Molecular Therapy
Methods & Clinical Development
Original Article

Prolonged activity of the transposase helper may raise safety concerns during DNA transposon-based gene therapy

Gergely Imre,[1,2,15] Bertalan Takács,[3,15] Erik Czipa,[4] Andrea Bakné Drubi,[1,2] Gábor Jaksa,[5] Dóra Latinovics,[6] Andrea Nagy,[1] Réka Karkas,[1,7] Liza Hudoba,[1] Bálint Márk Vásárhelyi,[1] Gabriella Pankotai-Bodó,[8] András Blastyák,[1] Zoltán Hegedűs,[9] Péter Germán,[1] Balázs Bálint,[10] Khaldoon Sadiq Ahmed Abdullah,[1,7] Anna Georgina Kopasz,[1] Anita Kovács,[11,12] László G. Nagy,[10] Farkas Sükösd,[8] Lajos Pintér,[5] Thomas Rülicke,[13] Endre Barta,[4] István Nagy,[6,14] Lajos Haracska,[3,5] and Lajos Mátés[1]

[1]Laboratory of Cancer Genome Research, Institute of Genetics, Biological Research Centre, Szeged, Hungary; [2]Doctoral School of Biology, University of Szeged, Szeged, Hungary; [3]HCEMM-BRC Mutagenesis and Carcinogenesis Research Group, Institute of Genetics, Biological Research Centre, Szeged, Hungary; [4]Department of Biochemistry and Molecular Biology, University of Debrecen, Debrecen, Hungary; [5]Delta Bio 2000 Ltd., Szeged, Hungary; [6]Seqomics Biotechnology Ltd, Mórahalom, Hungary; [7]Doctoral School of Multidisciplinary Medical Sciences, University of Szeged, Szeged, Hungary; [8]Institute of Pathology, University of Szeged, Szeged, Hungary; [9]Laboratory of Bioinformatics, Biological Research Centre, Szeged, Hungary; [10]Laboratory of Fungal Genomics and Evolution, Institute of Biochemistry, Biological Research Centre, Szeged, Hungary; [11]Wenzhou-Kean University, Wenzhou, China; [12]Wenzhou Institute, Chinese Academy of Sciences, Wenzhou, China; [13]Institute of Laboratory Animal Science, University of Veterinary Medicine Vienna, Vienna, Austria; [14]Sequencing Platform, Biological Research Centre, Szeged, Hungary

DNA transposon-based gene delivery vectors represent a promising new branch of randomly integrating vector development for gene therapy. For the side-by-side evaluation of the *piggyBac* and *Sleeping Beauty* systems—the only DNA transposons currently employed in clinical trials—during therapeutic intervention, we treated the mouse model of tyrosinemia type I with liver-targeted gene delivery using both transposon vectors. For genome-wide mapping of transposon insertion sites we developed a new next-generation sequencing procedure called streptavidin-based enrichment sequencing, which allowed us to identify approximately one million integration sites for both systems. We revealed that a high proportion of *piggyBac* integrations are clustered in hot regions and found that they are frequently recurring at the same genomic positions among treated animals, indicating that the genome-wide distribution of *Sleeping Beauty*-generated integrations is closer to random. We also revealed that the *piggyBac* transposase protein exhibits prolonged activity, which predicts the risk of oncogenesis by generating chromosomal double-strand breaks. Safety concerns associated with prolonged transpositional activity draw attention to the importance of squeezing the active state of the transposase enzymes into a narrower time window.

INTRODUCTION

The successful introduction of therapeutic genes into target cells is a prerequisite for classical gene therapy. For achieving long-term gene expression and clinical benefit, integration of these therapeutic genes to the chromosomal DNA of the target cells is also required. From the early stages of gene therapy, viral vectors have represented the majority of vehicles for transferring genes to human cells. This is due to their high transduction rate and the ability to stably integrate their cargo into the host genome. Most of the integrating viral vectors used for gene therapy originate from γ-retroviruses and more recently from lentiviruses (according to the database of the *Journal of Gene Medicine*, http://www.abedia.com/wiley/).

Insertional mutagenesis is a natural outcome of random vector integrations. As a consequence, during the first decade of the 2000s, insertional oncogenesis has emerged as a major limitation in retrovirus-based gene therapy protocols treating primary immunodeficiencies.[1] In addition to the potential disruption of host genetic elements, an even more dangerous phenomenon, the transcriptional deregulation of genes near the sites of vector integrations has been shown to contribute greatly to insertional oncogenesis. This phenomenon is caused by genetic elements present on the vector or on the cargo DNA. Virus-specific transcriptional enhancers within the retroviral long terminal repeats have been identified as one of the major determinants of nearby host gene activation.[2] Thus, two factors, biased genomic vector integration and the potential of the vector to transcriptionally deregulate nearby genes, collectively increase the likelihood of insertional oncogenesis. Interestingly, the AAV-based vectors that are presently blooming in popularity may also be affected by insertional oncogenesis. When using rAAV vectors, even though

Received 6 December 2022; accepted 8 March 2023;
https://doi.org/10.1016/j.omtm.2023.03.003.

