
	79.6	20.4	
750 stuffer	5.7	34.5	59.8
	14.3	85.7	
750 rDNA	52.3	12.5	35.2
	80.7	19.3	
1000 stuffer	1.1	44.4	54.5
	2.4	97.6	
1000 rDNA	40.2	17.2	42.5
	70.0	30.0	

✱H-H – Head-to-Head; H-T – Head-to-Tail; ET -- Early Termination;

n=3 animals/group (80-100 sequences)

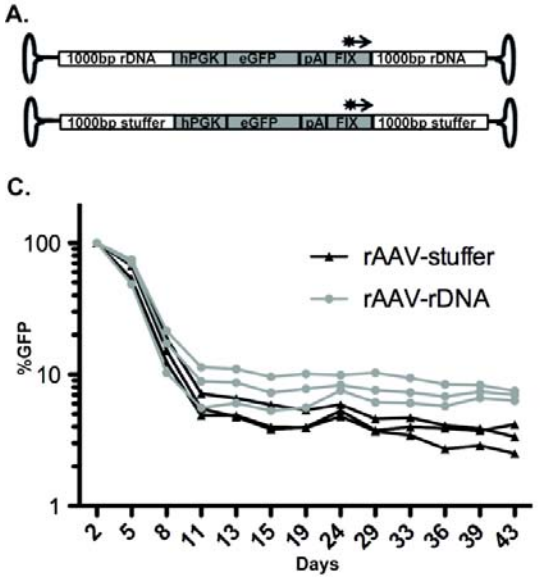
Table 3.1. LAM-PCR analysis.

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. Non-informative sequences include internal-amplicons (see Figure 3.1E), 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.

The summary of LAM-PCR data is shown in Table 3.1. Table 3.1a shows data for unique sequences only, so if more than one LAM-PCR clone had exactly the same sequence, it was counted only once. On the other hand, Table 3.1b summarizes all the

sequences, so if a given sequence was found more than once, they were counted as individual events. The reason for this dual analysis is the fact that rDNA sequences are highly repetitive and rDNA loci exist in many locations throughout the genome. It was believed, that if the vector integrates in the same rDNA sequence (in the same or different loci) this would likely lead to identical or very similar amplicons, which if counted only once would underestimate the efficiency of targeted integration. Furthermore, we wanted to eliminate the risk of scoring the same integration site multiple times, if a single amplicon generated multiple TOPO colonies. However, when comparing Table 3.1a and 3.1b it becomes clear that regardless of the type of analysis the final comparisons were similar.

In the 500bp group, the frequency of rDNA specific integration increased 9.2 times between rAAV-stuffer and rAAV-rDNA (7.1% rDNA integration in the control group vs. 65.2% in the rDNA group) when only the informative sequences were taken into account, or 11.4 times higher if all sequences were analyzed (37.5% *vs*. 3.3%). In the 750bp group, the rDNA flanking sequences increased the specificity of rDNA integration among informative clones by 7.6 fold (11.1% for control *vs.* 83.9% for rDNA vector) or by 11 fold if informative and non-informative clones were counted (61.9% *vs*. 5.6%). In the 1000bp group the difference was the largest – 13.2 fold (5.9% *vs*. 77.8%, stuffer vs. rDNA) or 18.3 fold when all clones were included in the analysis (43.8% *vs*. 2.4%). Using a plasmid rescue approach, we have previously established that about 3% of rAAV vector genomes integrations fall into rDNA loci (Inagaki et al., 2007). Based on that, we can roughly estimate that 23-54% of rDNA vector integrations were targeted into the rDNA locus.



B.

	AAV-stuffer		
	Cell confluency	MOI	%GFP Day2
1	80%	8000	65
2	100%	9000	77
3	100%	5000	56

	AAV-rDNA		
	Cell confluency	MOI	%GFP Day2
1	80%	9000	61
2	100%	17000	81
3	100%	6000	46

D.

	AAV-stuffer	AAV-rDNA	Fold increase
1	3.36	7.49	2.2
2	2.52	6.32	2.5
3	4.16	7.03	1.7

E.

	Informative		Non-informative
	%rDNA	Random	(H-H/H-T/ET)
1000 stuffer	5.1	39	55.9
	11.5	88.5	
1000 rDNA	23.2	19.6	57.2
	54.2	45.8	

Figure 3.7. rAAV-rDNA vs. rAAV-stuffer comparison in NIH3T3 cells. (A) Graphic representaion of vectors used in the study. Vectors encoded hPGK-GFP-pA cassette flanked by 1000bp rDNA or "stuffer" control DNA. Fragment of hF.IX cDNA was cloned downstream of polyA to keep the overall size of the vectors identical to the 1000bp rAAV-rDNA/stuffer vectors used in the *in vivo* studies (see Figure 3.3A). The black arrow indicates the biotinylated oligos used during primer extension step of LAM-PCR. (**B)** Comparison of different conditions used to transduce the NIH3T3 cells. Cells were transduced at 80% or 100% confluency (as indicated) and different multiplicities of infection (MOI). The resulting GFP levels at the first time point (day 2 post transduction) are indicated. (**C)** GFP levels in cells transduced with 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. Fold difference is shown in red. (**E)** LAM-PCR data. DNA samples extracted from cells on day 43 were analyzed by LAM-PCR. Percentages of informative and non-informative sequences are shown (see explanation in legend to Table 3.1).

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-1000bp constructs and replaced the TTR-hF.IX-pA cassette with hPGK-eGFP-pA cassette, which allowed for quick and efficient transgene expression analysis by FACS. As the new transgene cassette was smaller than the hF.IX cassette, we introduced 400bp of hF.IX cDNA as a DNA spacer between polyA and 3'rDNA/stuffer to keep the vector size the same as in the *in vivo* studies (Figure 3.7a). 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 1000bp "stuffer" sequences in place of rDNA flanking sequences (Figure 3.7a). Both vectors were packaged into AAVDJ(Giering et al., 2008) capsids 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 MOI, there might be a difference in initial levels of GFP expression between rDNA and stuffer control vectors (Figure 3.7b-d). Thus cells were transduced at two densities (80% and 100% confluency) 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 3.7b shows cell density, MOI and GFP levels on day 2 for 3 groups transduced with rAAV-stuffer/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 3.3b-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 3.7c). By TaqMan PCR analyses, the VCN paralleled the GFP levels (data not shown). As shown in Figure 3.4c and d, on day 43 the percentage of GFP positive cells in groups treated with rDNA vectors were 1.7-2.5 times higher 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 if 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 3.7e shows the summary of LAM-PCR results. 11.5% of the rAAV-stuffer vector integration sites were located in rDNA loci. In contrast, 54.2% of the rAAV-rDNA vector integration sites were found to be in rDNA, or a 4.7 fold increase in the rDNA specific integration when compared to the control vector. While the differences were not as great as the *in vivo* results, enhanced rDNA integration was observed in cultured cells.

Southern blot analysis

Liver samples obtained during partial hepatectomy as well as at the end of the study from animals that underwent liver regeneration and from control animals followed for the same period of time were analyzed by Southern blot analysis using double-,

single- and zero-cutter enzymes to digest the genomic DNA. The use of a double-cutter allowed for quantification of average vector copy number (VCN) per haploid genome. DNA from naïve control mouse spiked with plasmid DNA and digested with the same enzyme served as vector copy standard. A double cutter does not allow for the discrimination of integrated and episomal rAAV DNA or between monomeric and concatemeric episomal DNA forms.

The use of an enzyme that cuts only once within the vector (a single-cutter) allowed us to distinguish episomal and integrated monomeric genomes from concatemeric forms in head-to-head (H-H) and head-to-tail (H-T) orientations. The episomal genomes will only give bands corresponding to head-to-head and head-to-tail arrangements of monomers within a concatemer. However, if the vector integrated into genomic DNA, each integration site will produce a band of unique size depending on where in the neighboring genomic DNA the enzyme cuts. If the vector integrated as a monomer, each copy of the vector will give a unique signal on the blot, but if vector integrated as a larger concatemer, only the single most outside monomer within the concatermer will produce a unique band, while all internal genomes will produce bands corresponding to head-to-head and/or head-to-tail organization of the concatemer.

Based on the location of the enzyme recognition site in the vectors, for the rAAV-stuffer, used in the pilot study, the H-H band was expected to be 7.4kb and H-T 4.1kb, while for the rAAV-rDNA vector, the corresponding bands were expected at 6.0kb and 4.1kb. As shown in Figure 3.2b, all visible bands in samples from control and PH/CCl_4 animals correspond to H-H and H-T orientation, with the exception of a 1.6kb band visible in control animals only. For the vectors used in the second part of the study,

the expected bands for the H-T orientation of AAV genomes in the concatemer were 4.5kb, 4.0kb, and 3.5kb for the rAAV-rDNA/Stuffer-1000bp, 750bp and 500bp vectors, respectively, while the expected H-H bands were 6.5kb, 6.0kb, and 5.5kb (Figure 3.2b and c). The non-specific 1.6kb band discussed earlier was present in the samples from rAAV-rDNA and rAAV-stufer treated animals and did not change in size with the increase of total vector sizes.

The use of zero cutter, which cuts only in genomic DNA and not inside of the provector (Figure 3.2c), allowed for detection of specific bands corresponding to monomeric and concatemeric forms of vector genomes in samples. A zero-cutter should produce a unique size band for each provector integration, independently if the vector integrated as a monomer or a concatemer. Unintegrated episomal genomes would produce bands corresponding to a single monomeric form (~4.1kb) or multiplications of 4.1kb for higher concatemers. Due to the fact that the episomes are circular DNA, the episomes migrate at a faster rate in the agarose gel than linear standards or DNA ladder, causing the specific bands to appear smaller than expected (~3kb instead of expected 4.1kb band for a single monomeric form) (Figure 3.2c). For the vectors used in the second *in vivo* study, the size of a monomer was 4.5kb, 4.0kb, and 3.5kb for the 1000bp, 750bp, and 500bp vectors, respectively.

LAM-PCR analysis of provector integration sites

As it was impossible to predict the exact mechanism of genomic integration, making it hard to predict what sequences of the vector will be present in the integrated provector, and thus what sequences could be used as a template for LAM-PCR specific

oligos, we modified the standard LAM-PCR protocol. The biotinylated oligo, used for the linear primer extension step of LAM-PCR, was not placed close to the end of the vector as is frequently done in standard LAM/LM-PCR. Instead, the biotinylated primer annealed to an internal region of the vector in the enhancer/promoter region of the expression cassette for the hF.IX vectors (Figure 3.1e) (or within the hF.IX cassette in the rAAV-hPGK-eGFP-pA vector). Thus only provectors containing the promoter region could serve as a template for LAM-PCR. It is important to keep in mind that this does not exclude the possibility of amplifying highly rearranged provectors that still contain promoter/enhancer sequences but lost the ability to express transgene. However, this approach should greatly increase the frequency of amplification of integrated and structurally intact, and thus potentially functional, transgene expression cassettes. As a consequence of such primer placement, the LAM-PCR amplicon will have to be significantly longer than in a standard LAM-PCR in order to include genomic sequences outside of vector encoded rDNA/stuffer sequences. Three different biotinylated oligos binding to the enhancer/promoter region of the vector were tested and the oligo that produced the lowest unspecific background was selected for further studies. Due to the fact that rDNA sequences present in the vector and at the predicted site of homologous recombination are very GC rich and thus difficult PCR templates, we tested four different DNA polymerases for LAM-PCR. Only one polymerase was able to amplify rDNA sequences in PCR tests on rDNA templates and only that one polymerase was able to amplify provector sequences in test studies. Thus only that one selected polymerase was used in all the studies shown here. Furthermore, to lower potential bias and increase sensitivity of LAM-PCR, two different endonucleases were used in the pilot study to

digest LAM-PCR double-stranded amplicons prior to linker ligation. As the results obtained with both enzymes were very similar, in the subsequent studies only one endonuclease was used.

In the second part of the study when vectors with different size flanks were used, the size of the minimal amplicon required to include genomic DNA flanking the integration site will be different between different vector groups. The 500bp group most probably would require the shortest PCR amplicon in order to obtain DNA sequence outside of the integrated provector, and the 1000bp group would be predicted to require the longest PCR product in order to be informative. This increases the chances of non-informative PCR clones that could be caused by premature PCR termination on GC rich rDNA sequences in 750 and 1000bp rDNA vectors.

Furthermore, as shown in Figure 3.1e, due to the fact that AAV often integrates as concatemers, only amplicons that originated at the most outside rAAV genome within a larger concatemer contained genomic sequences and thus were informative, while those that originated within internal genomes amplified H-T or H-H junctions and thus were not informative.

The vectors used in the NIH3T3 cells did not contain the same enhancer/promoter region as the hF.IX vectors, and thus a new biotinylated oligo had to be used. As described earlier, we did not want to bias the analysis by placing the oligo in the ITRs or in the flanking regions closer to the 5’ or 3’ ends of the vector. Furthermore, in order to avoid potential bias we wanted to avoid using two different oligos for the rDNA and stuffer vectors, and thus it was important to select a region that is common to vectors, rAAV-rDNA and rAAV-stuffer. As it was of interest to place the oligo used for linear

primer extension step of LAM-PCR as close to the end of the vector as possible, we concentrated on hPGK and hF.IX sequences instead of the eGFP coding region. An oligo placed in hPGK produced a lot of non-specific background (data not shown) and thus hF.IX was chosen as the region for biotinylated oligo for the LAM-PCR primer extension step (Figure 3.1e).

V 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(Inagaki et al., 2007; Miller et al., 2005; Nakai et al., 2003), an observation that might be related to the fact that rAAV vectors seem to uncoat and locate near the nucleoli(Johnson and Samulski, 2009). The nucleoli form in discrete locations determined by the position of nucleolus organizer regions (NORs)(Berraondo et al., 2005), a chromosomal loci where multiple copies of the rRNA genes are located in tandemly repeated head-to-tail units and span millions of basepairs (Ritossa and Spiegelman, 1965; Wallace and Birnstiel, 1966). This not only further increasing the chance of targeted integration but, due to the redundant nature of the rDNA loci, also possibly lowering the risk of unintended toxicity related to disruption of some of the rRNA genes (Johansen et al., 2007; Kobayashi, 2006). 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 (Kobayashi et al., 1998). 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 (McEachern et al., 2006). 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 hF.IX 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 are stable.

When looking at the efficiency of rAAV-rDNA vector integration into rDNA loci (Table 3.1 and Figure 3.3e), it is important to keep in mind that the numbers represent percentages of sequenced clones with the expected *vs.* random integration and not a percentage of cells with targeted integration. In fact, as LAM-PCR is not quantitative, the only correct way to interpret the results is to look at the relative increase in rDNA *vs.* random integrations for a given vector pair. In the *in vivo* studies we observed 8 – 18 fold increase in rDNA specific integration between rAAV-stuffer and rAAV-rDNA vectors. Using our previous data showing that about 3% of rAAV vectors integrate randomly into rDNA (Nakai et al., 2005b), we can calculate that the novel rAAV-rDNA vectors integrate into rDNA at 23-54%. This is in agreement to data generated in Chapter 2, in which we used rAAV-rDNA-FAH vectors in Fah-deficient mice and showed that rAAV-rDNA vectors integrate at rate 10-30times higher than control vectors. Using a plasmid rescue strategy, although the recovery yield was small, we found that integrations into the rDNA locus at a frequency of 33% are consistent with the estimation by LAM-PCR of 23 to 54%.

During the course of our studies we noticed that prior to liver regeneration rAAV-rDNA vectors tend to express at lower levels than rAAV-stuffer control vectors (Figure 3.3b-d). This drop in expression correlated with the length of rDNA flanks, with the vectors containing 1000bp rDNA expressing at the lowest level. Southern blot and TaqMan analyses on 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 3.2a and 3.5). Furthermore, as shown in Figure 3.6b-c and 3.2, all vector genomes were present in the expected conformations, suggesting that rDNA sequences

had no apparent effect on vector uncoating and downstream steps leading to the episomal rAAV forms and/or vector integration. The only unexpected, and still unexplained, observation from Southern blot analysis was the presence of an unspecific 1.6kb band in all samples except negative control when using a single cutter. Furthermore, the band was present in prePH and postPH samples and the size of the band does not change with the increase of vector size, and thus it was not a supercoiled episomal form of the vector.