[15]These authors contributed equally

Correspondence: Lajos Mátés, Laboratory of Cancer Genome Research, Institute of Genetics, Biological Research Centre, Szeged, Hungary.
E-mail: mates.lajos@brc.hu

This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

they are non-integrating by design, chromosomal integration of the vector occurs at significant frequencies producing the long-term expression of transgenes.[3] Thus, similarly to other integrating vectors, insertional oncogenesis is a concern with rAAV, as illustrated by studies where gene therapy treatments were associated with tumorigenesis in mice resulting from the integration of rAAV vectors.[4,5]

To increase vector safety, researchers started to develop new types of randomly integrating vectors. DNA transposon-based gene delivery vectors represent a promising new branch of randomly integrating vector development for gene therapy. Recently, we and others have been able to increase the gene transfer efficiency of two DNA transposon systems, the *Sleeping Beauty* (SB)[6] and the *piggyBac* (PB)[7] to the viral range. Thus, these two systems may be used as substitutes for viral vectors in most current clinical protocols.[8] Immunogenicity problems are much less pronounced for transposon vectors due to the lack of viral capsid. Chromosomal gene transfer requires only the presence of a single protein, the transposase helper. Since transposon inverted terminal repeat (ITR) sequences, especially those of the SB system do not have a significant effect on the genes in their vicinity,[9–11] the use of transposon-based vectors is also expected to reduce the risk of insertional oncogenesis.

It has long been known that the genome-wide integration pattern of randomly integrating vectors cannot be considered completely random.[12] Biased genomic vector integration patterns occur because viruses and transposons from which those gene delivery vectors of gene therapy are derived have certain preferences for their sites of integration in the host genome. The driving force behind such preferences is to ensure the proper expression of the newly integrated virus or transposon copy. These preferences in whole or in part are transmitted to the derivative gene delivery vectors as well. The supportive or inhibitory nature, regarding expression, of a given chromosomal locus is determined by a variety of chromatin resident factors such as transcriptional regulators and histone modifications.

In this study, to better understand DNA transposon-mediated therapeutic gene delivery and to assess its risks we carried out a high-resolution analysis of the PB and SB transposon integration sites during the gene therapeutic treatment of the mouse model of tyrosinemia type I (TT1).[13] We applied the SB1006 and hyperPB7 transposases, as they provide hyperactive gene delivery systems that may replace viral vectors in the majority of clinical protocols. A prerequisite for conducting the analysis at the appropriate depth was the development of a new next-generation sequencing (NGS) procedure that is well suited to locate vector integration sites from very heterogeneous samples. Using this procedure, we were able to detect up to 120,000 transposon integration events from a single gene therapy-treated organ. We identified and examined approximately one million integration sites for both systems. We explored several interesting and hitherto unknown features of the mode of action of the PB and SB systems in mouse hepatocytes such as hot regions affected by a high frequency of vector integration and the clustering of integrations in larger genomic regions, indicating the occurrence of "local hopping."[14] From the analysis of the genome-wide distribution of vector integration sites, it was clear that the integration pattern of the SB system is much closer to random.

Comparing the supporting NGS read numbers of vector integrations in the short- and long-term monitored groups of PB- and SB-treated experimental animals we revealed no signs of insertional oncogenesis using either transposon system. However, using time-shifted delivery experiments of the transposase and transposon components we also proved that, upon introduction to the mouse liver, the PB system exhibits prolonged transpositional activity as compared with SB. This feature of the PB system represents a potential new risk factor for DNA transposons, the underlying hazards of which are not yet fully recognized. This work offers a potential mechanism to explain the adverse events seen in the clinical trial of first-in-human administration of PB-modified chimeric antigen receptor (CAR) T cells.[15,16]

RESULTS

Hyperactive transposon-based *in vivo* gene therapeutic treatment of TT1 mice

For the side-by-side evaluation of the two hyperactive DNA transposon systems potentially applicable in gene therapy, we chose to treat the preclinical mouse model of the human TT1 disease[17] using both types of vectors. TT1 is a fatal inherited disease with progressive liver dysfunction, which in humans is caused by germline mutations in the fumarylacetoacetate hydrolase (*FAH*) gene. The *FAH* gene product is the last enzyme of the tyrosine degradation pathway, and in its absence hepatocytes are killed due to the accumulation of succinylacetone and other harmful metabolites arising from the blocked tyrosine catabolism (supplemental material 1: Figure S1A). *Fah*-deficient mice, if left untreated, also die neonatally due to liver dysfunction. The fatal liver phenotype of the TT1 disease can be cured by a drug called nitisinone (NTBC). NTBC blocks tyrosine catabolism upstream of *Fah* and thereby prevents the buildup of hepatotoxic metabolites[18,19] (supplemental material 1: Figure S1A). We created a *Fah* gene knockout (KO) mouse strain that carries a deletion spanning over exons 2 to 5, and exhibits a TT1 phenotype similar to that of the original KO strain[13] (supplemental material 1: Figures S1B and S1C).