Most importantly, in each tested vector pair, based on LAM-PCR results, there was a significantly higher percentage of locus specific integration of the rAAV-rDNA vectors when compared to the rAAV-stuffer controls (Table 3.1). The ratio of specific rDNA integrations was the highest for rAAV-rDNA-1000 vectors and represented an 18-fold increase when compared to the rAAV-stuffer-1000 control vector. 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 (Daya and Berns, 2008). 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*. It is also important to mention, that our LAM-PCR analysis was complicated by the specific nature of rDNA sequences, which are repetitive sequences with high GC content, making it an especially difficult PCR template. Out of 4 different polymerases tested, only one was found to amplify DNA from rDNA template and was used in this study. However, it is impossible to exclude the possibility that the choice of polymerase introduced biases when used for LAM-PCR amplification of rAAV-rDNA *vs.* rAAV-stuffer integration sites. We feel confident, however, that should such biases occur, it would apply to all samples. Furthermore, it

would be expected that if any biases existed, it would most probably cause less efficient amplification of the rAAV-rDNA provector sequences and not the rAAV-stuffer, and thus would lead to underrepresentation and underestimation of rAAV-rDNA site-specific integrations. It would also be expected that such biases would be higher with the increase of rDNA flanking regions present in the vector and thus any premature termination caused by longer rDNA sequences would cause increase in the number of non-informative clones between 500bp, 750bp and 1000bp rDNA vectors. As shown in Table 3.1, the percentages of non-informative clones did not change significantly with the increase of rDNA flanking sequences present in the vector (42.5%, 26.2% and 43.7% for the 500bp, 750bp and 1000bp rDNA vectors). It is, however, important to mention that based on sequencing results it was impossible to distinguish between non-informative clones caused by early termination of the polymerase, and non-informative clones caused by amplification of internal sequences (head-to-head and head-to-tail) within larger concatemers. None of the tested polymerases were able to amplify through AAV ITR sequences, leading to termination at or near the ITRs. Thus, if the size of concatemers differed between tested rAAV-rDNA vectors, with smaller concatemers in the 1000bp group, it would still be possible that provectors with longer rDNA sequences could lead to a higher percentage of early terminations solely based on the length of rDNA sequences. The Southern blot analysis of vector genomes did not reveal any significant decrease in concatemer size with the increase of rDNA sequences present in the vector, further supporting the conclusion that longer rDNA sequences did not introduce any biases in LAM-PCR analysis. On the other hand, surprisingly, when looking at percentage of non-informative clones between rAAV-rDNA and rAAV-stuffer for each

group, there is a noticeable decrease of non-informative clones for the rAAV-rDNA vectors when compared to corresponding stuffer vectors. This could indicate that rAAV-rDNA vectors tend to form smaller concatemers despite such effect not being apparent from Southern blot analysis. This could also provide further prove to support the observation that higher percentage of rAAV-rDNA genomes undergo targeted recombination, known to utilize monomeric forms rather than larger contatamers (Russell et al., 1994) (see also below).

When comparing our results from *in vivo* and *in vitro* studies (compare Table 3.1 to table in Figure 3.7e) it becomes apparent that the effects the rDNA flanking sequences had on locus specific integration were more apparent in the *in vivo* studies. Since the *in vitro* studies were performed in mouse cells, we can exclude any species-specific differences, but not cell type specific differences (hepatocytes *vs*. fibroblasts). However, there are other factors that require further consideration. Presence of different transgene cassettes in the *in vivo* and *in vitro* studies could have affected recombination efficiency, as previously reported(Hirata et al., 2002), as well as the amount of time from transduction to cell division was significantly different in those two studies (86 days *in vivo vs*. 72 hrs *in vitro*). The last major difference was the AAV serotype used to pseudotype the vectors. AAV8 was used in the *in vivo* studies, while AAVDJ (Grimm et al., 2008) was used *in vitro* due to its enhanced transduction profile on tissue culture cells. Extensive work from the Russell group indicates that only the single-stranded rAAV genomes participate in homologous recombination (Hendrie et al., 2003; Hirata and Russell, 2000; Russell et al., 1994). The use of different AAV serotypes could lead to

differences in kinetics of AAV uncoating and conversion of the ssDNA genome to dsDNA(High et al., 2005).

The exact mechanism of rAAV-rDNA vectors integration still needs to be investigated. It 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. Despite all the unanswered questions and the need to further optimize transduction conditions in rapidly dividing cells, our data clearly show that the novel rAAV-rDNA vector breach the gap between rAAV and lentivirala vectors and open a door for the use of rAAV vectors in context of dividing cells *in vivo* and *in vitro/ex vivo*.

.

Chapter 4

Large-scale integration sites analysis of AAV-rDNA vectors

Zhongya Wang, Hirayuki Nakai, Carl Pelz, Markus Grompe

Manuscript in preparation.

I. Abstract

AAV-rDNA vectors have been confirmed as integrating vectors with a preference for rDNA repeats. Despite their rDNA preference for integration, other preferential integration sites remain unknown. Here we used a shuttle vector strategy and LAM-PCR based next generation sequencing (NGS) to analyze the integration sites in a large-scale manner. Overall we identified 12 integration sites. Three out of 12 (25 %) are within homologous rDNA fragments in the rDNA repeats and one integration site (1/12) is within 2 kb of the homologous fragment. In total the frequency of 4/12 (30 %) of integrations within rDNA repeats is about 100 times higher than random integration (0.3 %) based on the size fraction. Compared to previous results concerning regular AAV vector integration, there was 10-fold enrichment in rDNA repeats, which was consistent with functional results observed in treated mice. Large-scale analysis of integration sites by NGS showed that tag density in rDNA repeats was about 100 times more than other chromosomal regions, which confirmed the findings derived from plasmid-rescued integration sites. The large-scale integration site analysis also confirmed that the AAV-rDNA integration pattern resembles that of regular AAV vectors, except for the enriched rDNA preference. Both vectors demonstrated significant preference for transcription start sites, CpG islands and RefSeq genes. Large-scale analysis of integration sites also revealed that AAV-rDNA vectors had no preference for tumor related genes, which indicated that AAV-rDNA vectors were as safe as, if not more safe, than regular AAV vectors.

II. Introduction

Gene therapy for metabolic diseases requires long-term transgene expression, which is often provided by integrating viral vectors. Integration into host genomes is a double-edged sword, which can give consistent transgene expression while causing insertional mutagenesis and tumorgenesis, which has been confirmed in X-linked severe combined immune deficiency patients. These patients were treated with a retroviral vector and eventually developed leukemia due to insertional mutagenesis near the LIM domain only 2 (*LMO2*) oncogene (Hacein-Bey-Abina et al., 2003b).

Large-scale analysis of integration sites will shed the light on insertional mutagenesis of viral vectors, as well as providing integration preference in the host genome. For example, the preferential integration sites of *murine leukemia viruses* (MLV) are close to the start of transcriptional units, while *human immunodeficiency virus* (*HIV*) preferentially integrates within the transcriptional region of active genes, especially activated genes after HIV infection (Mitchell et al., 2004; Schroder et al., 2002; Wu et al., 2003). These data also suggested that different viral vectors have different integration patterns. AAV vectors have been considered one of the most promising viral gene therapy vectors. One of the most appealing features is its capability of integrating into the host genome (McCarty et al., 2004). AAV integration has been one of the central interests in AAV gene therapy. Although previous studies showed that the frequency of rAAV integration into the host genome *in vivo* is low (Chen et al., 2005; Miao et al., 1998; Nakai et al., 2001; Schnepp et al., 2003), integration site identification remains

necessary for AAV gene therapy. Several studies have been conducted to dissect AAV integration preference in mouse liver and in human cell lines, which showed that active transcription regions were the preferential integration sites of AAV and that rDNA repeats were hot spots (Miller et al., 2005; Nakai et al., 2003; Nakai et al., 2005b). Large-scale analysis of integration sites not only provided information on AAV integration preference, which might highlight the AAV integration mechanism, but also helped to determine the prospects of future AAV clinical trials.

Rescue strategy

In order to identify the integration sites used by the new AAV-rDNA vector for comparison with previous-generation vectors, a plasmid rescue strategy was employed. The rescue strategy is to rescue a plasmid replicon (*ori*) bearing fragment from the junction of the insertion within the flanking genomic DNA. This strategy was used to clone genomic DNA sequence adjacent to integrating vectors in *Dictyostelium* (Hildebrandt and Nellen, 1991; Maniak and Nellen, 1989; McMahon et al., 1996). Rescue strategy has been successfully used to identify the AAV integration sites in mouse liver and human cell lines (Miller et al., 2005; Nakai et al., 2003; Nakai et al., 2005b), which provided insightful information for chromosomal effects on AAV integration, and AAV integration pattern. Here in this dissertation I have applied rescue strategy to identify AAV-rDNA integration sites.

Next generation sequencing (NGS)

The automated Sanger sequencing method, which dominated DNA sequencing for many years, is considered a 'first-generation' sequencing technology (Hutchison, 2007). Newer methods of high-throughput DNA sequencing, are generally grouped as next generation sequencing (NGS). Compared with automated Sanger sequencing, NGS platforms have significantly elevated throughput and substantially lower reagent costs (Metzker, 2009). NGS technologies have changed the scientific approaches in basic, applied and clinical research. For example, in transcriptome studies, sequencing based methods are now replacing microarrays, as sequencing based methods can identify and quantify unknown transcripts, even at a low level and can provide splicing and sequence variation information (Wang et al., 2009; Wold and Myers, 2008). The capability of sequencing whole genomes provides the possibility of large-scale comparative and evolutionary studies. Next generation sequencing has also been applied to identifying the viral integration sites (Duncavage et al., 2011). For example, NGS has been successfully used to identify the HIV virus integration sites in host cells (Ciuffi and Barr, 2010; Lefebvre et al., 2011). The application of NGS in HIV studies provided important information concerning HIV integration compared to regular LAM-PCR (Linear amplification-mediated PCR) and LM-PCR (ligation-mediated PCR) methods used previously (Hayakawa et al., 2009; Hooker et al., 2009; Podsakoff et al., 2005). In this dissertation, I have designed a new method to take advantage of NGS and LAM-PCR to sequence the integration sites of AAV-rDNA vectors.

III. Materials and methods

Shuttle Vector Strategy

rAAV vector production and animal handling

All animal experiments were performed according to an approved IACUC protocol for animal care at OHSU.

The plasmid pAAV-rDNA-I-*Sce*I-KO was constructed as following: *Afl*II (CTTAAG), I-*Sce*I (TAGGGATAACAGGGTAAT) and *Bam*HI (GGATCC) recognition sites were introduced into the shuttle cassette by forward primer: CTAA*CTTAAGTAGGGATAACAGGGTAATGGATCC*ATTGCACAAGATAAAA ATATATCATCATGA; *Afl*II was introduced into the shuttle cassette by reverse primer: TCTT*CTTAAG*TCGGTCCGCACCATGTGAGCAAAAGGCCAG. The shuttle cassette was amplified from plasmid pNKan with the two primers above. The shuttle cassette was subcloned into the *Afl*II site of pAAV-rDNA to make pAAV-rDNA-I-*Sce*I-KO. AAV-rDNA-I-*Sce*I-KO serotype 8 was produced and purified in the Nakai lab at the University of Pittsburgh (Inagaki et al., 2007; Nakai et al., 2005a)

AAV8-rDNA-I-*Sce*I-KO was injected into wild type mice at a dose of $3x10^{11}$ vg/mouse. These mice were killed four weeks after AAV injection and fresh livers were immediately frozen in liquid nitrogen. Genomic DNA was extracted from liver as previously described in Chapter 2.

Generation of rAAV8 integration libraries

20 ug of high molecular weight liver gDNA was digested with I-*Sce*I, which recognizes the TAGGGATAACAGGGTAAT sequence in the shuttle vector, to separate concatemers, and was followed by treatment with calf intestinal phosphatase (CIP; New England Biolabs, Beverly, Mass.) at 50 °C for 1 h to prevent ligation of free DNA ends, including the I-*Sce*I ends. After the CIP treatment, the enzyme was heat inactivated at 65 °C for 1 h. After heat inactivation, the DNA was extracted by phenol-chloroform and precipitated by ethanol, followed by *Bam*HI/*Bgl*II digestion. *Bgl*II generates compatible ends and does not cut the AAV-rDNA-I*Sce*-I-KO genome. Completely digested genomic DNA was then separated in agarose gels overnight to confirm the digestion and achieve best separation efficiency. Then the digested liver gDNA was recovered and self-ligated by T4 ligase (Roche, Indianapolis, IN). The self-ligated product was used to transform *Escherichia coli* (ElectroMAX DH10B; Invitrogen, Carlsbad, Calif.) with 2.5 to 5 ug of ligated products to make libraries. DH10B transformed with plasmid libraries were plated on Luria-Bertani (LB) agar plates containing kanamycin (50 μg/ml).

Integration sites screening and sequencing

As can be seen in Figure 4.1, when *Bam*H I digest AAV8-rDNA-I-*Sce*I-KO infected liver gDNA, there will be self-ligation from *Bam*HI digested head to tail concatemers. Primers: TGAACAGCCTCTGGCATG TTGGAA and TCGCACCTGATTGCCCGA CATTAT were used to identify concatemers formed by self-ligation products. All positive PCRs of the right size indicate the rescued plasmids were from the self-ligated concatemers with no integration site information. Sequences of screened shuttle vector bearing host genome information were determined with a 3730x

DNA analyzer (Life Technology, formerly Applied Biosystems, Carlsbad, California) using the CCTTCGATGTCGGCTCTTCCTAT primer.

Mapping of integration sites

After obtaining the sequence, we isolated the proviral cellular junction sequence, and then aligned the junction sequence against the public mouse genome database (July 2007, NCBI37/*mm9*) with both the University of California - Santa Cruz (UCSC) BLAT program and the National Center for Biotechnology Information (NCBI) BLAST program. Sequence matches were only considered to be authentic if a sequence match extended over the length of the high-quality sequence with >95 % identity. For rDNA sequence we mapped the virus cellular junction against the rRNA gene repeats when there were no mouse genomic regions matched with high identity.

Determination of breakpoints of the vector genome and chromosomal breakpoint

The breakpoint was determined as the nucleotide positions where rAAV2 vector and mouse genomes were rejoined and recombined. When there were microhomologies around the recombination sites, we took a nucleotide position in microhomologies away from the vector genome as the breakpoint. For chromosomal break points, if there were microhomologies around the recombination sites, we took a nucleotide position in microhomologies away from chromosomal sequence.

Bioinformatics

We defined a genomic region between transcriptional start and stop boundaries of a RefSeq gene as a "gene," as previously reported (Nakai et al., 2003; Wu et al., 2003). To identify cancer related genes, we searched several databases, including the Tumor Gene Database (http://www.tumor-gene.org/TGDB/tgdb.html), and the Retrovirus Tagged Cancer Gene Database (http://rtcgd.ncifcrf.gov/). Any gene matching one of these two databases was defined as an oncogene.

Statistical analysis

In all cases statistical significance was determined using the χ^2 test to compare AAV-rDNA vector integration site frequencies with those of randomly generated rDNA repeats hits in the genome, or with the published rDNA repeat integration frequency of regular AAV vectors (Nakai et al., 2005b). The size of a haploid mouse genome is $3x10^9$ bp and considering 200 copies of 45.3-kb rDNA repeats form a target of $9 x10^6$ bp in length, the expected probability of an integration landing in the rRNA gene repeats by a random integration model is $(9 x10^6)/(3 x10^9) = 0.003$. *P* values were determined using tables, and those less than 0.01 were considered significant.

NGS sequencing

Primer extension

Four 5'-biotinylated primers against the Fah expression cassette were designed.

RSVR1: Bio-CCGCTTTTCGCCTAAACACAC;

RSVR2: Bio- CGGAACGCTTCGTCGATCTG;

FahL1: Bio-AAAACTTCGGCTCCATGTTG;

FahL2: Bio-CTGGAAACAAAGGGCTCAAG

Single primer extension reactions were performed using the following program with 2 μg of genomic DNA from AAV-rDNA-FAH injected liver as template. 95 ^{o}C x 5' ; (95 ^{o}C x15'', 55 ^{o}C x 20'', 72 ^{o}C x2') x100; 72 ^{o}C x 20'.

The KOD DNA Polymerase (EMD, Billerica, MA) was used according to the manufacturer's protocol.