In vivo transposon-based gene therapy was applied to treat *Fah* KO mice using either the PB or the SB vector system. In both cases, the *Fah* therapeutic gene was inserted between the transposon ITRs and the hyperactive transposase helper was administered as a separate plasmid. We treated 12 mice using each type of vector, of which 6 animals were sacrificed at 2 months after treatment (short-term monitoring) and the remaining 6 animals at 7 months after treatment (long-term monitoring) (supplemental material 1: Table S1). The transposon and the transposase plasmid components of both systems were hydrodynamically delivered into the liver of *Fah* KO mice, then NTBC was withdrawn (Figure 1A). Control mice hydrodynamically injected only with transposon plasmids died after NTBC withdrawal, whereas those injected with either of the

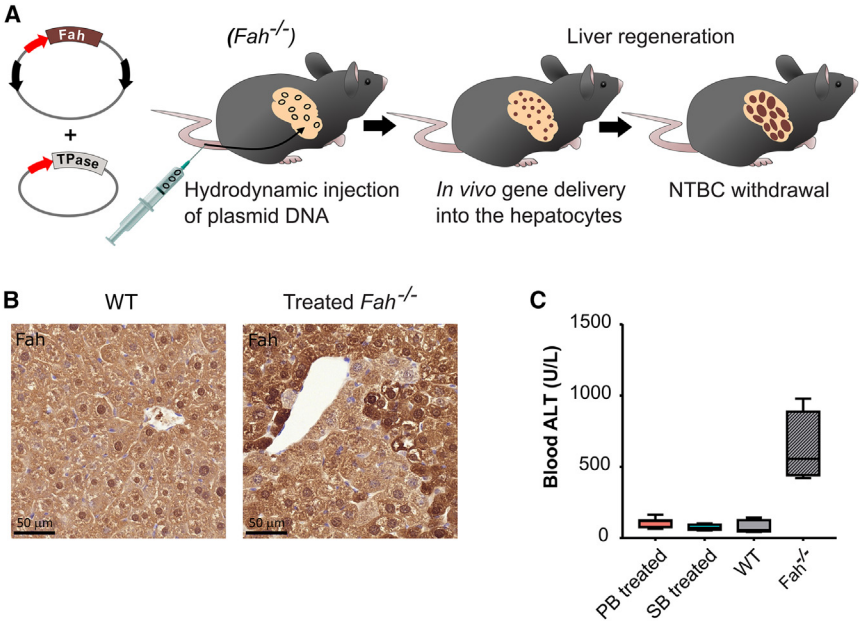

Figure 1. *In vivo* transposon-based gene therapeutic treatment of TT1 mice

(A) Schematic diagram of animal treatment. (B) Fah immunostaining of liver sections 2 months after NTBC withdrawal. (C) Alanine aminotransferase (ALT) levels measured in blood samples.

transposon/transposase cocktails started to regain weight after a short period of weight loss, and soon reached the weight of healthy controls (supplemental material 1: Figure S1D). Hydrodynamic plasmid delivery primarily targets hepatocytes by the enhancement of their membrane permeability.[20] In those hepatocytes where hydrodynamic transfection was successful, the hyperactive transposase helper enzymes could catalyze the "cut and paste" transposition reactions[21] (supplemental material 1: Figure S2); the excision of the transposon followed by its integration into the host chromosomes. Already after 2 months, most hepatocytes were Fah-positive because *Fah* KO livers exerted selection pressure for genetically corrected Fah-expressing hepatocytes[20] (Figure 1B). Serum levels of alanine aminotransferase (ALT) also confirmed the healthy state of liver tissue in treated animals (Figure 1C). Remarkably, we saw no sign of tumorigenesis in the course of either short- or long-term monitoring of the experimental animals.

NGS identification of vector integration sites revealed that the genome-wide distribution of SB integration events is closer to random

At the end of the monitoring periods, genomic DNA was isolated from the 12 PB- and the 12 SB-treated livers. Next, using a qPCR method, we identified the average therapeutic gene dose in the organs as 1, 2 and 1,83 copies per diploid genome for the PB and for the SB systems, respectively (Figure 2A).

To more accurately assess the genome-wide consequences of therapeutic gene delivery mediated by hyperactive transposon systems, we identified vector insertion sites from the 12 PB- and the 12 SB-treated livers. For this purpose, we developed a novel NGS approach designated as streptavidin-based enrichment sequencing (SBE-seq), the key element of which was the efficient enrichment of vector-genome joint sequences using streptavidin-coated paramagnetic beads (supplemental material 1: Figure S3). The efficiency of SBE-seq was very similar for both transposon systems as the data were generated at very similar supporting read numbers per integration (supplemental material 1: Figure S4A). Using SBE-seq we found on average ~85,000 vector insertion sites mapping to unique genomic positions in a single treated organ in both cases (Figure 2B), and a total of approximately one million sites were identified for the PB and for the SB transposon systems as well (Table S3). We focused our analysis only on uniquely mapped integrations and performed additional filtering (see materials and methods for details) to avoid possible artifacts, even due to errors or imperfections in the mouse reference genome.