The primer extension reaction products were incubated with streptavidin-coupled magnetic beads (Life Technology, formerly Invitrogen, Grand Island, NY) in a binding solution at room temperature (18–25 °C) on a horizontal shaker at 300 r.p.m. overnight. Beads with attached single strand DNA were collected on an MPC (Life Technology, formerly Invitrogen, Grand Island, NY) and the supernatant was discarded. The collected beads were washed three times in washing buffer (Schmidt et al., 2007). After washing, beads were suspended in 50 μl of hexanucleotide priming mixture (Roche, Indianapolis, IN) and the reactions were incubated for 1 h at 37 ^{o}C to synthesize the second strand DNA. Then the beads were collected with an MPC, washed as above and re-suspended in 100 μl TE buffer. The DNA beads samples were sonicated 1x 70 % amplitude for 10 seconds with an Ultrasonic Generator System **(**Cole-Parmer Vernon Hills, IL**).** The samples were purified on a Qiagen PCR column (Qiagen, Valencia, CA) and treated for blunt-ends using DNA Terminator Kit (Lucigen, Middleton, WI) at room temp for 30 min. Blunt-ended DNA samples were purified with AMPure xp beads (Beckman Coulter Genomics**, Danvers, MA**). After TE elution, DNA was treated with Exo-minus Klenow DNA pol (Epicentre, Madison, WI) to add an A base to the 3' end and purified with AMPure xp beads. After purification, Solexa Genomic Adapters (Illumia, San Diego, CA)

were ligated to the A bearing DNA fragments end by T4 DNA ligase (Roche, Indianapolis, IN). The ligation product was purified by AMPure xp beads and amplified with Solexa Genomic PCR Primer (Illumia, San Diego, CA) for 10-20 cycles to generate a limitedly amplified DNA library. The quality of DNA library was checked and samples were sent to OHSU Illumia sequence core.

NGS Bioinformatics

Sequences (25 bp) derived from Solexa sequencing were collected and libraries with different primers were generated for further analysis. Bowtie program (Langmead et al., 2009) was used to map all sequences to the AAV-rDNA-FAH genome sequence, since tags within the viral genome do not provide any integration information. All sequences that mapped to AAV-rDNA-FAH were removed. The remaining sequences were mapped to the 45kb rDNA sequence using the Bowtie program (Langmead et al., 2009) to select the tags mapping to rDNA repeats. The rest of the tags were mapped to the mouse genome (UCSC *mm9*) with the Bowtie program. Next, all tags were screened by the peak detection program Findpeaks (Fejes et al., 2008), implanted in C++ with FDR 0.05, and only tags showing within peaks were counted as trusted tags. In total we generated 35361 informative tags. Next, we crossed libraries to screen tags that showed up in at least two libraries to minimize PCR random binding generated non-specificity. Total of 572 tags from peaks within a 2 kb range in at least two libraries were identified. The integration sites information contained tags fasta files on chromosome other than rDNA were submitted to GTSG-*QuickMap* (Gene Therapy Safety Group, http://www.gtsg.org/quickmap.jsp). The results were compared to previous published

integration sites of AAV vectors in livers (Nakai et al., 2003; Nakai et al., 2005b). We used the GTSG randomly simulated 1,000,000 integrations sites as a random integration reference control.

NGS Statistical analysis

In all cases, statistical significance was determined using the x^2 test to compare AAV-rDNA vector integration site frequencies with those of randomly generated genomic positions. *P* values were determined using tables, and those less than 0.01 were considered significant.

IV. Results

AAV-rDNA integration libraries and screening

AAV8-rDNA-I-*Sce*-KO viruses were injected into wild type mouse liver at $3x10^{11}$ vg/mouse via the retro orbital route. Four weeks after injection, mice (n = 4) were sacrificed and the livers were removed to make genomic DNA for the integration library. Because of head to tail concatemers generated through self-ligation, as shown in Figure 4.1, we designed screening primers to be able to identify concatemer-formed self-ligated plasmids. As summarized in Table 4.1, the ratio of concatemer-formed plasmids was around 90 %. In total we screened approximately 4000 clones, and sequenced 420 clones, out of which, we identified 12 integration events from four mice.

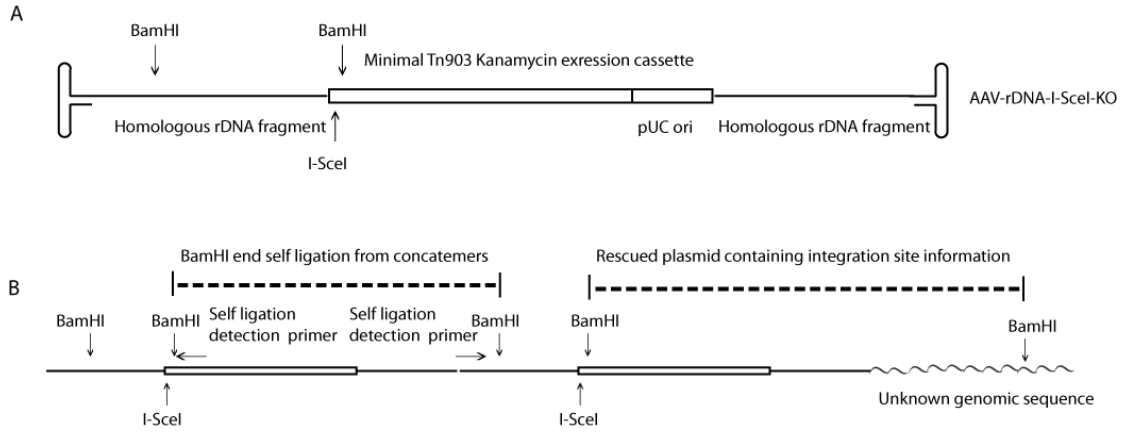


Figure 4.1 Scheme of shuttle vector strategy.

(A) Scheme of AAV-rDNA-I-*Sce*I-KO and enzymes used for rescuing plasmid.

(B) Scheme of concatemer based self-ligation and screening.

AAV-rDNA vectors prefer rDNA repeats as integration regions

After obtaining the sequences, we used BLAST to compare the sequences to the AAV-rDNA-I-*Sce*I-KO sequence and to the mouse genome (NCBI37/*mm9*). We determined eight independent integration events, as shown in Table 4.2, each of which mapped to one unique mouse genome position, with different proviral structures.

Total clones screened	Sequenced clones	Aligned to mouse genome
4000[a]	420[b]	12[c]

Table 4.1 summary of clone screening.

[a] Approximate number. Some of the clones might have been re-screened. [b] This is th total number of clones sent for sequencing. [c] Total number of sequenced results that matched to the mouse genome (*mm9* or rDNA gene) with a BLAST score > 95 %.

For sequences that were not aligned to a specific position in the mouse genome, we aligned them to the pre-45S rRNA genes. In total, we identified four independent integration events in rDNA repeats (Figure 4.2 B). Junction sequences of these integration events (Figure 4.2 C) displayed multiple junction patterns. Integration in 16752 and 17554 showed TC and GC microhomology respectively. Integration in 16702 showed a two nucleotide (CT) insertion, while integration in 19145 showed directly joining without deletion or insertion. The existence of various junction patterns indicated that integration did not occur by homologous recombination. However, the presence of homologous rDNA fragments in the AAV-rDNA vectors did increase integration into rDNA repeats about 10 times (4/12 vs 0.003, $P < 0.0001$). Among all 12 integrations sites,

3/12 (25 %) were intergenic and 5/12 (41.7 %) were within transcriptional units. Excluding the four integration sites, 5/8 (62.5 %) of the integration sites were within RefSeq genes (transcription units), which was significantly higher than the random set (about 36.2 % in randomly simulation, $P < 0.0001$). The observed frequency (62.5 %) of integration sites within RefSeq genes was similar to the previously reported frequency (53.1 % and 59.3 %) of regular AAV (Nakai et al., 2003; Nakai et al., 2005b). Thus, AAV-rDNA had a preference for transcription regions similar to regular AAV vectors. We checked all RefSeq genes identified as integration sites with the Tumor Gene Family Data Base and the Retrovirus Tagged Cancer Gene Database, and none of them were listed as oncogenes and tumor suppressor.

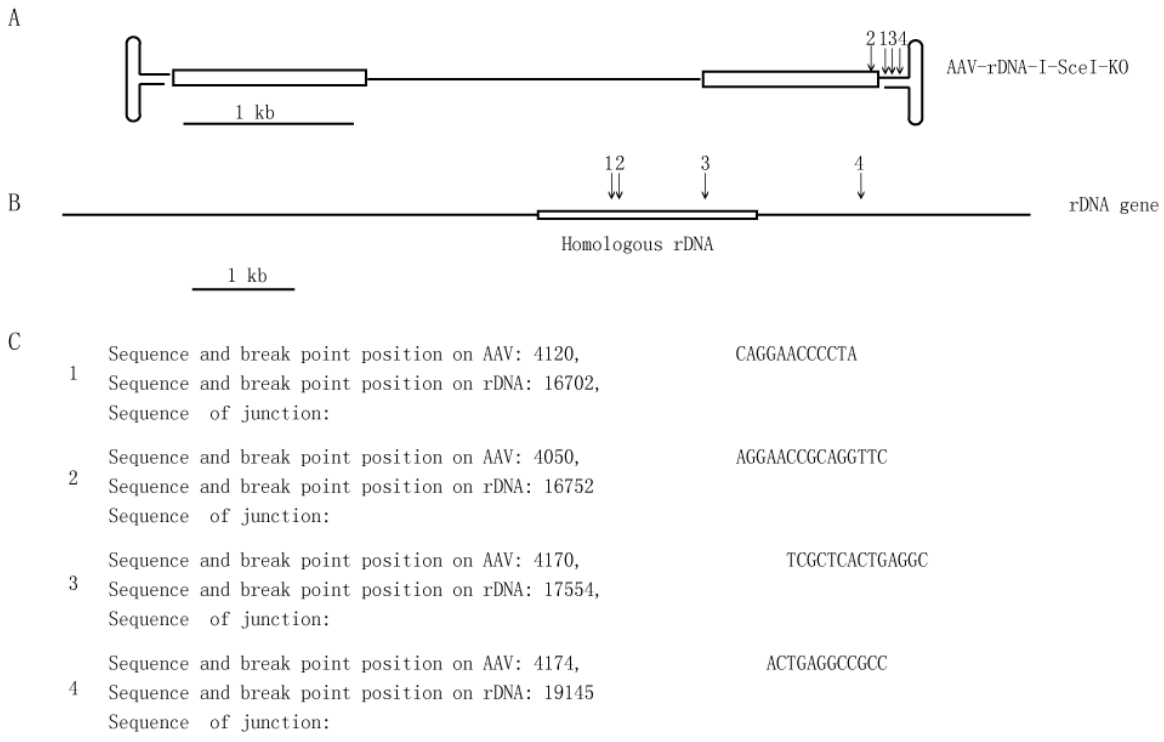


Figure 4.2. Analysis of AAV-rDNA integration sites in rDNA region.

(A) Scheme of breakpoint of integration on AAV-rDNA vector. The AAV-rDNA vector genome is 4354 nucleotides in length. The terminal 145 nucleotides (positions 1-145 and 4110-4254) represent inverted terminal repeats (ITRs). **(B)** Scheme of breakpoint of integration on rDNA region and distance to homology fragment (position on rDNA: 13412-15451) of AAV-rDNA fragment. Position is based on rDNA sequence (GenBank: X82564.1). **(C)** Junction sites sequence alignment with AAV-rDNA vector and rDNA locus.

AAV-rDNA vectors integration sites	Chromosome[a]	DNA strand	Location	Refseq gene	Gene function	Tumor gene[b]
1	6	-	121806565	Mug1	Endopeptidase inhibitor activity	No
2	8	+	42212855	Mtus1	Molecular function unknown	No[c]
3	9	+	59322717	Arih1	Ubiquitin-protein ligase activity	No
4	11	-	7113263	Igfbp3	Insulin-like growth factor binding	No
5	13	-	50529615	Intergenic	NA	NA
6	15	-	3438032	Ghr	Growth factor binding	No
7	x	+	139531170	Intergenic	NA	NA
8	x	-	113127425	Intergenic	NA	NA
9	NA	-	16702[d]	rDNA	NA	NA
10	NA	-	16752[d]	rDNA	NA	NA
11	NA	-	17554[d]	rDNA	NA	NA
12	NA	-	19145[d]	rDNA	NA	NA

Table 4.2 integration sites identified.

[a] rDNA repeats are spread over several chromosomes, and identification of the exact rDNA gene where integration occurred was not attempted. Therefore these are labeled as NA for rDNA integration. [b] For tumor genes, we used the Tumor Gene Family Data Base

(http://www.tumor-gene.org/TGDB/tgdb.html) and the Retrovirus Tagged Cancer Gene Database (http://rtcgd.ncifcrf.gov/). [c] Mtus1, although named mitochondrial tumor suppressor 1, it is not listed in the Tumor Gene Family Data Base or the Retrovirus Tagged Cancer Gene Database, Accordingly, it is listed as not a "tumor gene". [d] Positions on rDNA are based on *M.musculus* 45S pre rRNA gene (GenBank: X82564.1)

NGS generated integrations sites confirmed rDNA preference of AAV-rDNA vectors

To analyze the integration preference of AAV-rDNA vectors in a large-scale manner, we combined linear amplification-mediated PCR (LAM-PCR) (Schmidt et al., 2007) and next generation sequencing (Metzker, 2009; Wold and Myers, 2008). Firstly we used four biotinylated primers targeting the FAH expression cassette (two pointing in each direction) to perform primer extension on genomic DNA from AAV-rDNA-FAH infected livers. Then we collected the primer extension products by bead pull-down and performed second strand DNA synthesis. After that we followed the standard Chip-Seq protocol (Illumina/*Solexa* technology, Illumina, San Diego, CA) to generate four libraries for each primer used. We found common tags within peaks using the program Findpeaks in the C++ environment (Fejes et al., 2008). In total we identified 35361 tags in chromosome regions other than rDNA repeats.

	rDNA repeats	rDNA transcription unit (first 13.7 kb)	rDNA intergenic region (30 kb)	Chromosome region
Total tags	17544	4093	13451	35361
Short repeats[a]	3100	0	3100	0
Short repeats[b]	1908	0	1908	0
Short repeats[c]	1768	0	1768	0
Copy number	200	200	200	1
Nucleotides	45.3 kb[d]	11.7 kb[e]	31.6 kb	$3x10^9$ bp[f]
Tag density (tag/kb)	1.19[g]	1.75[h]	1.06	0.0118
Enrichment	100	142	89.6	1[i]

Table 4.3 Tag density in rDNA repeats and other mouse genome.

Tags in rDNA represent all tags with short repeats (which might map to multiple locations in the mouse genome) manually subtracted.

[a] $(TG)^n$ repeats. [b] $(CT)^n$ repeats. [c] $(AC)^n$ repeats. [d] rDNA repeats represent full length rDNA that does not exclude the 2 kb homologous sequence, although all tags match this region have been filtered off. [e] The transcription unit is 13.7 kb, because of the interruption of the 2 kb homologous fragment, we only counted the first 11.7kb as the transcription region. *f* The actual haploid mouse genome actual size is around 2.7 gb, here we used 3 gb to facilitate comparison with previous publications (Nakai et al., 2005b). [g] The tag density represented the first 10 kb of the transcription region. [h] the tag density represented each rDNA repeat ($9.0x10^6$ in total) other than one rDNA repeat (45 kb). [i] The enrichment for chromosomal region has been arbitrarily set as one.

To measure the rDNA preference for integration, we used the tag density to compare rDNA repeats and other chromosomal regions. As shown in Table 4.3, the tag density in rDNA repeats was 100 times more overall than other chromosomal regions. Comparing to 10 times (3 % in rDNA *vs.* 0.3 % randomly, $P < 0.0001$) enrichment in rDNA repeats for regular AAV vector (Nakai et al., 2005b), the preference of AAV-rDNA vectors in rDNA repeats was enriched 100-folds higher, which was about 10 times more than regular AAV vectors. This also matched the functional data (10-30 times better than regular AAV) and plasmid rescue data (30 % *vs.* 3 %). Here we should note that since all tags that matched the AAV-rDNA-FAH sequence were excluded, the integrations within homologous fragments were unable to be identified.

Integration preference of AAV-rDNA vector over gene regulatory sequences and RefSeq genes

Previously it has been reported that AAV2 vectors preferentially integrate into active transcription regions of the mouse genome (Nakai et al., 2003). To determine the association of AAV-rDNA integration sites with transcription regions or regulatory regions, we firstly searched different libraries to find the common tags within same tag peaks that Findpeaks generated. In total, we identified 572 common tags in at least two libraries and submitted a Fasta file containing the 572 integration sites to GTSG-*QuickMap to perform the analysis.* These analyses revealed a total of 396 integration sites that matched to unique chromosomal locations with over a 95 % Blast score.

Percentage (%)	**AAV-rDNA –FAH vector**	**AAV-vector** ***(Nakai, et.al. JVI, 2005)***	**AAV-vector** ***(Nakai, et.al. Nature Genetics, 2003)***	**Random Set**[c]	**P value**[d]
Within RefSeq genes	52.3	53.1	59.3	36.2	<0.0001
<5kb of RefSeq genes	12.6	25.8	7.4	7.6	<0.0001
Near TSS[a] (< 5 kb)	21.8	27.3	25.9	0.6	<0.0001
CpG island	3.0	24.9	7.4	0.4	<0.0001
< 5 kb CpG island	20.8	49.3	19.6	5.2	<0.0001
Oncogene related[b]	2.3	3.5		1.4	0.2482[f]
Total integrations	**396**	**337**	**29**	**1,000,000**	

Table 4.4. AAV-rDNA integration pattern and comparison of AAV-rDNA integration with regular AAV vectors.