Strikingly, we found 138,587 and 18,496 PB and SB vector integration sites in the mouse genome, respectively, which occurred more than once in the 12 PB- and 12 SB-treated animals. These unique genomic positions resulted in 411,712 and 37,923 recurrent vector integration events identified in the 12 PB- and 12 SB-treated animals, corresponding to 37.19% and 4.01% of all identified PB and SB integration events, respectively (supplemental material 1: Figure S4C). The one order of magnitude fewer recurrent integrations observed for the SB system suggests that the genome-wide distribution of vector integration sites upon gene delivery with this system is closer to random. The number of recurrent integrations and their number of repetitions are listed in Table S3. A single vector integration site was identified that reached the maximum detectable repetition number of 24. This means that we found vector integrations in both orientations at this genomic position in all 12 animals treated with the given gene delivery system. This integration site was identified for the PB system on Chr6 at position 71170422-71170426, approximately 6 kb upstream of the fatty acid binding protein 1 (*Fabp1*) gene, which is highly expressed in the liver and intestine.

Of the approximately one million vector integration sites identified per gene delivery system, we found 4.61% and 1.99% non-canonical integration sites for the PB and SB systems (Table S3), respectively, which corresponds approximately to the literature data published so far.[21] Non-canonical integrations refer to transposon integrations into target sequences other than the canonical TTAA (PB) and TA (SB) target sites. The sequence composition of non-canonical target sites of the PB and SB systems are illustrated by sequence logo type

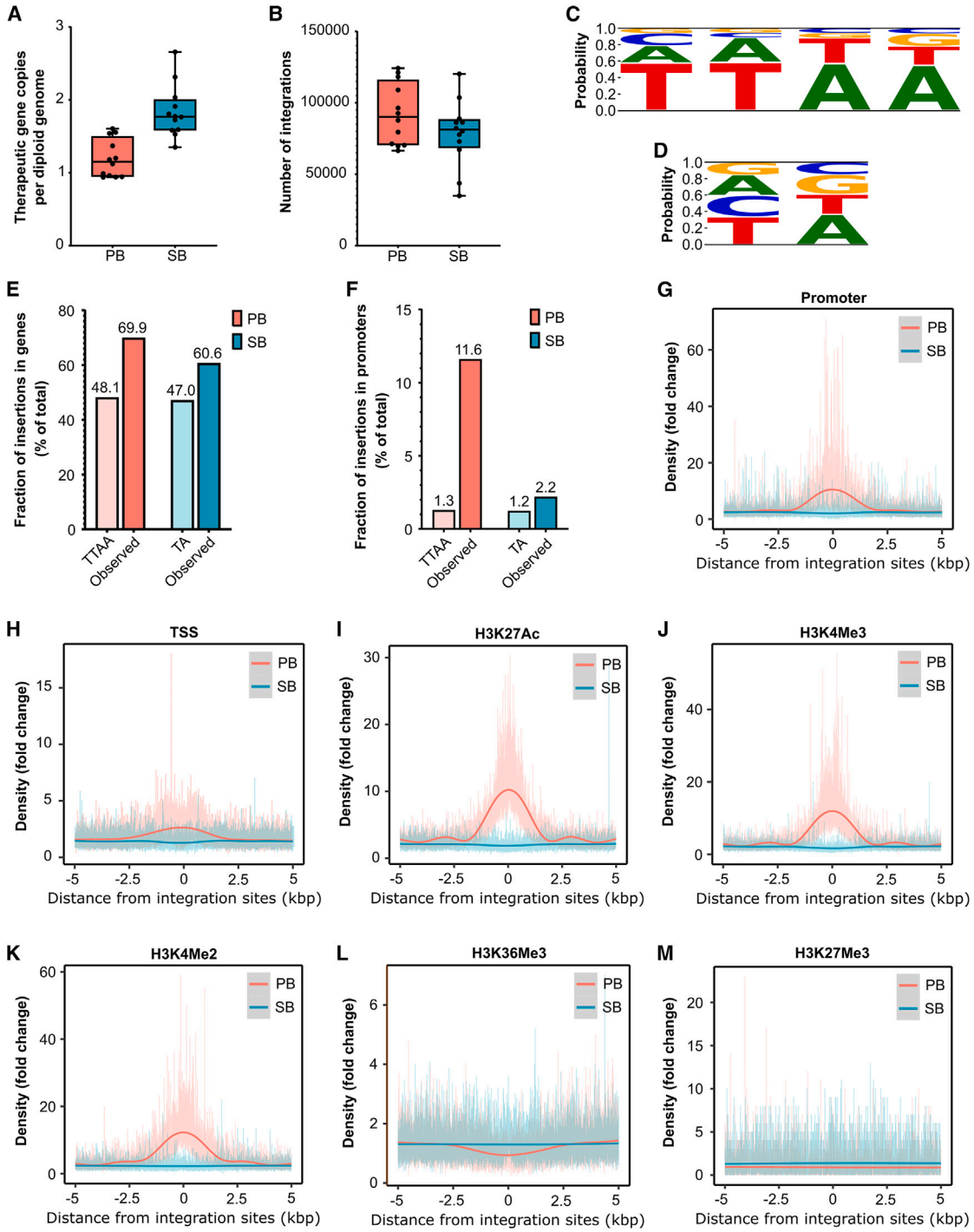

Figure 2. NGS identification of vector integration sites in the liver revealed deep features of transposon-based gene delivery