[a] Transcription start site. [b] Defined as any gene found in the Tumor Gene Family Data Base or Retrovirus Tagged Cancer Gene Database. [c] Program randomly generated 1,000,000 position on the whole mouse genome.

[d] All the statistical significance P were calculated between AAV-rDNA-FAH vector and the random set with Chi squared two-tailed P value. [f] Considering the RefSeq gene enrichment of AAV-rDNA vectors, no increased risk of oncogene integration was observed.

In this dissertation, I compared the observed frequency of integration regions of interest with 1,000,000 program-simulated and randomly generated chromosome locations to see which regions are preferable integration hot spots (Table 4.4). The first question we wanted to know was whether AAV-rDNA vector prefers RefSeq genes. As shown in Table 4.4, AAV-rDNA integrated into RefSeq genes with a frequency of

52.3 %, which was considered significantly ($P < 0.0001$) higher than the random set (36.2 %). Notably, the frequency was similar to previously identified AAV vectors in hepatocytes (53.1 % and 59.1 %, respectively) (Nakai et al., 2003; Nakai et al., 2005b), Next, we wanted to see if AAV-rDNA vectors prefer regulatory sequences (*i.e.* near transcription start sites or, in CpG islands) as regular AAV vectors do. As shown in Table 4.4, the frequency of integration within 5 kb of transcription start sites was 21.8 %, surprisingly about 400 times higher ($P < 0.0001$) than the random set (0.6%) but similar to the values of 25.3 % and 25.9 % seen for regular AAV vectors(Nakai et al., 2003; Nakai et al., 2005b). For CpG islands, we also found eight-fold enrichment compared to the random set (3 % *vs*. 0.4 %, $P < 0.0001$). However, if we compared to previously findings (24.9 % and 7.4 % in CpG islands, respectively), the enrichment is lower than regular AAV vectors. This might be due to technical differences in site assessment. Since NGS is based on PCR, the PCR for GC rich regions, like CpG islands, is notoriously hard, while there is no or a reduced effect on rescue strategy. This is partially proved as within 5 kb of CpG islands, the frequency is 20.8 %, 49.3 % and 19.6% for AAV-rDNA, AAV vector large-scale and AAV vector *in vivo (Nakai et al., 2003; Nakai et al., 2005b).* About two-fold difference (20.8 % *vs.* 49.3 %) in 5 kb adjacent to CpG island was less than eight-fold difference in CpG islands (3 % *vs*. 24.9 %), and actually it was almost the same for AAV-rDNA and AAV vectors *in vivo* (20.8 % *vs.*19.6 %) (Nakai et al., 2003). Comparing the 5.2 % frequency within 5 kb of CpG islands in the random set, there was almost four times enrichment for AAV-rDNA vectors ($P < 0.0001$).

With a strong preference for integration in CpG islands and in adjacent TSS, which are the main transcription regulatory features, AAV-rDNA vectors have demonstrated a

similar integration pattern to regular AAV vectors. The preference for active transcription regions might be due to higher frequency of double-strand breaks or lesions present during transcription, or active transcription provides easier accession to host genomes than the non-active regions.

Another important issue is whether AAV-rDNA vectors have a preference for tumor related genes. The frequency at which AAV-rDNA vectors hit tumor related genes was 2.3 %, which was not significantly higher than the 1.4 % frequency for the random set (Table 4.4). Considering the preference of integration for RefSeq genes (52.3 % *vs.* 36.2 %, $P < 0.0001$), AAV-rDNA vectors had negligibly increased (1.1 times, 2.3 out of 52.3 *vs.* 1.4 out of 36.2) preference for tumor related genes. Notably, as for the frequency of integration within tumor related genes, AAV-rDNA was slightly lower (2.3 % *vs.* 3.4 %) than regular AAV vectors with similar preference for RefSeq genes (52.3 % *vs.* 53.1 %). All of these suggest that AAV-rDNA is a safe gene therapy vector as are regular AAV vectors.

Preference of AAV-rDNA vectors for integration into rDNA transcribed regions

To elucidate the AAV-rDNA integration pattern in rDNA, we mapped all the tags to rDNA repeats. We can see that from Figure 4.3, the first 13.7 kb (transcription region) had a higher tag density than the intergenic spacer. Because of the interruption of 2 kb homologous sequence in the rDNA in the first 13.7 kb, we calculated the first 11.7 kb tags (tags/kb) density, as shown in Table 4.3: 1.72 tags/kb, which was about 1.6 times more than other rDNA regions (1.06 tags/kb), which suggested that AAV-rDNA vectors prefer rDNA genes over the intergenic spacer regions ($P < 0.0001$). This feature was

consistent with regular AAV vectors (Miller et al., 2005; Nakai et al., 2005b), which showed that AAV vectors prefer transcription regions.

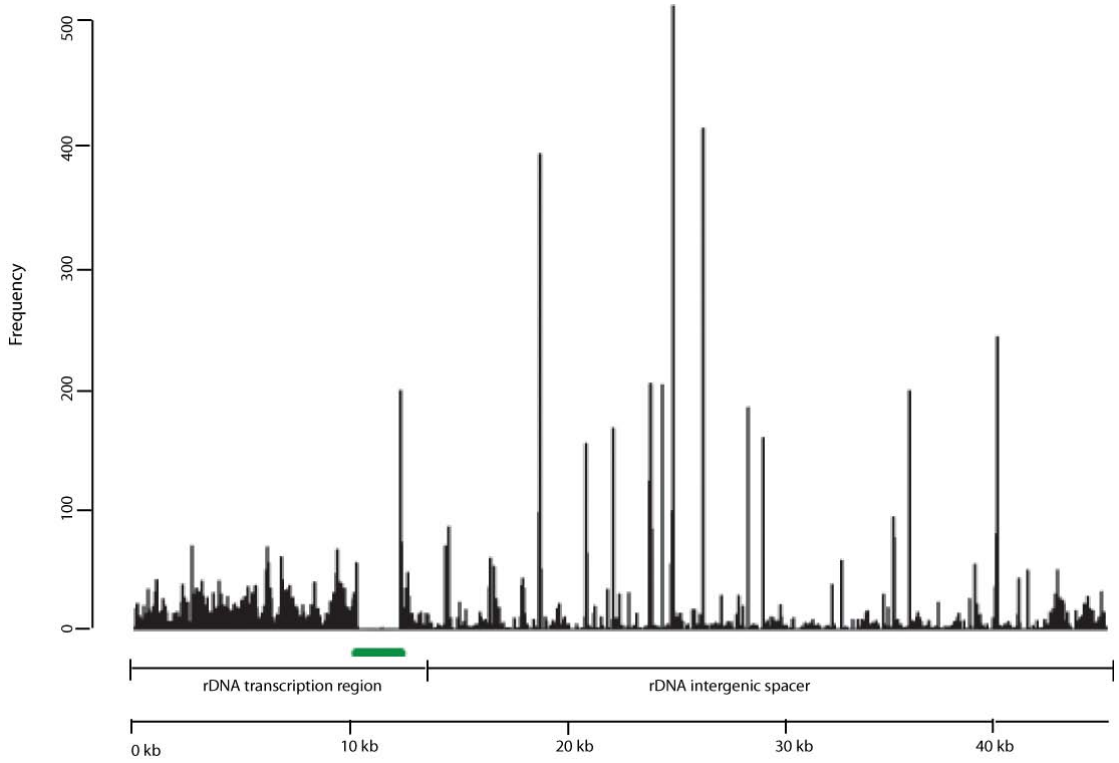


Figure 4.3. Integration distribution on rDNA repeats.

Tag distribution on rDNA repeats. The green line represents the homologous rDNA fragment. The Y axis stands for tag frequency at each location of rDNA repeats. The 13.3 kb transcribed region is labeled, which includes the 18S, 5.8S, and 23S rRNAs. The 29.7-kb intergenic spacer is depicted as a single bold line as labeled.

V. Discussion

AAV are one of the most promising gene therapy vectors (Mingozzi and High, 2011). There have been successful clinical trials for diseases such as Leber's congenital amaurosis and hemophilia B (Maguire et al., 2008; Simonelli et al., 2009). In this report we have used an AAV shuttle vector system and PCR based next generation sequencing (NGS) to identify integration sites. This allowed us to compare the results from different methods to see if they provide unbiased results. For the shuttle vector strategy, there are several advantages. Firstly, the size and quality of junction sequences is not limited by PCR efficiency, this is particularly important for AAV-rDNA vectors, because of the two homologous arms, it requires PCR amplification of adjacent gDNA regions of homologous rDNA sequence to get integration site information. The rescued long and accurate sequence allows us to map the genome with more specificity. Secondly, the shuttle vector strategy is able to provide both ends of the integrations sites, although we didn't address this feature in this dissertation. For PCR-generated next generation sequencing, there are several advantages. Firstly, it is a high throughput method, which can provide a massive amount of integration information. Secondly, introduction of sonication will minimize the bias brought by genomic distribution of restriction enzymes, which is one of the drawbacks of the shuttle rescue strategy. However, the disadvantage of PCR based NGS is that the accuracy of results depends on PCR specificity, which depends highly on the primer binding specificity, annealing temperature and template complexity. The majority of these factors are driven by random (non-specific) primer template binding. In this study, to maximally eliminate the non-specific binding issue, we designed four primers pointing in two directions and only evaluate integrations showing

up in at least in two independently primer (PCR) generated libraries. The basis for this design is that the chance of two different primers binding to same position randomly is extremely low. The primers are about 20 bp length and the total mouse genome is $3x10^9$ bp. In a random binding model, the chance of two primers binding to one specific location is $20/3x10^9$, which is approximately zero. The similar results obtained by the shuttle vector strategy and PCR based NGS results supported our theory.

In total we identified 12 integration sites by rescue strategy from approximately 4000 clones at the start. The ratio of positive rescue was significantly lower than we expected and than previously published results (Miller et al., 2005; Nakai et al., 2003; Nakai et al., 2005b). The reason for this is, as shown in Figure 4.1B, that the presence of a *Bam*HI site in the rDNA homology of the vector allowed head to tail concatemers formed by two *Bam*HI ends to ligate together to form self-replicatable plasmids. As we showed in Figure 2.6 of Chapter 2, concatemeric integrations largely exist not only in primary livers but also in transplanted livers, which suggest that even in the integrated virus genome, there are still head to tail concatemers, although the copy number might be lower than in primary livers as shown in Figure 2.6. The possible resolution for this might be to construct *Bam*HI disrupted AAV-rDNA rescue vectors. However, whether the disruption of *Bam*HI site will change the integration frequency and integration pattern of AAV-rDNA vectors still remains unknown. As previously reported, point mutations breaking the homologous sequence length has a significant effect on the homologous recombination rate (Bollag et al., 1989). Although the mechanism of AAV-rDNA vectors mediated integration still remains to be elucidated, the homology of rDNA sequence seems important for efficient integration.

With the NGS identified integration sites, we have found that AAV-rDNA has a very similar integration pattern as reported previously for regular AAV vectors (Nakai et al., 2003; Nakai et al., 2005b) except for two things. The first is AAV-rDNA vectors have about 10-folds higher preference for rDNA repeats integration. The second is that AAV-rDNA vectors have a much lower integration frequency in CpG islands than regular AAV vectors (3.0 % *vs.* 24.9 % as in Table 4.4). Interestingly, in human cell lines, Miller *et.al.* reported a 4.3 % integration frequency within CpG islands by the plasmid rescue method as Nakai *et. al.* used (Miller et al., 2005; Nakai et al., 2005b), which is very similar to our PCR based NGS method generated integration frequency. Also, this might be a method-generated difference. Since NGS is based on PCR, and amplification of GC rich regions is challenging, such sites are likely underreported relative to the rescue strategy for AAV vectors. This is partially proved as within 5 kb of CpG islands, the frequency is 20.8 %, 49.3 % and 19.6 % for AAV-rDNA, AAV vector large-scale and AAV vector *in vivo*, respectively *(Nakai et al., 2003; Nakai et al., 2005b).* The two-fold difference (20.8 % *vs.* 49.3 %) in 5 kb adjacent to CpG island is less than in CpG islands (8 times, 3 % *vs*. 24.9 %), although actually it is almost the same for AAV-rDNA and AAV vector *in vivo* (20.8 % *vs.*19.6 %). Compared to the 5.2 % frequency within 5 kb of CpG islands in the random set, there is an almost four-fold enrichment for AAV-rDNA vectors.

Both the shuttle vector strategy and PCR-based NGS did not show that AAV-rDNA vectors had any preference for tumor related genes, as for AAV regular vectors, which once again prove that AAV-rDNA is a safe vector for gene therapy.

One thing worth noting is that when we compared the rescued integration sites with the PCR based NGS generated integration sites, we did nott find any matches. This might be due to two reasons. Firstly, the strict requirement for the presence of an integration in multiple libraries (in the NGS sequencing strategy) likely excludes many legitimate sites from analysis. Secondly, the PCR based NGS methods were used to identify the integration sites of AAV-rDNA-hFah, while the rescue strategy was used to identify the integration sites of AAV-rDNA-I-*Sce*I-KO. The integration sites between two vectors might be slightly different, although the total integration patterns are similar.

The large-scale investigation of AAV-rDNA integration in this chapter shows about 10-30 times more integration, with about 30 % of integration within rDNA repeats, which is about 10 times more than regular AAV vectors (Nakai et al., 2005b). We have also shown that AAV-rDNA vectors have very similar integration patterns as regular AAV vectors, except for about 10 times higher rDNA preference. Also the similarity between PCR generated NGS and shuttle vector strategy suggested that the PCR based library matching NGS method that we used is a powerful and reliable method to correct for PCR stringency generated errors.

Chapter 5

Summary and Conclusions

I Summary

The recombinant adeno-associated viral (rAAV) vector is currently one of the most promising gene therapy modalitites. It has been used with success in many animal models, and in over 80 human clinical trials (Mingozzi and High, 2011). Long-term AAV-mediated transgene expression at therapeutic levels would be highly desirable for genetic liver diseases. However, in animal models the vast majority of rAAV genomes in the liver are episomal and less than 5–10% of their expression persists after partial hepatectomy (Nakai et al., 2001). In healthy humans, hepatocytes are known to divide every 200–300 days (Macdonald, 1961) and this rate is predicted to be significantly higher in cases with liver injury such as viral hepatitis and other infections.

To date, however, limited progress has been made in terms of developing rAAV vectors capable of more efficient chromosomal integration.

rDNA preference and the strategy

Large-scale analysis of AAV integration sites in hepatocytes showed that about 60% of rAAV vector integrations occurred in active genes (Nakai et al., 2003). In addition, integration 'hot spots' were found and one surprising finding was that 3–8% of integrated vectors resided in ribosomal DNA repeats. Even though this sequence has ~200 copies/diploid genome, the observed AAV integration frequency corresponds to a strong overrepresentation in terms of integration specificity (Miller et al., 2005; Nakai et al., 2005b). Given the preference of unmodified rAAV vectors for integration into rDNA and knowing that uncoating of the virus occurs in the nucleus (Johnson and Samulski, 2009)

we reasoned that introduction of rDNA homology into rAAV vectors would increase their integration frequency, especially into rDNA repeats, by homologous recombination or non-homologous end-joining.

AAV-rDNA-FAH achieved 10–30 times more integration than regular AAV vectors in HT1 mouse model

It is essential to find a disease model to study AAV integration *in vivo*. A good mouse model needs to be able to: 1) positively select the integrated AAV genome *in vivo* while eliminating episomal AAV genomes; 2) easily distinguish cells containing AAV genomes from cells without AAV genomes; 3) quantitatively measure the AAV genomes and measure the therapeutic effect *in vivo*; 4) survive long enough to demonstrate the long-term effect of AAV integration. It is extremely important to evaluate the safety of AAV integration.

The HT1 mouse model is a selective mouse model in which FAH-expressed hepatocytes repopulated Fah-deficient livers (Overturf et al., 1996). In HT1 mice, FAH-positive nodules gradually grow over the Fah-negative hepatocytes and finally repopulate the whole liver.

1) 20–30 times higher integration frequency than regular AAV.