(A) Copy number of the therapeutic gene in the livers of PB- and SB-treated animals. Liver DNA samples were tested using a *Fah* therapeutic gene-specific qPCR assay. Results were normalized to measurements of the olfactory receptor 16 (*Olfr16*) gene as an input control, and values are presented relative to one diploid genome. (B) Number of vector integrations identified in the livers of PB- and SB-treated animals. (C) Sequence logo type representation of the sequence composition of non-canonical PB integration sites. (D) Sequence logo type representation of the sequence composition of non-canonical SB integration sites. (E) Fraction of vector integrations found in genes. Here, the gene body was defined as the entire gene from the transcription start site to the end of the transcript. Control values were determined from the distribution of all potential TTAA (PB) and TA (SB) target sites in the mouse genome. (F) Fraction of vector integrations found in promoters. Control values were determined from the distribution

(legend continued on next page)

representations (Figures 2C and 2D). The position weight matrices for which the sequence logos were plotted are provided in Table S3. We also detected the presence of 7.47% and 1.75% of recurrent non-canonical integration events of all non-canonical integration events identified in 12 PB- and 12 SB-treated animals, respectively (Table S3). For both systems, this is significantly lower than the recurrence rate observed in the entire population of identified integration events.

We observed specific regions in the murine hepatocyte genome that were often targeted by transpositions. To refine this picture, we defined a so-called hot region as a region with an insertion density of greater than one per kilobase, with at least five insertions in total, and a less than 1 kb maximum distance between the adjacent insertions. These criteria had to be independently met in at least two animals in the given experimental panel over the same genomic region. We identified 1,438 and 8 such hot regions for the PB and for the SB transposon systems, respectively (Table S3) and 0.07% of the PB and 12.5% of the SB hot regions were shared (overlapping) with the other system. Of all identified integrations, 5.32% and 0.03% fell into these hot regions for the PB and for the SB systems, respectively. The much less hot regions identified for the SB system is further supporting that the genome-wide distribution of vector integration sites upon gene delivery with this system is closer to random. We filtered the gene annotations overlapping hot regions (± 5 kb) using the Catalogue of Somatic Mutation in Cancer (COSMIC) Cancer Gene Census (CGC) database[22] currently containing over 700 genes with a documented role in cancer. Among the genes overlapping with hot regions, we found CGC annotations of 72 genes for the PB system and 1 for the SB system (Table S3).

We investigated the genome-wide distribution of the identified vector insertion sites, and found that 69.86% of the PB and 60.59% of the SB integrations fell into genes, which, in both cases, but especially for the PB system, is higher than the expected random distribution (Figure 2E). When looking at the subregions within genes, the most striking difference between the two gene delivery systems studied appeared to be the proportion of vector integrations falling in promoter regions. Of all PB integrations, 11.60% landed in promoter sequences and this was not due to enrichment of potential TTAA target sites in promoter sequences, while SB integrations did not show significant enrichment in the latter regions (Figure 2F). By conducting a more detailed analysis of the distribution of integration events in the vicinity of promoter regions we confirmed the strong positive correlation of PB integrations to this element (Figure 2G). However, the previously published bimodal distribution of PB integration events around transcription start sites (TSSs)[23] is, according to our study, largely due to the primary distribution of potential TTAA target sites around the TSS as the phenomenon is only seen in the absence of correction with a random TTAA control sample (Figure 2H; supplemental material 1: Figure S5A).

We also compared our integration datasets with the data of seven available ChIP-seq experiments performed on healthy mouse liver (Table S4) to determine coincidence with different histone modifications occurring in the promoter and transcribed regions of active or repressed genes. We found a strong positive correlation of PB transposon integrations with H3K27Ac and H3K4Me3 modifications (Figures 2I and 2J) accumulating in the promoter region of active genes.[24,25] The other forms of H3K4 methylations, H3K4Me2 and H3K4Me1, occur in the promoter and transcribed regions of active genes.[25] Both modifications, especially H3K4Me2, correlated positively with the integrations of the PB system (Figure 2K; supplemental material 1: Figure S5B). Interestingly, we were also able to show a negative correlation of the H3K36Me3 modifications accumulating in the transcribed regions of active genes[25] with PB integrations (Figure 2L). Finally, we investigated the H3K27Me3 and H3K9Me3 modifications that occur in the promoter and transcribed regions of repressed genes,[25] where H3K9Me3 directs DNA methylation and heterochromatin formation.[26] We found no correlation with these modifications in the vicinity of integrations of either one of the gene delivery systems examined (Figure 2M; supplemental material 1: Figure S5B).