Firstly, we used the HT1 mouse model to characterize this vector. At the DNA level, AAV-rDNA-FAH produced 20–30 times more integrated viral genomes than control AAV-stuffer-FAH at the dosages of $3x10^{11}$ and $1x10^{10}$ vg/mouse according to real time PCRs. We measured the total integrated AAV genome without the selection

pressure on transgene expression cassettes. The 20–30-fold difference in expression of the integrated transgene cassette between AAV-rDNA and AAV-stuffer illustrated the difference in integration capability of these two vectors, but did not indicate whether or not the integrated transgene expresses functional protein or whether the integrated transgene expression cassettes are silenced. To measure the expression of the integrated cassettes, we used anti-FAH immunostaining to stain the AAV-infected liver and found that AAV-rDNA-FAH produced 18 times more FAH-positive nodules than control vectors. The detection of FAH-positive hepatocytes confirmed that that AAV had integrated and also expressed enough FAH protein to be detected by immunostaining, which is comparison of functional transgene integration. At the functional level, AAV-rDNA-FAH vectors were able to rescue HT1 mice, which would die without treatment, at a 10–30-fold lower dose than comparable controls. Importantly, rescued *Fah*$^{-/-}$ mice all showed normal liver function similar to that of wild-type mice irrespective of dose. These data provide evidence that AAV-rDNA vectors allow about 10–30 times more viral vector integration with functional transgene expression than regular AAV vectors. Up to 30 times more integration will provide long-lasting transgene expression in the liver. In some genetic liver diseases, such as HT1, Crigler-Najjar syndrome I , hepatocyte injury will induce liver regeneration, which means that only integrated AAV genomes replicate with the host genome. In these diseases, AAV-rDNA vectors that integrate efficiently will be able to provide long-term therapy, unlike regular AAV vectors. Furthermore, the 30-fold better therapeutic effect means that we can use a much lower dose of AAV-rDNA vectors to treat disease. Gene therapy with a low dose

of vector might also be a strategy for alleviating the host immune response while maintaining or improving the therapeutic effect.

2) Site-specific integration

The work done for this thesis also confirmed the site-specific integration using PCR and sequencing. Although the 1.3% site-specific integration ratio is low, this demonstrated that the strategy we were using was on the right track to achieve site-specific integration. Generating a double-strand break in the homologous rDNA region would significantly increase the integration frequency along with the site-specific integration ratio. Confirmation of site-specific integration is central to the application and improvement of AAV-rDNA vectors. Site-specific integration, as one of the central interests of AAV gene therapy in recent years, not only has a highly efficient therapeutic effect, but also overcomes concerns about mutagenesis and tumorgenesis due to random integration.

3) Reasonable viral copy number

In viral gene therapy, one of the key questions is how many copies of viral genomes are there in each host cell? We used southern blotting to measure the integrated transgene expression cassettes. The copy number for AAV2-rDNA-hFah in primary hepatocytes was about 2; for AAV8 in primary hepatocytes the copy number was about 8. It is essential to maintain enough integrated AAV-mediated transgene expression cassettes expressing functional protein to correct the functional deficiency while avoiding excessive integration that could lead to genotoxicity and potential mutagenicity. We

also analyzed the effects of low dose ($1x10^9$vg/mouse) AAV-rDNA-FAH liver treatment. After successfully correcting *Fah*$^{-/-}$ mice and repopulating the *Fah*$^{-/-}$ liver, the copy number was only ~0.6 copies/diploid cell. This low copy number indicated that the AAV-rDNA-FAH vector is capable of providing efficient therapy while limiting the potential genotoxicity and mutagenicity, which would be highly desirable in clinical trials.

4) Enriched rDNA integration sites

To identify the integration preference of the AAV-rDNA vectors, we applied a shuttle vector strategy and identified 12 integration sites. Four out of 12 integration sites were within rDNA repeats and three out of these four were within the homologous sequence contained within the vector, indicating a strong preference for rDNA, as expected. This proportion (30%) of total integration sites within the rDNA repeats equates to a 100-fold enrichment in comparison to random integration based on the size fraction (200x45000 bp/$3x10^9$ bp, 0.3%). In addition, the percentage of total integration sites within rDNA repeats (30%) was 10-fold higher than when using the regular AAV-vector, for which the equivalent rate was ~3% (Nakai et al., 2005b). Five out of eight integration sites within other chromosomal regions were inside the Refseq genes, a rate that was significantly higher than that achieved by random integration (36.2%, $p<0.001$) and similar to that achieved by regular AAV vectors (53.1%), as previously reported (Nakai et al., 2005b). The preference for Refseq genes might be due to the fact that active transcription regions have more double-strand breaks, which are preferential integration sites for AAV vectors (Miller et al., 2004). The significant enrichment in

rDNA repeats and preference for Refseq genes indicated that the integration pattern of AAV-rDNA vectors was similar to that of regular AAV vectors except the significantly enrichment in rDNA repeats. Notably, none of the 12 integration sites were within tumor-related genes, a desirable feature for clinical trials.

AAV-rDNA-hF.IX achieved higher hF.IX persistence in hemophilia disease model

1) Almost 40% persistent hFIX expression

To test whether this strategy is limited to the HT1 mouse model, given that the HT1 mouse model is a selection model (Overturf et al., 1997b), we collaborated with Mark Kay's lab at Stanford University to test whether AAV-rDNA-hF.IX gave more persistent hF.IX expression. We showed that after *in vivo* delivery the rAAV-rDNA vectors provided sustained therapeutic levels (~40%) of serum hF.IX even after repeated manipulation to induce liver regeneration. This was a significant improvement over regular AAV, which provides less than 10% F.IX after the same treatment. These persistent levels of FIX will enable long-term therapy of hemophilia B patients, in which regular AAV-mediated transient F.IX expression was not enough to last for 2 months in one clinical trial (Manno et al., 2006). The results also suggest that the rAAV-rDNA vectors integrate into the genomic rDNA locus 8–18 times more frequently than control vectors, indicating that 23–54% of integrations are specific to the rDNA locus. The enrichment in rDNA regions was due to the homologous rDNA arm, since integration of the control AAV-stuffer vector did not lead to enrichment in the rDNA repeats.

2) Enriched rDNA integration

To identify whether the entire 2 kb homologous rDNA is necessary for rDNA integration and increased integration frequency, we constructed different AAV-rDNA vectors with rDNA arms of 500 bp, 750 bp and 1000 bp. *In vivo* studies showed that AAV-rDNA with 1000 bp or 750 bp rDNA arms were capable of rDNA integration and increased integration frequency, while the AAV-rDNA with 500 bp arms showed no significant improvement over the sizable control AAV stuffer. The 4.7 kb packaging limit of the AAV is generally considered a disadvantage for AAV gene therapy, so is the AAV-rDNA (Grieger and Samulski, 2005). For AAV-rDNA, the size of the transgene expression cassette is limited to 2.7 kb. Identifying the minimal length for rDNA integration and increased integration frequency will be of significance for possible applications in diseases that require large transgene expression cassettes.

The results from the hemophilia B disease model provided further evidence that the AAV-rDNA vector is superior to regular AAV vectors in terms of persistence of transgene expression. As the HT1 mouse model is a selective mouse model and the hemophilia B mouse model is one of the most extensively studied disease models in the gene therapy field, our success in two different models strongly indicated that AAV-rDNA vectors could be used in most disease models.

Large-scale analysis of integration sites of AAV-rDNA vectors showed similar integration pattern as regular AAV vectors

To identify the integration sites of AAV-rDNA vectors, we used two high-throughput strategies: the shuttle vector strategy and a novel strategy in which we

combined LAM-PCR and Solexa NGS technology. The main advantage of the shuttle vector strategy is its high fidelity, such that the results obtained using this method are very reliable with high specificity, and also provide both ends of the integration sites. That is because the shuttle vector strategy does not involve the random primer annealing of PCR-based methods. The disadvantage of this strategy is that the results are biased by the distribution of restriction sites. The advantage of the LAM-PCR-based NGS strategy is the large amount of sequence (integration sites) data created. The disadvantage is that many of the integration sites identified using this strategy are not real integration sites due to random primer binding. Here we applied both methods and compared the results of each to identify the integration preference and integration patterns of AAV-rDNA as faithfully as possible.

Shuttle vector strategy

Large-scale analysis of the integration sites suggested about a 10-fold preference for integration into rDNA repeats, which are relatively safe loci, while the overall integration pattern was similar tohat of regular AAV vectors, with the exception of a stronger rDNA preference. Importantly, the large-scale integration analysis showed no preference of AAV-rDNA vectors for tumor-related genes.

1) 10-fold higher enrichment in rDNA repeats than regular AAV vectors

Using the shuttle vector strategy, we identified 12 integration sites from over 4000 clones. Four out of 12 integration sites were within rDNA repeats, accounting for 30% of all integration sites. Previously Nakai *et al.* showed that 3% of integration sites of AAV2

were within rDNA repeats, thus demonstrating about a 10-fold enrichment compared with random integration (Nakai et al., 2005b). By comparing AAV-rDNA vectors and AAV vectors, we can conclude that AAV-rDNA vectors show about 10-fold higher enrichment in rDNA repeats than regular AAV vectors. The introduction of rDNA sequences into AAV vector can significantly improve integration into rDNA repeats as well as the total integration frequency, which is highly desirable for genetic diseases.

2) Preference for transcription regions

To better understand the integration pattern of AAV-rDNA and to determine whether AAV-rDNA has a similar integration preference to regular AAV, we considered the integrations within rDNA repeats separately from integrations elsewhere. Of those integration sites outside rDNA repeats, five out of eight were within RefSeq genes, that is, a significantly higher proportion (62.5% within RefSeq genes) than in random integration (36.2% within RefSeq genes, $P<0.0001$)(Nakai et al., 2005b). Nakai has previously shown that AAV2 preferentially integrates into RefSeq genes in the mouse liver (Nakai et al., 2003; Nakai et al., 2005b). AAV-rDNA vectors had a similar integration preference for RefSeq genes, which are often referred to as active transcription regions. This might be due to the fact that there are more double-strand breaks in active transcription regions, which are thus preferred for AAV integration (Miller et al., 2004).

3) Breakpoint analysis indicated different integration mechanism

We analyzed the breakpoint of four integration sites within rDNA repeats to determine whether integration within rDNA repeats was mediated by homologous

recombination. Four breakpoints analysis showed that one with nucleotides addition, two with micro-homologous sequence and one with direct joining, suggesting that there might be more than one mechanism involved in AAV-rDNA integrations, even within rDNA repeats, particularly as homologous recombination only generates unique breakpoints, as discussed in Chapter II.

LAM-PCR-based cross-library NGS

As the shuttle vector strategy only generated 12 integration sites, precluding any conclusion on the preference of AAV-rDNA vectors for integration, we decided to combine LAM-PCR and Solexa NGS technology in order to conduct a large-scale integration site analysis. To avoid false-positive integration sites generated by non-specific primer binding, we designed three primers to generate three independent libraries in order to identify common integration sites within the three libraries. These were presumed to be real integration sites, as the chance that two different primers would randomly bind to the same location is very low (about $20/3x10^9$). Using this innovative large-scale strategy, we analyzed the integration sites and compared the results with those generated by the reliable shuttle vector.

1) Integration frequency within rDNA repeats was about 100 times higher than that achieved by random integration according to tag density.

To measure the chromosomal preference of integrations, we used sequencing tag density as an indicator. Since each tag represented one possible integration site, the presence of more tags within one defined chromosomal region suggested that this region

was preferred for integration. In this study, we compared regions with rDNA repeats with other chromosomal regions. The tag density in rDNA repeats was about 100 times that of other chromosomal regions, which suggested that AAV-rDNA vectors preferred rDNA repeats about 100 times more than other chromosomal regions. This 100 times' enrichment was consistent with the results of the shuttle vector strategy. Both strategies indicated that AAV-rDNA vectors preferred rDNA repeats 100 times more than random integration and about 10 times more than regular AAV vectors. This also suggested that homologous rDNA fragments in AAV-rDNA vectors play important roles in rDNA integration, which proved the original design to integrate into these rDNA regions. Importantly, as rDNA gene had about 200 copies in each host cells and was not considered as tumor related genes, the increased rate of integration in these regions will not have any significant implications for mutagenesis and tumorgenesis.

2) AAV-rDNA vectors have a preference for RefSeq genes (transcription regions)

Each virus has its own integration preferences, for example, HIV prefers to integrate into active transcription regions while MLV was found to favor integration into CpG island-represented promoters (Mitchell et al., 2004). To investigate whether AAV-rDNA has a similar integration preference to regular AAV, which has been shown to prefer integration into RefSeq genes as well as CpG islands and transcription start sites, we analyzed all integration sites and mapped these to the RefSeq library. In comparison with randomly generated control integration sites, AAV-rDNA was found favor integration into transcription regions (52.5% *vs.* 36.2%, $P<0.0001$). Importantly, the preference for RefSeq genes was similar to that of regular AAV, which suggested that the

introduction of the homologous rDNA fragment did not change the basic integration pattern of AAV. Thus both the shuttle vector strategy and LAM-PCR-based cross-library NGS strategy confirmed that AAV-rDNA vectors have a similar integration preference to regular AAV. The similarity of the results also indicated that the LAM-PCR-based cross-library NGS strategy is a reliable method for the large-scale analysis of integration sites.

The possible mechanisms of preference for transcription regions might be 1) active transcription regions are more accessible to the AAV genome, given that when the chromosome forms condensed chromatin, the AAV genome will not be able to access the DNA in the chromosome; 2) during active transcription, transcription of a DNA sequence increases the formation of DNA breaks or DNA lesions that are processed into DNA breaks, which are favored by AAV integrations (Gonzalez-Barrera et al., 2002; Miller et al., 2004).

3) AAV-rDNA vectors have a preference for CpG islands and transcription start sites (TSS).

To further analyze the integration preferences of AAV-rDNA vectors, we compared the integration sites with a CpG islands library and TSS library and found that AAV-rDNA integrated into CpG islands ~10 times more frequently than the random set and integrated into regions near TSSs (within 5 KB) about ~40 times more frequently than the random set. The preferential integration into CpG islands and TSS regions by AAV-rDNA was also similar to that of regular AAV, which has been reported to prefer integration into both regions (Nakai et al., 2005b). This analysis provided further

evidence that AAV-rDNA shares similar integration preferences to regular AAV. AAV-rDNA vectors thus show the same safety features as regular AAV vectors.

4) AAV-rDNA vectors have no preference for tumor-related genes

Retroviral integration near proto-oncogenes has been reported to contribute to tumorigenesis by activating their expression in animal models (Coffin et al., 1997). Importantly, insertional activation has been reported in two patients treated with retroviral gene therapy (Check, 2002). AAV with a strong promoter has been reported to be associated with HCC in neonatal mice (Donsante et al., 2007). As AAV-rDNA is an integrating vector, it was important to determine whether AAV-rDNA has a preference for integration in tumor-related genes. We mapped all the integration sites to the Tumor Gene Database (http://www.tumor-gene.org /TGDB/tgdb.html), and the Retrovirus Tagged Cancer Gene Database (http://rtcgd.ncifcrf.gov/). Any genes matching one of these two databases were defined as oncogenes. However, comparison of the random set and AAV-rDNA integration sites showed no preference for tumor-related genes. Interestingly, the ratio of AAV-rDNA integration sites within tumor-related genes was less than the ratio of regular AAV integration sites (Nakai et al., 2005b). Together, this suggested that AAV-rDNA is an integrating vector with a higher integration frequency with a preference for integration in rDNA regions but does not increase the risk of tumorigenesis and mutagenesis, which is one of the greatest concerns in gene therapy and clinical trials.

Large-scale integration site analysis revealed that AAV-rDNA vectors show a 10-fold higher preference for integration into rDNA repeats compared with regular AAV,

which not only provided the explanation for the higher integration frequency of AAV-rDNA, but also indicated the feasibility of a site-specific integration strategy. The analysis also demonstrated that the AAV-rDNA vector, regardless of its higher integration frequency and enriched rDNA integration, shared similar, if not the same, integration patterns as regular AAV. Since regular AAV vectors are widely used in clinical trials (Mingozzi and High, 2011), AAV-rDNA should be considered when long-term transgene expression is required.

Physiological and pathological perspective

The nucleolus, as a dynamic sub-nuclear structure in ribosome subunit biogenesis, mediates cell-stress responses (Rubbi and Milner, 2003) and regulates cell growth (Carmo-Fonseca et al., 2000). Metabolic conditions will also change the proteome and structure of the nucleolus (Carmo-Fonseca et al., 2000). The nucleolus has been found to interact with many viruses including DNA viruses, retroviruses and RNA viruses. As an increasing number of studies on the nucleolus have found that the nucleolus is associated with cell stress (Boisvert et al., 2007; Boulon et al., 2010; Hiscox et al., 2010), it will be interesting to determine how the nucleolus responds to AAV infection. In addition, some RNA viruses have been found to disrupt nucleolar function or use nucleolar proteins to help virus replication (Hiscox, 2007). Does the localization of AAV in the nucleolus have some physiological impact on host cells?

II Conclusions

The work described herein designed a new integrating AAV-rDNA vector and proved that this new vector has at least a 10-fold higher capacity for integration with persistent transgene expression than regular AAV vectors, without long-term safety concerns. AAV-rDNA vectors were applied in two disease models and, as proof-of-concept, in a human fibroblast cell line, all of which showed about 10-fold higher functional integration. The increased integration frequency of such vectors is desirable for the treatment of genetic diseases. AAV-rDNA vectors should be considered when therapeutic transgenes are small enough to accommodate the inclusion of rDNA homology arms.

Practically, AAV-rDNA vectors will aid the clinical application of AAV, as the improved integration efficiency will significantly reduce the dosage and costs in the clinical situation, both of these being barriers to the clinical application of AAV (Grimm and Kleinschmidt, 1999; Samulski, 2003).

In summary: 1) we developed an integrating vector with a 10–30-fold higher integration frequency but no increase in safety concerns, providing a distinct choice for gene therapy in genetic liver diseases. This vector was shown to be efficient in two disease models. 2) We introduced a new strategy that could be used for the development and application of AAV vectors in human gene therapy applications in the near future. With a controllable double-strand break generator like zinc-finger and an ideal integration candidate, a higher-than-therapeutic-level integration vector can be achieved in the reasonably near future. Furthermore, with the introduction of other homologous sequences, the targeting efficiency might be higher and closer to the therapeutic level. 3) We invented a new strategy (LAM-PCR-based cross-library NGS sequencing) to identify

open-end integration sites with very high stringency, which solved the issues of low stringency in the large-scale analysis method.

REFERENCES

Argast, G.M., Stephens, K.M., Emond, M.J., and Monnat, R.J., Jr. (1998). I-PpoI and I-CreI homing site sequence degeneracy determined by random mutagenesis and sequential in vitro enrichment. J Mol Biol *280*, 345-353.

Bain, M.D., Purkiss, P., Jones, M., Bingham, P., Stacey, T.E., and Chalmers, R.A. (1990). Dietary treatment eliminates succinylacetone from the urine of a patient with tyrosinaemia type 1. Eur J Pediatr *149*, 637-639.

Baum, C., Dullmann, J., Li, Z., Fehse, B., Meyer, J., Williams, D.A., and von Kalle, C. (2003). Side effects of retroviral gene transfer into hematopoietic stem cells. Blood *101*, 2099-2114.

Bengtsson, C., Gredmark, T., Hallberg, L., Hallstrom, T., Isaksson, B., Lapidus, L., Lindquist, O., Lindstedt, S., Lurie, M., Nystrom, E., *et al.* (1989). The population study of women in Gothenburg 1980-81--the third phase of a longitudinal study. Comparison between participants and non-participants. Scand J Soc Med *17*, 141-145.

Berns, K.I., and Giraud, C. (1996). Biology of adeno-associated virus. Curr Top Microbiol Immunol *218*, 1-23.

Berraondo, P., Ochoa, L., Crettaz, J., Rotellar, F., Vales, A., Martinez-Anso, E., Zaratiegui, M., Ruiz, J., Gonzalez-Aseguinolaza, G., and Prieto, J. (2005). IFN-alpha gene therapy for woodchuck hepatitis with adeno-associated virus: differences in duration of gene expression and antiviral activity using intraportal or intramuscular routes. Mol Ther *12*, 68-76.

Blomer, U., Naldini, L., Kafri, T., Trono, D., Verma, I.M., and Gage, F.H. (1997). Highly efficient and sustained gene transfer in adult neurons with a lentivirus vector. J Virol *71*, 6641-6649.

Boisvert, F.M., van Koningsbruggen, S., Navascues, J., and Lamond, A.I. (2007). The multifunctional nucleolus. Nat Rev Mol Cell Biol *8*, 574-585.

Bollag, R.J., Waldman, A.S., and Liskay, R.M. (1989). Homologous recombination in mammalian cells. Annu Rev Genet *23*, 199-225.

Boulon, S., Westman, B.J., Hutten, S., Boisvert, F.M., and Lamond, A.I. (2010). The nucleolus under stress. Mol Cell *40*, 216-227.

Brisco, M.J., Latham, S., Bartley, P.A., and Morley, A.A. (2010). Incorporation of measurement of DNA integrity into qPCR assays. Biotechniques *49*, 893-897.

Buning, 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.

Calcedo, R., Morizono, H., Wang, L., McCarter, R., He, J., Jones, D., Batshaw, M.L., and Wilson, J.M. (2011). Adeno-associated virus antibody profiles in newborns, children, and adolescents. Clin Vaccine Immunol *18*, 1586-1588.

Carmo-Fonseca, M., Mendes-Soares, L., and Campos, I. (2000). To be or not to be in the nucleolus. Nat Cell Biol *2*, E107-112.

Cerone, R., Holme, E., Schiaffino, M.C., Caruso, U., Maritano, L., and Romano, C. (1997). Tyrosinemia type III: diagnosis and ten-year follow-up. Acta Paediatr *86*, 1013-1015.

Check, E. (2002). Gene therapy: shining hopes dented - but not dashed. Nature *420*, 735.

Chen, C.L., Jensen, R.L., Schnepp, B.C., Connell, M.J., Shell, R., Sferra, T.J., Bartlett, J.S., Clark, K.R., and Johnson, P.R. (2005). Molecular characterization of adeno-associated viruses infecting children. J Virol *79*, 14781-14792.
Chen, Y.H., Chang, M., and Davidson, B.L. (2009). Molecular signatures of disease brain endothelia provide new sites for CNS-directed enzyme therapy. Nat Med *15*, 1215-1218.
Chiorini, J.A., Kim, F., Yang, L., and Kotin, R.M. (1999). Cloning and characterization of adeno-associated virus type 5. J Virol *73*, 1309-1319.
Chiorini, J.A., Yang, L., Liu, Y., Safer, B., and Kotin, R.M. (1997). Cloning of adeno-associated virus type 4 (AAV4) and generation of recombinant AAV4 particles. J Virol *71*, 6823-6833.
Choi, V.W., Asokan, A., Haberman, R.A., and Samulski, R.J. (2007a). Production of recombinant adeno-associated viral vectors. Curr Protoc Hum Genet *Chapter 12*, Unit 12 19.
Choi, V.W., Asokan, A., Haberman, R.A., and Samulski, R.J. (2007b). Production of recombinant adeno-associated viral vectors for in vitro and in vivo use. Curr Protoc Mol Biol *Chapter 16*, Unit 16 25.
Ciuffi, A., and Barr, S.D. (2010). Identification of HIV integration sites in infected host genomic DNA. Methods *53*, 39-46.
Clarke, J.K., McFerran, J.B., McKillop, E.R., and Curran, W.L. (1979). Isolation of an adeno associated virus from sheep. Brief report. Arch Virol *60*, 171-176.
Coffin, J.M., Hughes, S.H., and Varmus, H.E. (1997). The Interactions of Retroviruses and their Hosts.
Coria, M.F., and Lehmkuhl, H.D. (1978). Isolation and identification of a bovine adenovirus type 3 with an adenovirus-associated virus. Am J Vet Res *39*, 1904-1906.
Dawson, G.J., Yates, V.J., Chang, P.W., and Oprandy, J.J. (1982). Is avian adeno-associated virus an endogenous virus of chicken cells? Nature *298*, 580-582.
Daya, S., and Berns, K.I. (2008). Gene therapy using adeno-associated virus vectors. Clin Microbiol Rev *21*, 583-593.
De Braekeleer, M., and Larochelle, J. (1990). Genetic epidemiology of hereditary tyrosinemia in Quebec and in Saguenay-Lac-St-Jean. Am J Hum Genet *47*, 302-307.
DiPrimio, N., Asokan, A., Govindasamy, L., Agbandje-McKenna, M., and Samulski, R.J. (2008). Surface loop dynamics in adeno-associated virus capsid assembly. J Virol *82*, 5178-5189.
Donoho, G., Jasin, M., and Berg, P. (1998). Analysis of gene targeting and intrachromosomal homologous recombination stimulated by genomic double-strand breaks in mouse embryonic stem cells. Mol Cell Biol *18*, 4070-4078.
Donsante, A., Miller, D.G., Li, Y., Vogler, C., Brunt, E.M., Russell, D.W., and Sands, M.S. (2007). AAV vector integration sites in mouse hepatocellular carcinoma. Science *317*, 477.
Duncan, A.W., Hickey, R.D., Paulk, N.K., Culberson, A.J., Olson, S.B., Finegold, M.J., and Grompe, M. (2009). Ploidy reductions in murine fusion-derived hepatocytes. PLoS Genet *5*, e1000385.
Duncan, A.W., Taylor, M.H., Hickey, R.D., Hanlon Newell, A.E., Lenzi, M.L., Olson, S.B., Finegold, M.J., and Grompe, M. (2010). The ploidy conveyor of mature hepatocytes as a source of genetic variation. Nature *467*, 707-710.

Duncavage, E.J., Magrini, V., Becker, N., Armstrong, J.R., Demeter, R.T., Wylie, T., Abel, H.J., and Pfeifer, J.D. (2011). Hybrid capture and next-generation sequencing identify viral integration sites from formalin-fixed, paraffin-embedded tissue. J Mol Diagn *13*, 325-333.
Dyall, J., Szabo, P., and Berns, K.I. (1999). Adeno-associated virus (AAV) site-specific integration: formation of AAV-AAVS1 junctions in an in vitro system. Proc Natl Acad Sci U S A *96*, 12849-12854.
Fejes, A.P., Robertson, G., Bilenky, M., Varhol, R., Bainbridge, M., and Jones, S.J. (2008). FindPeaks 3.1: a tool for identifying areas of enrichment from massively parallel short-read sequencing technology. Bioinformatics *24*, 1729-1730.
Fisher, K.J., Jooss, K., Alston, J., Yang, Y., Haecker, S.E., High, K., Pathak, R., Raper, S.E., and Wilson, J.M. (1997). Recombinant adeno-associated virus for muscle directed gene therapy. Nat Med *3*, 306-312.
Flotte, T.R., Afione, S.A., and Zeitlin, P.L. (1994). Adeno-associated virus vector gene expression occurs in nondividing cells in the absence of vector DNA integration. Am J Respir Cell Mol Biol *11*, 517-521.
Frank, K.M., Hogarth, D.K., Miller, J.L., Mandal, S., Mease, P.J., Samulski, R.J., Weisgerber, G.A., and Hart, J. (2009). Investigation of the cause of death in a gene-therapy trial. N Engl J Med *361*, 161-169.
Friedrich, G., and Soriano, P. (1991). Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. Genes Dev *5*, 1513-1523.
Fuller, S.A., Takahashi, M., and Hurrell, J.G. (2001). Cloning of hybridoma cell lines by limiting dilution. Curr Protoc Mol Biol *Chapter 11*, Unit11 18.
Gao, G., Vandenberghe, L.H., and Wilson, J.M. (2005). New recombinant serotypes of AAV vectors. Curr Gene Ther *5*, 285-297.
Gao, G., Wang, Q., Calcedo, R., Mays, L., Bell, P., Wang, L., Vandenberghe, L.H., Grant, R., Sanmiguel, J., Furth, E.E., *et al.* (2009). Adeno-associated virus-mediated gene transfer to nonhuman primate liver can elicit destructive transgene-specific T cell responses. Hum Gene Ther *20*, 930-942.
Gao, G.P., Alvira, M.R., Wang, L., Calcedo, R., Johnston, J., and Wilson, J.M. (2002). Novel adeno-associated viruses from rhesus monkeys as vectors for human gene therapy. Proc Natl Acad Sci U S A *99*, 11854-11859.
Giering, J.C., Grimm, D., Storm, T.A., and Kay, M.A. (2008). Expression of shRNA from a tissue-specific pol II promoter is an effective and safe RNAi therapeutic. Mol Ther *16*, 1630-1636.
Girod, A., Ried, M., Wobus, C., Lahm, H., Leike, K., Kleinschmidt, J., Deleage, G., and Hallek, M. (1999). Genetic capsid modifications allow efficient re-targeting of adeno-associated virus type 2. Nat Med *5*, 1438.
Gluecksohn-Waelsch, S. (1979). Genetic control of morphogenetic and biochemical differentiation: lethal albino deletions in the mouse. Cell *16*, 225-237.
Gonzalez-Barrera, S., Garcia-Rubio, M., and Aguilera, A. (2002). Transcription and double-strand breaks induce similar mitotic recombination events in Saccharomyces cerevisiae. Genetics *162*, 603-614.
Greelish, J.P., Su, L.T., Lankford, E.B., Burkman, J.M., Chen, H., Konig, S.K., Mercier, I.M., Desjardins, P.R., Mitchell, M.A., Zheng, X.G., *et al.* (1999). Stable restoration of

the sarcoglycan complex in dystrophic muscle perfused with histamine and a recombinant adeno-associated viral vector. Nat Med *5*, 439-443.
Grieger, J.C., and Samulski, R.J. (2005). Packaging capacity of adeno-associated virus serotypes: impact of larger genomes on infectivity and postentry steps. J Virol *79*, 9933-9944.
Grimm, D. (2002). Production methods for gene transfer vectors based on adeno-associated virus serotypes. Methods *28*, 146-157.
Grimm, D., and Kleinschmidt, J.A. (1999). Progress in adeno-associated virus type 2 vector production: promises and prospects for clinical use. Hum Gene Ther *10*, 2445-2450.
Grimm, D., Lee, J.S., Wang, L., Desai, T., Akache, B., Storm, T.A., and Kay, M.A. (2008). In vitro and in vivo gene therapy vector evolution via multispecies interbreeding and retargeting of adeno-associated viruses. J Virol *82*, 5887-5911.
Grompe, M. (2001). The pathophysiology and treatment of hereditary tyrosinemia type 1. Semin Liver Dis *21*, 563-571.
Grompe, M. (2006). Principles of therapeutic liver repopulation. J Inherit Metab Dis *29*, 421-425.
Grompe, M., al-Dhalimy, M., Finegold, M., Ou, C.N., Burlingame, T., Kennaway, N.G., and Soriano, P. (1993). Loss of fumarylacetoacetate hydrolase is responsible for the neonatal hepatic dysfunction phenotype of lethal albino mice. Genes Dev *7*, 2298-2307.
Grompe, M., Caskey, C.T., and Fenwick, R.G. (1991). Improved molecular diagnostics for ornithine transcarbamylase deficiency. Am J Hum Genet *48*, 212-222.
Grompe, M., Lindstedt, S., al-Dhalimy, M., Kennaway, N.G., Papaconstantinou, J., Torres-Ramos, C.A., Ou, C.N., and Finegold, M. (1995). Pharmacological correction of neonatal lethal hepatic dysfunction in a murine model of hereditary tyrosinaemia type I. Nat Genet *10*, 453-460.
Hacein-Bey-Abina, S., von Kalle, C., Schmidt, M., Le Deist, F., Wulffraat, N., McIntyre, E., Radford, I., Villeval, J.L., Fraser, C.C., Cavazzana-Calvo, M., *et al.* (2003a). A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency. N Engl J Med *348*, 255-256.
Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M., McCormack, M.P., Wulffraat, N., Leboulch, P., Lim, A., Osborne, C.S., Pawliuk, R., Morillon, E., *et al.* (2003b). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. Science *302*, 415-419.
Hartman, Z.C., Kiang, A., Everett, R.S., Serra, D., Yang, X.Y., Clay, T.M., and Amalfitano, A. (2007). Adenovirus infection triggers a rapid, MyD88-regulated transcriptome response critical to acute-phase and adaptive immune responses in vivo. J Virol *81*, 1796-1812.
Hasbrouck, N.C., and High, K.A. (2008). AAV-mediated gene transfer for the treatment of hemophilia B: problems and prospects. Gene Ther *15*, 870-875.
Hayakawa, J., Washington, K., Uchida, N., Phang, O., Kang, E.M., Hsieh, M.M., and Tisdale, J.F. (2009). Long-term vector integration site analysis following retroviral mediated gene transfer to hematopoietic stem cells for the treatment of HIV infection. PLoS One *4*, e4211.