To better assess the potential risk of insertional oncogenesis's association with the application of the hyperactive PB and SB transposon systems, we also compared the supporting NGS read numbers of vector integrations in the short- and long-term monitored experimental groups for both transposon systems. We looked for genes (± 5 kb) in the region of which the supporting read number of vector integrations were significantly increased in the long-term monitored group compared with the short-term monitored one. Such a situation would suggest a steady growth of hepatocyte colonies carrying integrations within those genes. For the SB system only four genes were found where vector integrations had more than five times higher average supportive read numbers in the group of animals monitored on the long-term and at least one individual integration had a supportive read count above 300 (Table S3). None of these genes are recorded in the CGC database. No genes meeting these criteria were found for the PB system.

Delayed transposon introduction proved the prolonged activity of the PB system

It is well known that transposons remobilized from chromosomal positions tend to land into *cis*-linked sites in the vicinity of the donor

of all potential TTAA (PB) and TA (SB) target sites in the mouse genome. (G) Distribution of promoter elements in the ±5 kb environment of integration sites. From (G–M) the values at each point were divided by the *in silico* control values to obtain the fold change. Control values were determined by examining randomly selected sets of potential PB and SB target sites of similar size. (H) Distribution of TSS elements in the ±5 kb environment of integration sites. (I) Distribution of H3K27Ac histone modifications in the ±5 kb environment of integration sites. (J) Distribution of H3K4Me3 histone modifications in the ±5 kb environment of integration sites. (K) Distribution of H3K4Me2 histone modifications in the ±5 kb environment of integration sites. (L) Distribution of H3K36Me3 histone modifications in the ±5 kb environment of integration sites. (M) Distribution of H3K27Me3 histone modifications in the ±5 kb environment of integration sites.

locus. This phenomenon is called "local hopping" and is common to both the PB[27] and the SB[28] transposons. Although during *in vivo* therapeutic gene delivery transpositions primarily start from the transposon donor plasmids, the presence of local hopping may indicate differences in the timing of transpositions between the two systems. Therefore, we investigated vector insertion density in 0.5 Mb windows starting from each integration event, and identified those in which the density of integrations was significantly greater than predicted by binomial test. This procedure was used to test the extent of potential clustering of integration sites in chromosomal regions much larger than the hot regions defined in the previous section. The first and last integrations of the overlapping windows, for which this condition was met, gave the 5′ and 3′ endpoints of the given region showing integration clustering. To get rid of the distortion effect of hot regions, the number of integrations in the areas of the previously defined hot regions was artificially reduced to one in each animal. To control the potential effect of the strong positive correlation of PB integrations with some histone modifications, we used a histone modification-biased control dataset when evaluating clustering of PB integration sites (see materials and methods for details).

Strikingly, the number of genomic regions showing clustering of vector integrations was about 4-fold higher in the real PB and SB integration sets defined per experimental animal than in their corresponding controls (Figure 3A). Apparently, PB-treated animals had the highest clustering of identified integrations, and this was predominantly not due to the correlation of PB insertions with histone modifications (Figure 3A). The average size of regions positive for integration clustering were 407,853 and 489,407 bp for PB and SB systems, respectively. When the size distribution of the identified regions was examined in more detail, an increase in the number of regions was observed for both transposon systems in the 0–2 Mbp interval in real samples compared with controls (supplemental material 1: Figure S6). Thus, the integrations of the gene delivery systems used here are clustered in large regions of similar size to the local hopping windows, which may indicate, among other things, their long-lasting transposition activity.

To prove this, first we compared the amount of transposon excision products over time for both vector systems in gene therapeutically treated mice. Excision is the first step of transposition during which the transposon is removed from the donor plasmid by the transposase enzyme (supplemental material 1: Figure S2). Donor plasmids that have lost the transposon carry a double-strand break (DSB), which is repaired predominantly by the non-homologous end-joining (NHEJ) DNA repair pathway. The amount of such NHEJ repair products can be quantified using a specific qPCR assay (Figure 3B). We examined three time points after treatment and found a striking difference in the behavior of the two gene delivery systems, with the amount of excision products generated by the PB transposase significantly increasing from day 3 to day 7 (Figure 3C). This finding supported the theory that the PB system has a prolonged transpositional activity. Since both transposase proteins were expressed from the same type of plasmid, there is little chance that the differences observed in the timing of excision would be due to differences in transposase expression. Nevertheless, the expression of transposases was monitored at the RNA and protein levels as well. Expression of both transposases showed a similar declining tendency from day 3 at both RNA and protein levels (Figures 3E–3G). At day 7, both proteins were readily detectable and the amount of SB100 was slightly higher than that of hyperPB, while by day 14 they became barely detectable (Figure 3G). Thus, both transposase proteins are readily available 7 days after treatment, and differences in their expression cannot explain the observed differences in their excision characteristics.