Hedner, U., Ginsburg, D., Lusher, J.M., and High, K.A. (2000). Congenital Hemorrhagic Disorders: New Insights into the Pathophysiology and Treatment of Hemophilia. Hematology Am Soc Hematol Educ Program, 241-265.
Held, P.K., Olivares, E.C., Aguilar, C.P., Finegold, M., Calos, M.P., and Grompe, M. (2005). In vivo correction of murine hereditary tyrosinemia type I by phiC31 integrase-mediated gene delivery. Mol Ther *11*, 399-408.
Hendrie, P.C., Hirata, R.K., and Russell, D.W. (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.
High, K. (2002). AAV-mediated gene transfer for hemophilia. Genet Med *4*, 56S-61S.
High, K.A. (2003). Gene transfer as an approach to treating hemophilia. Semin Thromb Hemost *29*, 107-120.
High, K.A., and Aubourg, P. (2011). rAAV Human Trial Experience. Methods Mol Biol *807*, 429-457.
High, K.P., Bradley, S., Loeb, M., Palmer, R., Quagliarello, V., and Yoshikawa, T. (2005). A new paradigm for clinical investigation of infectious syndromes in older adults: assessment of functional status as a risk factor and outcome measure. Clin Infect Dis *40*, 114-122.
Hildebrandt, M., and Nellen, W. (1991). Library-independent cloning of genomic fragments adjacent to vector integration sites: isolation of the EB4-PSV gene from a Dictyostelium gene disruption transformant. Biochem Biophys Res Commun *181*, 884-888.
Hirata, R., Chamberlain, J., Dong, R., and Russell, D.W. (2002). Targeted transgene insertion into human chromosomes by adeno-associated virus vectors. Nat Biotechnol *20*, 735-738.
Hirata, R.K., and Russell, D.W. (2000). Design and packaging of adeno-associated virus gene targeting vectors. J Virol *74*, 4612-4620.
Hiscox, J.A. (2007). RNA viruses: hijacking the dynamic nucleolus. Nat Rev Microbiol *5*, 119-127.
Hiscox, J.A., Whitehouse, A., and Matthews, D.A. (2010). Nucleolar proteomics and viral infection. Proteomics *10*, 4077-4086.
Holme, E., and Lindstedt, S. (1998). Tyrosinaemia type I and NTBC (2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione). J Inherit Metab Dis *21*, 507-517.
Hooker, D.J., Gorry, P.R., Ellett, A.M., Wesselingh, S.L., and Cherry, C.L. (2009). Measuring and monitoring apoptosis and drug toxicity in HIV patients by ligation-mediated polymerase chain reaction. J Cell Mol Med *13*, 948-958.
Horwich, A.L. (1991). Inherited hepatic enzyme defects as candidates for liver-directed gene therapy. Curr Top Microbiol Immunol *168*, 185-200.
Hutchison, C.A., 3rd (2007). DNA sequencing: bench to bedside and beyond. Nucleic Acids Res *35*, 6227-6237.
Iannacone, M., Sitia, G., Isogawa, M., Marchese, P., Castro, M.G., Lowenstein, P.R., Chisari, F.V., Ruggeri, Z.M., and Guidotti, L.G. (2005). Platelets mediate cytotoxic T lymphocyte-induced liver damage. Nat Med *11*, 1167-1169.
Inagaki, K., Lewis, S.M., Wu, X., Ma, C., Munroe, D.J., Fuess, S., Storm, T.A., Kay, M.A., and Nakai, H. (2007). DNA palindromes with a modest arm length of greater,

similar 20 base pairs are a significant target for recombinant adeno-associated virus vector integration in the liver, muscles, and heart in mice. J Virol *81*, 11290-11303.
Ishiwata, A., Mimuro, J., Mizukami, H., Kashiwakura, Y., Takano, K., Ohmori, T., Madoiwa, S., Ozawa, K., and Sakata, Y. (2009). Liver-restricted expression of the canine factor VIII gene facilitates prevention of inhibitor formation in factor VIII-deficient mice. J Gene Med *11*, 1020-1029.
Jennings, K., Miyamae, T., Traister, R., Marinov, A., Katakura, S., Sowders, D., Trapnell, B., Wilson, J.M., Gao, G., and Hirsch, R. (2005). Proteasome inhibition enhances AAV-mediated transgene expression in human synoviocytes in vitro and in vivo. Mol Ther *11*, 600-607.
Johansen, S.D., Haugen, P., and Nielsen, H. (2007). Expression of protein-coding genes embedded in ribosomal DNA. Biol Chem *388*, 679-686.
Johnson, J.S., and Samulski, R.J. (2009). Enhancement of adeno-associated virus infection by mobilizing capsids into and out of the nucleolus. J Virol *83*, 2632-2644.
Kalpana, G.V. (1999). Retroviral vectors for liver-directed gene therapy. Semin Liver Dis *19*, 27-37.
Kay, M.A., and Fausto, N. (1997). Liver regeneration: prospects for therapy based on new technologies. Mol Med Today *3*, 108-115.
Kay, M.A., Rothenberg, S., Landen, C.N., Bellinger, D.A., Leland, F., Toman, C., Finegold, M., Thompson, A.R., Read, M.S., Brinkhous, K.M., *et al.* (1993). In vivo gene therapy of hemophilia B: sustained partial correction in factor IX-deficient dogs. Science *262*, 117-119.
Kelsey, G., Ruppert, S., Beermann, F., Grund, C., Tanguay, R.M., and Schutz, G. (1993). Rescue of mice homozygous for lethal albino deletions: implications for an animal model for the human liver disease tyrosinemia type 1. Genes Dev *7*, 2285-2297.
Kelsey, G., and Schutz, G. (1993). Lessons from lethal albino mice. Curr Opin Genet Dev *3*, 259-264.
Klebig, M.L., Russell, L.B., and Rinchik, E.M. (1992). Murine fumarylacetoacetate hydrolase (Fah) gene is disrupted by a neonatally lethal albino deletion that defines the hepatocyte-specific developmental regulation 1 (hsdr-1) locus. Proc Natl Acad Sci U S A *89*, 1363-1367.
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.
Kobayashi, T., Heck, D.J., 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.
Kogure, K., Urabe, M., Mizukami, H., Kume, A., Sato, Y., Monahan, J., and Ozawa, K. (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.
Kotin, R.M., Siniscalco, M., Samulski, R.J., Zhu, X.D., Hunter, L., Laughlin, C.A., McLaughlin, S., Muzyczka, N., Rocchi, M., and Berns, K.I. (1990). Site-specific integration by adeno-associated virus. Proc Natl Acad Sci U S A *87*, 2211-2215.

Kruger, L., Eskerski, H., Dinsart, C., Cornelis, J., Rommelaere, J., Haberkorn, U., and Kleinschmidt, J.A. (2008). Augmented transgene expression in transformed cells using a parvoviral hybrid vector. Cancer Gene Ther *15*, 252-267.
Kubo, S., Sun, M., Miyahara, M., Umeyama, K., Urakami, K., Yamamoto, T., Jakobs, C., Matsuda, I., and Endo, F. (1998). Hepatocyte injury in tyrosinemia type 1 is induced by fumarylacetoacetate and is inhibited by caspase inhibitors. Proc Natl Acad Sci U S A *95*, 9552-9557.
Kvittingen, E.A., Rootwelt, H., Berger, R., and Brandtzaeg, P. (1994). Self-induced correction of the genetic defect in tyrosinemia type I. J Clin Invest *94*, 1657-1661.
Kvittingen, E.A., Rootwelt, H., Brandtzaeg, P., Bergan, A., and Berger, R. (1993). Hereditary tyrosinemia type I. Self-induced correction of the fumarylacetoacetase defect. J Clin Invest *91*, 1816-1821.
Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.
Lee, H.C., Fang, S.B., Yeung, C.Y., and Tsai, J.D. (2005). Urinary tract infections in infants: comparison between those with conjugated vs unconjugated hyperbilirubinaemia. Ann Trop Paediatr *25*, 277-282.
Lefebvre, G., Desfarges, S., Uyttebroeck, F., Munoz, M., Beerenwinkel, N., Rougemont, J., Telenti, A., and Ciuffi, A. (2011). Analysis of HIV-1 expression level and sense of transcription by high-throughput sequencing of the infected cell. J Virol *85*, 6205-6211.
Lin, J., and Vogt, V.M. (1998). I-PpoI, the endonuclease encoded by the group I intron PpLSU3, is expressed from an RNA polymerase I transcript. Mol Cell Biol *18*, 5809-5817.
Lindblad, B., Lindstedt, S., and Steen, G. (1977). On the enzymic defects in hereditary tyrosinemia. Proc Natl Acad Sci U S A *74*, 4641-4645.
Lindstedt, S., Holme, E., Lock, E.A., Hjalmarson, O., and Strandvik, B. (1992). Treatment of hereditary tyrosinaemia type I by inhibition of 4-hydroxyphenylpyruvate dioxygenase. Lancet *340*, 813-817.
Lipiec, A., Malecki, M., and Hajdukiewic, K. (2009). [Serotypes of adeno-associated viruses]. Postepy Biochem *55*, 95-102.
Lozier, J.N., and High, K.A. (1990). Molecular basis of hemophilia. Hematol Pathol *4*, 1-26.
Lozier JN, K.C. (2005). Clinical aspects and therapy of hemophilia. (New York, Elsevier Churchill Livingstone).
Macdonald, R.A. (1961). "Lifespan" of liver cells. Autoradio-graphic study using tritiated thymidine in normal, cirrhotic, and partially hepatectomized rats. Arch Intern Med *107*, 335-343.
Maguire, A.M., High, K.A., Auricchio, A., Wright, J.F., Pierce, E.A., Testa, F., Mingozzi, F., Bennicelli, J.L., Ying, G.S., Rossi, S., *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.
Maguire, A.M., Simonelli, F., Pierce, E.A., Pugh, E.N., Jr., Mingozzi, F., Bennicelli, J., Banfi, S., Marshall, K.A., Testa, F., Surace, E.M., *et al.* (2008). Safety and efficacy of gene transfer for Leber's congenital amaurosis. N Engl J Med *358*, 2240-2248.

Maniak, M., and Nellen, W. (1989). pISAR, a tool for cloning genomic sequences adjacent to the site of vector integration. Nucleic Acids Res *17*, 4894.
Manno, C.S., Chew, A.J., Hutchison, S., Larson, P.J., Herzog, R.W., Arruda, V.R., Tai, S.J., Ragni, M.V., Thompson, A., Ozelo, M., *et al.* (2003). AAV-mediated factor IX gene transfer to skeletal muscle in patients with severe hemophilia B. Blood *101*, 2963-2972.
Manno, C.S., Pierce, G.F., Arruda, V.R., Glader, B., Ragni, M., Rasko, J.J., Ozelo, M.C., Hoots, K., Blatt, P., Konkle, B., *et al.* (2006). Successful transduction of liver in hemophilia by AAV-Factor IX and limitations imposed by the host immune response. Nat Med *12*, 342-347.
Masurel-Paulet, A., Poggi-Bach, J., Rolland, M.O., Bernard, O., Guffon, N., Dobbelaere, D., Sarles, J., de Baulny, H.O., and Touati, G. (2008). NTBC treatment in tyrosinaemia type I: long-term outcome in French patients. J Inherit Metab Dis *31*, 81-87.
Mattar, C.N., Nathwani, A.C., Waddington, S.N., Dighe, N., Kaeppel, C., Nowrouzi, A., McIntosh, J., Johana, N.B., Ogden, B., Fisk, N.M., *et al.* (2011). Stable Human FIX Expression After 0.9G Intrauterine Gene Transfer of Self-complementary Adeno-associated Viral Vector 5 and 8 in Macaques. Mol Ther *19*, 1950-1960.
McCarty, D.M., Young, S.M., Jr., and Samulski, R.J. (2004). Integration of adeno-associated virus (AAV) and recombinant AAV vectors. Annu Rev Genet *38*, 819-845.
McEachern, K.A., Nietupski, J.B., Chuang, W.L., Armentano, D., Johnson, J., Hutto, E., Grabowski, G.A., Cheng, S.H., and Marshall, J. (2006). AAV8-mediated expression of glucocerebrosidase ameliorates the storage pathology in the visceral organs of a mouse model of Gaucher disease. J Gene Med *8*, 719-729.
McMahon, T.L., Wilczynska, Z., Barth, C., Fraser, D.J., Pontes, L., and Fisher, P.R. (1996). Replicon rescue: a novel strategy to clone the genomic DNA flanking insertions of integrating shuttle vector DNA. Nucleic Acids Res *24*, 4096-4097.
Metzker, M.L. (2009). Sequencing technologies - the next generation. Nat Rev Genet *11*, 31-46.
Miao, C.H., Snyder, R.O., Schowalter, D.B., Patijn, G.A., Donahue, B., Winther, B., and Kay, M.A. (1998). The kinetics of rAAV integration in the liver. Nat Genet *19*, 13-15.
Michelfelder, S., and Trepel, M. (2009). Adeno-associated viral vectors and their redirection to cell-type specific receptors. Adv Genet *67*, 29-60.
Miller, D.G., Petek, L.M., and Russell, D.W. (2003). Human gene targeting by adeno-associated virus vectors is enhanced by DNA double-strand breaks. Mol Cell Biol *23*, 3550-3557.
Miller, D.G., Petek, L.M., and Russell, D.W. (2004). Adeno-associated virus vectors integrate at chromosome breakage sites. Nat Genet *36*, 767-773.
Miller, D.G., Rutledge, E.A., and Russell, D.W. (2002). Chromosomal effects of adeno-associated virus vector integration. Nat Genet *30*, 147-148.
Miller, D.G., Trobridge, G.D., Petek, L.M., Jacobs, M.A., Kaul, R., and Russell, D.W. (2005). Large-scale analysis of adeno-associated virus vector integration sites in normal human cells. J Virol *79*, 11434-11442.
Miller, D.G., Wang, P.R., Petek, L.M., Hirata, R.K., Sands, M.S., and Russell, D.W. (2006). Gene targeting in vivo by adeno-associated virus vectors. Nat Biotechnol *24*, 1022-1026.

Mingozzi, F., and High, K.A. (2007). Immune responses to AAV in clinical trials. Curr Gene Ther *7*, 316-324.
Mingozzi, F., and High, K.A. (2011). Therapeutic in vivo gene transfer for genetic disease using AAV: progress and challenges. Nat Rev Genet *12*, 341-355.
Mingozzi, F., Maus, M.V., Hui, D.J., Sabatino, D.E., Murphy, S.L., Rasko, J.E., Ragni, M.V., Manno, C.S., Sommer, J., Jiang, H., *et al.* (2007). CD8(+) T-cell responses to adeno-associated virus capsid in humans. Nat Med *13*, 419-422.
Mitchell, A.M., Nicolson, S.C., Warischalk, J.K., and Samulski, R.J. (2010). AAV's anatomy: roadmap for optimizing vectors for translational success. Curr Gene Ther *10*, 319-340.
Mitchell, G.A., Grompe,M., Lambert, M., and Tanguay, R.M. (2001). Hypertyrosinemia (New York, MacGraw-Hill).
Mitchell, M.C., and Herlong, H.F. (1986). Alcohol and nutrition: caloric value, bioenergetics, and relationship to liver damage. Annu Rev Nutr *6*, 457-474.
Mitchell, R.S., Beitzel, B.F., Schroder, A.R., Shinn, P., Chen, H., Berry, C.C., Ecker, J.R., and Bushman, F.D. (2004). Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. PLoS Biol *2*, E234.
Moghrabi, N.N., Johnson, M.A., Yoshitomi, M.J., Zhu, X., Al-Dhalimy, M.J., Olson, S.B., Grompe, M., and Richards, C.S. (2009). Validation of Fanconi anemia complementation Group A assignment using molecular analysis. Genet Med *11*, 183-192.
Mohan, N., McKiernan, P., Preece, M.A., Green, A., Buckels, J., Mayer, A.D., and Kelly, D.A. (1999). Indications and outcome of liver transplantation in tyrosinaemia type 1. Eur J Pediatr *158 Suppl 2*, S49-54.
Monnat, R.J., Jr., Hackmann, A.F., and Cantrell, M.A. (1999). Generation of highly site-specific DNA double-strand breaks in human cells by the homing endonucleases I-PpoI and I-CreI. Biochem Biophys Res Commun *255*, 88-93.
Montini, E., Held, P.K., Noll, M., Morcinek, N., Al-Dhalimy, M., Finegold, M., Yant, S.R., Kay, M.A., and Grompe, M. (2002). In vivo correction of murine tyrosinemia type I by DNA-mediated transposition. Mol Ther *6*, 759-769.
Muramatsu, S., Mizukami, H., Young, N.S., and Brown, K.E. (1996). Nucleotide sequencing and generation of an infectious clone of adeno-associated virus 3. Virology *221*, 208-217.
Muzyczka, N., Samulski, R.J., Hermonat, P., Srivastava, A., and Berns, K.I. (1984). The genetics of adeno-associated virus. Adv Exp Med Biol *179*, 151-161.
Nakai, H., Fuess, S., Storm, T.A., Muramatsu, S., Nara, Y., and Kay, M.A. (2005a). Unrestricted hepatocyte transduction with adeno-associated virus serotype 8 vectors in mice. J Virol *79*, 214-224.
Nakai, H., Herzog, R.W., Hagstrom, J.N., Walter, J., Kung, S.H., Yang, E.Y., Tai, S.J., Iwaki, Y., Kurtzman, G.J., Fisher, K.J., *et al.* (1998). Adeno-associated viral vector-mediated gene transfer of human blood coagulation factor IX into mouse liver. Blood *91*, 4600-4607.
Nakai, H., Montini, E., Fuess, S., Storm, T.A., Grompe, M., and Kay, M.A. (2003). AAV serotype 2 vectors preferentially integrate into active genes in mice. Nat Genet *34*, 297-302.