To demonstrate that hyperPB still has transpositional activity at day 7 post-treatment, while SB100 does not, we performed an experiment in which the transposon plasmid carrying the therapeutic gene was introduced into the liver of experimental animals only on the seventh day after transposase delivery. With such an experimental setup, successful TT1 treatment can only be expected if the given transposase protein is still capable of carrying out transpositions on day 7 following its introduction. Ten TT1 mice were treated with both vector systems accordingly. The treatment of five mice was successful in the transposon-delayed PB panel, whereas only one animal survived in the SB panel. We identified vector insertion sites from the corresponding five PB- and one SB-treated livers using SBE-seq. We found on average ~25,000 vector insertion sites in the PB-treated livers, whereas only 4,893 integrations were identified in the liver of the single surviving SB-treated animal (Figure 3H). The depth of sequencing was again very similar for the two transposon systems, as data were generated at very similar supporting read numbers per integration (supplemental material 1: Figure S4B). Our results clearly demonstrated that the PB system exhibits prolonged transpositional activity relative to SB. This increases the likelihood of PB transpositions starting from transposon donors of already chromosomal location. Such transpositions lead to the formation of chromosomal DSBs,[29] which, if incorrectly repaired, can result in chromosomal deletions.

DISCUSSION

Even in TT1, gene therapy is an important treatment option, although NTBC medication is possible. NTBC treatment can improve prognosis and quality of life of TT1 patients, but many late complications persist.[30] In rodent models, the fatal liver phenotype of the disease can be managed with gene therapy interventions that rely on hydrodynamic nucleic acid delivery and primarily target the liver.[31–33] When choosing a preclinical model, we also kept in mind that in this rodent model, the *Fah* therapeutic gene is under positive selection pressure and that extensive liver regeneration is initiated from *Fah*-corrected cells (Figures 1A and 1B). As a result, the number of cells carrying the original gene delivery events is multiplied by cell divisions, which facilitates the detection of vector integrations. In addition, it may even allow the tracking of the translocation of the original integration events in the progeny cells, such as the detection of local hopping.

Among gene therapy procedures involving effective randomly integrating vectors, the transposon-based technology is a promising

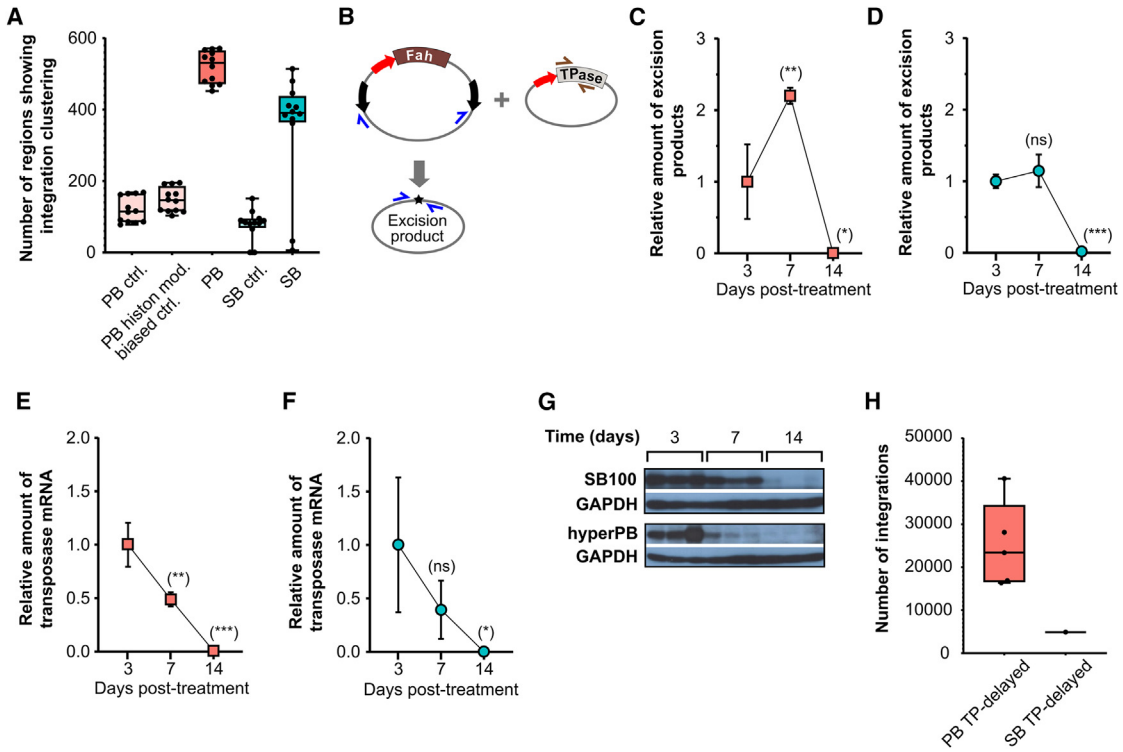

Figure 3. The PB system exhibits prolonged transposition activity compared with SB