Nakai, H., Storm, T.A., and Kay, M.A. (2000). Recruitment of single-stranded recombinant adeno-associated virus vector genomes and intermolecular recombination are responsible for stable transduction of liver in vivo. J Virol *74*, 9451-9463.
Nakai, H., Wu, X., Fuess, S., Storm, T.A., Munroe, D., Montini, E., Burgess, S.M., Grompe, M., and Kay, M.A. (2005b). Large-scale molecular characterization of adeno-associated virus vector integration in mouse liver. J Virol *79*, 3606-3614.
Nakai, H., Yant, S.R., Storm, T.A., Fuess, S., Meuse, L., and Kay, M.A. (2001). Extrachromosomal recombinant adeno-associated virus vector genomes are primarily responsible for stable liver transduction in vivo. J Virol *75*, 6969-6976.
Nathwani, A.C., Cochrane, M., McIntosh, J., Ng, C.Y., Zhou, J., Gray, J.T., and Davidoff, A.M. (2009). Enhancing transduction of the liver by adeno-associated viral vectors. Gene Ther *16*, 60-69.
Nathwani, A.C., Tuddenham, E.G., Rangarajan, S., Rosales, C., McIntosh, J., Linch, D.C., Chowdary, P., Riddell, A., Pie, A.J., Harrington, C., *et al.* (2011). Adenovirus-Associated Virus Vector-Mediated Gene Transfer in Hemophilia B. N Engl J Med.
Naylor, S.L., Carritt, B., Boileau, C., Beroud, C., Alexander, C., Allderdice, P., Alimov, A., Ashworth, T., Bonifas, J., Bugert, P., *et al.* (1996). Report of the sixth international workshop on human chromosome 3 mapping 1995. Cytogenet Cell Genet *72*, 255-270.
Niswander, L., Kelsey, G., Schedl, A., Ruppert, S., Sharan, S.K., Holdener-Kenny, B., Rinchik, E.M., Edstrom, J.E., and Magnuson, T. (1991). Molecular mapping of albino deletions associated with early embryonic lethality in the mouse. Genomics *9*, 162-169.
Overturf, K., al-Dhalimy, M., Ou, C.N., Finegold, M., and Grompe, M. (1997a). Serial transplantation reveals the stem-cell-like regenerative potential of adult mouse hepatocytes. Am J Pathol *151*, 1273-1280.
Overturf, K., al-Dhalimy, M., Ou, C.N., Finegold, M., Tanguay, R., Lieber, A., Kay, M., and Grompe, M. (1997b). Adenovirus-mediated gene therapy in a mouse model of hereditary tyrosinemia type I. Hum Gene Ther *8*, 513-521.
Overturf, K., Al-Dhalimy, M., Tanguay, R., Brantly, M., Ou, C.N., Finegold, M., and Grompe, M. (1996). Hepatocytes corrected by gene therapy are selected in vivo in a murine model of hereditary tyrosinaemia type I. Nat Genet *12*, 266-273.
Paulk, N.K., Wursthorn, K., Wang, Z., Finegold, M.J., Kay, M.A., and Grompe, M. (2010). Adeno-associated virus gene repair corrects a mouse model of hereditary tyrosinemia in vivo. Hepatology *51*, 1200-1208.
Pien, G.C., Basner-Tschakarjan, E., Hui, D.J., Mentlik, A.N., Finn, J.D., Hasbrouck, N.C., Zhou, S., Murphy, S.L., Maus, M.V., Mingozzi, F., *et al.* (2009). Capsid antigen presentation flags human hepatocytes for destruction after transduction by adeno-associated viral vectors. J Clin Invest *119*, 1688-1695.
Podsakoff, G.M., Engel, B.C., Carbonaro, D.A., Choi, C., Smogorzewska, E.M., Bauer, G., Selander, D., Csik, S., Wilson, K., Betts, M.R., *et al.* (2005). Selective survival of peripheral blood lymphocytes in children with HIV-1 following delivery of an anti-HIV gene to bone marrow CD34(+) cells. Mol Ther *12*, 77-86.
Ponder, K.P. (2006). Gene therapy for hemophilia. Curr Opin Hematol *13*, 301-307.
Preuss, S., and Pikaard, C.S. (2007). rRNA gene silencing and nucleolar dominance: insights into a chromosome-scale epigenetic on/off switch. Biochim Biophys Acta *1769*, 383-392.

Qiu, J., Cheng, F., and Pintel, D. (2006). Molecular characterization of caprine adeno-associated virus (AAV-Go.1) reveals striking similarity to human AAV5. Virology *356*, 208-216.
Raake, P.W., Hinkel, R., Muller, S., Delker, S., Kreuzpointner, R., Kupatt, C., Katus, H.A., Kleinschmidt, J.A., Boekstegers, P., and Muller, O.J. (2008). Cardio-specific long-term gene expression in a porcine model after selective pressure-regulated retroinfusion of adeno-associated viral (AAV) vectors. Gene Ther *15*, 12-17.
Rebuffat, A., Harding, C.O., Ding, Z., and Thony, B. (2009). Comparison of adeno-associated virus pseudotype 1, 2, and 8 vectors administered by intramuscular injection in the treatment of murine phenylketonuria. Hum Gene Ther *21*, 463-477.
Recchia, A., Parks, R.J., Lamartina, S., Toniatti, C., Pieroni, L., Palombo, F., Ciliberto, G., Graham, F.L., Cortese, R., La Monica, N., *et al.* (1999). Site-specific integration mediated by a hybrid adenovirus/adeno-associated virus vector. Proc Natl Acad Sci U S A *96*, 2615-2620.
Reisler, D.M., Strong, W.B., and Mosley, J.W. (1967). Transaminase levels in the postconvalescent phase of infectious hepatitis. JAMA *202*, 131-135.
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.
Ritossa, F.M., and Spiegelman, S. (1965). Localization of DNA Complementary to Ribosomal Rna in the Nucleolus Organizer Region of Drosophila Melanogaster. Proc Natl Acad Sci U S A *53*, 737-745.
Rizzuto, G., Gorgoni, B., Cappelletti, M., Lazzaro, D., Gloaguen, I., Poli, V., Sgura, A., Cimini, D., Ciliberto, G., Cortese, R., *et al.* (1999). Development of animal models for adeno-associated virus site-specific integration. J Virol *73*, 2517-2526.
Ros, J., Vilaseca, M.A., Lambruschini, N., Mas, A., Lindstedt, S., and Holme, E. (1999). NTBC as palliative treatment in chronic tyrosinaemia type I. J Inherit Metab Dis *22*, 665-666.
Rosenberg, N., and Jolicoeur, P. (1997). Retroviral Pathogenesis.
Rubbi, C.P., and Milner, J. (2003). Disruption of the nucleolus mediates stabilization of p53 in response to DNA damage and other stresses. EMBO J *22*, 6068-6077.
Russell, D.W., Miller, A.D., and Alexander, I.E. (1994). Adeno-associated virus vectors preferentially transduce cells in S phase. Proc Natl Acad Sci U S A *91*, 8915-8919.
Russell, L.B., Russell, W.L., and Kelly, E.M. (1979). Analysis of the albino-locus region of the mouse. I. Origin and viability. Genetics *91*, 127-139.
Rutledge, E.A., Halbert, C.L., and Russell, D.W. (1998). Infectious clones and vectors derived from adeno-associated virus (AAV) serotypes other than AAV type 2. J Virol *72*, 309-319.
Sakai, K., Ohta, T., Minoshima, S., Kudoh, J., Wang, Y., de Jong, P.J., and Shimizu, N. (1995). Human ribosomal RNA gene cluster: identification of the proximal end containing a novel tandem repeat sequence. Genomics *26*, 521-526.
Sambrook, j., E.F.Fritsch (1989). Molecular cloning : a laboratory manual. (Spring Harbor, N.Y., Cold Spring Harbor Laboratory).
Samulski, R.J. (2003). AAV vectors, the future workhorse of human gene therapy. Ernst Schering Res Found Workshop, 25-40.

Samulski, R.J., Berns, K.I., Tan, M., and Muzyczka, N. (1982). Cloning of adeno-associated virus into pBR322: rescue of intact virus from the recombinant plasmid in human cells. Proc Natl Acad Sci U S A *79*, 2077-2081.
Samulski, R.J., Zhu, X., Xiao, X., Brook, J.D., Housman, D.E., Epstein, N., and Hunter, L.A. (1991). Targeted integration of adeno-associated virus (AAV) into human chromosome 19. EMBO J *10*, 3941-3950.
Schmidt, M., Schwarzwaelder, K., Bartholomae, C., Zaoui, K., Ball, C., Pilz, I., Braun, S., Glimm, H., and von Kalle, C. (2007). High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). Nat Methods *4*, 1051-1057.
Schnepp, B.C., Clark, K.R., Klemanski, D.L., Pacak, C.A., and Johnson, P.R. (2003). Genetic fate of recombinant adeno-associated virus vector genomes in muscle. J Virol *77*, 3495-3504.
Schroder, A.R., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. Cell *110*, 521-529.
Schulz, A., Ort, O., Beyer, P., and Kleinig, H. (1993). SC-0051, a 2-benzoyl-cyclohexane-1,3-dione bleaching herbicide, is a potent inhibitor of the enzyme p-hydroxyphenylpyruvate dioxygenase. FEBS Lett *318*, 162-166.
Simonelli, F., Maguire, A.M., Testa, F., Pierce, E.A., Mingozzi, F., Bennicelli, J.L., Rossi, S., Marshall, K., Banfi, S., Surace, E.M., *et al.* (2009). Gene therapy for Leber's congenital amaurosis is safe and effective through 1.5 years after vector administration. Mol Ther *18*, 643-650.
Smih, F., Rouet, P., Romanienko, P.J., and Jasin, M. (1995). Double-strand breaks at the target locus stimulate gene targeting in embryonic stem cells. Nucleic Acids Res *23*, 5012-5019.
Snyder, R.O., Samulski, R.J., and Muzyczka, N. (1990). In vitro resolution of covalently joined AAV chromosome ends. Cell *60*, 105-113.
Stilwell, J.L., and Samulski, R.J. (2003). Adeno-associated virus vectors for therapeutic gene transfer. Biotechniques *34*, 148-150, 152, 154 passim.
Sun, M.S., Hattori, S., Kubo, S., Awata, H., Matsuda, I., and Endo, F. (2000). A mouse model of renal tubular injury of tyrosinemia type 1: development of de Toni Fanconi syndrome and apoptosis of renal tubular cells in Fah/Hpd double mutant mice. J Am Soc Nephrol *11*, 291-300.
Taymans, J.M., Vandenberghe, L.H., Haute, C.V., Thiry, I., Deroose, C.M., Mortelmans, L., Wilson, J.M., Debyser, Z., and Baekelandt, V. (2007). Comparative analysis of adeno-associated viral vector serotypes 1, 2, 5, 7, and 8 in mouse brain. Hum Gene Ther *18*, 195-206.
Trautwein, C., Rakemann, T., Brenner, D.A., Streetz, K., Licato, L., Manns, M.P., and Tiegs, G. (1998a). Concanavalin A-induced liver cell damage: activation of intracellular pathways triggered by tumor necrosis factor in mice. Gastroenterology *114*, 1035-1045.
Trautwein, C., Rakemann, T., Malek, N.P., Plumpe, J., Tiegs, G., and Manns, M.P. (1998b). Concanavalin A-induced liver injury triggers hepatocyte proliferation. J Clin Invest *101*, 1960-1969.
Tsunoda, H., Hayakawa, T., Sakuragawa, N., and Koyama, H. (2000). Site-specific integration of adeno-associated virus-based plasmid vectors in lipofected HeLa cells. Virology *268*, 391-401.

Urosevic, M., Fujii, K., Calmels, B., Laine, E., Kobert, N., Acres, B., and Dummer, R. (2007). Type I IFN innate immune response to adenovirus-mediated IFN-gamma gene transfer contributes to the regression of cutaneous lymphomas. J Clin Invest *117*, 2834-2846.
Van Vliet, K.M., Blouin, V., Brument, N., Agbandje-McKenna, M., and Snyder, R.O. (2008). The role of the adeno-associated virus capsid in gene transfer. Methods Mol Biol *437*, 51-91.
Vandenberghe, L.H., Wang, L., Somanathan, S., Zhi, Y., Figueredo, J., Calcedo, R., Sanmiguel, J., Desai, R.A., Chen, C.S., Johnston, J., *et al.* (2006). Heparin binding directs activation of T cells against adeno-associated virus serotype 2 capsid. Nat Med *12*, 967-971.
Vandenberghe, L.H., Wilson, J.M., and Gao, G. (2009). Tailoring the AAV vector capsid for gene therapy. Gene Ther *16*, 311-319.
Wallace, H., and Birnstiel, M.L. (1966). Ribosomal cistrons and the nucleolar organizer. Biochim Biophys Acta *114*, 296-310.
Walter, J., You, Q., Hagstrom, J.N., Sands, M., and High, K.A. (1996). Successful expression of human factor IX following repeat administration of adenoviral vector in mice. Proc Natl Acad Sci U S A *93*, 3056-3061.
Wang, X., Montini, E., Al-Dhalimy, M., Lagasse, E., Finegold, M., and Grompe, M. (2002). Kinetics of liver repopulation after bone marrow transplantation. Am J Pathol *161*, 565-574.
Wang, Y., Huang, F., Cai, H., Wu, Y., He, G., and Tan, W.S. (2010). The efficacy of combination therapy using adeno-associated virus-TRAIL targeting to telomerase activity and cisplatin in a mice model of hepatocellular carcinoma. J Cancer Res Clin Oncol *136*, 1827-1837.
Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet *10*, 57-63.
Wang, Z., Zhu, T., Qiao, C., Zhou, L., Wang, B., Zhang, J., Chen, C., Li, J., and Xiao, X. (2005). Adeno-associated virus serotype 8 efficiently delivers genes to muscle and heart. Nat Biotechnol *23*, 321-328.
Weinberg, A.G., Mize, C.E., and Worthen, H.G. (1976). The occurrence of hepatoma in the chronic form of hereditary tyrosinemia. J Pediatr *88*, 434-438.
Weitzman, M.D., Young, S.M., Jr., Cathomen, T., and Samulski, R.J. (2003). Targeted integration by adeno-associated virus. Methods Mol Med *76*, 201-219.
Wold, B., and Myers, R.M. (2008). Sequence census methods for functional genomics. Nat Methods *5*, 19-21.
Wu, P., Xiao, W., Conlon, T., Hughes, J., Agbandje-McKenna, M., Ferkol, T., Flotte, T., and Muzyczka, N. (2000). Mutational analysis of the adeno-associated virus type 2 (AAV2) capsid gene and construction of AAV2 vectors with altered tropism. J Virol *74*, 8635-8647.
Wu, X., Li, Y., Crise, B., and Burgess, S.M. (2003). Transcription start regions in the human genome are favored targets for MLV integration. Science *300*, 1749-1751.
Wu, Z., Asokan, A., Grieger, J.C., Govindasamy, L., Agbandje-McKenna, M., and Samulski, R.J. (2006a). Single amino acid changes can influence titer, heparin binding, and tissue tropism in different adeno-associated virus serotypes. J Virol *80*, 11393-11397.

Wu, Z., Asokan, A., and Samulski, R.J. (2006b). Adeno-associated virus serotypes: vector toolkit for human gene therapy. Mol Ther *14*, 316-327.
Young, S.M., Jr., McCarty, D.M., Degtyareva, N., and Samulski, R.J. (2000a). Roles of adeno-associated virus Rep protein and human chromosome 19 in site-specific recombination. J Virol *74*, 3953-3966.
Young, S.M., Jr., Xiao, W., and Samulski, R.J. (2000b). Site-specific targeting of DNA plasmids to chromosome 19 using AAV cis and trans sequences. Methods Mol Biol *133*, 111-126.
Zincarelli, C., Soltys, S., Rengo, G., and Rabinowitz, J.E. (2008). Analysis of AAV serotypes 1-9 mediated gene expression and tropism in mice after systemic injection. Mol Ther *16*, 1073-1080.