(A) The number of genomic regions showing vector integration clustering in treated livers. Control values were determined by examining randomly selected sets of potential PB and SB target sites of similar size. Of elements of the histone modification-biased PB control set, 29% were collected from genomic regions harboring the histone modifications preferred by the PB system. Individual data points were plotted on a box diagram (see Table S5 for individual data values and statistics). (B) Schematic representation of plasmid constructs and primers used for qPCR assays. Red arrows, promoters; black arrows, transposon ITRs; TPase, CDS of transposase helper proteins. (C) Kinetics of PB transposon excision following *in vivo* gene delivery. Liver DNA samples at 3, 7, and 14 days post-treatment were tested using PB excision product specific qPCR assay. Results were normalized to measurements of the olfactory marker protein (*Omp*) gene as input control and data are presented as the mean ± SD (n = 3) of relative values compared with the value of day 3 (see Table S5 for individual data values and statistics). (D) Kinetics of SB transposon excision following *in vivo* gene delivery. Liver DNA samples at 3, 7, and 14 days post-treatment were tested using SB excision product specific qPCR assay. Results were normalized to measurements of the olfactory marker protein (*Omp*) gene as input control and data are presented as the mean ± SD (n = 3) of relative values compared with the value of day 3 (see Table S5 for individual data values and statistics). (E) Decay kinetics of PB transposase mRNA following *in vivo* gene delivery. Liver RNA samples at 3, 7, and 14 days post-treatment were tested using PB transposase-specific qRT-PCR assay. Results were normalized to measurements of the ribosomal protein L27 (*Rpl27*) transcript as input control and data are presented as the mean ± SD (n = 3) of relative values compared with the value of day 3 (see Table S5 for individual data values and statistics). (F) Decay kinetics of SB transposase mRNA following *in vivo* gene delivery. Liver RNA samples at 3, 7, and 14 days post-treatment were tested using SB transposase-specific qRT-PCR assay. Results were normalized to measurements of the ribosomal protein L27 (*Rpl27*) transcript as input control and data are presented as the mean ± SD (n = 3) of relative values compared with the value of day 3 (see Table S5 for individual data values and statistics). (G) Decay kinetics of transposase proteins following *in vivo* gene delivery. Liver extracts prepared at 3, 7, and 14 days post-treatment were tested by western blot assay for the presence of either HA-tagged PB or HA-tagged SB transposase proteins. Transposase proteins were identified based on their anti-HA reactivity. GAPDH served as a loading control. (H) Number of vector integrations identified in the livers of mice treated with delayed transposon introduction (see Table S5 for individual data values and statistics).

option as it is expected to reduce the risk of insertional oncogenesis. Therefore, we carried out side-by-side comparison of PB and SB transposon-based gene therapeutic treatment of TT1 mice. Based on our detailed molecular biology and bioinformatics analysis, we assessed the consequences and potential risk factors of the hyperactive transposon-mediated gene delivery. Our conclusions will be applicable to other transposon-based gene therapy protocols.

Central to our approach was that we performed high resolution analysis of vector integration sites to assess stable chromosomal gene delivery properties of the applied SB1006 and hyperPB7 hyperactive transposases. Mass identification of vector integration sites following hyperactive transposon-mediated gene therapy in the TT1 model is challenging due to the multi-colonial structure of the liver. At the end of the regeneration period, about 2 months after therapeutic gene delivery, a treated organ is made up of 50–100,000 *Fah*-corrected hepatocyte colonies (Figure 1B), each consisting of 1–2,000 cells, yielding the entire murine hepatocyte population made of approximately 100,000,000 cells.[34,35] The 50–100,000 hepatocyte colonies each carry distinct—but not necessarily unique—vector integration events. As a result, the prevalence of an individual vector integration event in the liver DNA samples is approximately 100,000 times lower

than the prevalence of any unique endogenous locus. Therefore, although similar techniques have already been described for the analysis of less heterogeneous tumor samples,[36] we had to set up a new technology to be able to examine the regenerated liver samples at sufficient depth. Our NGS procedure, SBE-seq, is designed to efficiently locate vector integration sites in highly heterogeneous samples, and can be used with small modifications to mass determine integration sites of any randomly integrating vector, whether from preclinical models or clinical specimens. This is indicated by the fact that similar NGS read numbers were generated when similar high numbers of integrations of the two transposon systems were determined in parallel (supplemental material 1: Figure S4A).

We identified approximately one million integration sites for both systems and by their bioinformatics analysis we uncovered fundamental aspects of therapeutic gene delivery, addressing a number of factors that directly affect the safety of gene therapy. Strikingly, we found that 37.19% and 4.01% of all identified vector integration events occurred more than once in the 12 PB- and 12 SB-treated animals, respectively (supplemental material 1: Figure S4C). In a given sample from a given animal, we can only distinguish two recurrent vector integrations for an individual target site, these are integrations in opposite orientations on the + and − DNA strands. This allows the unequivocal detection of a maximum of 24 (2 × 12) recurrent integration events for the same genomic target site in the 12 animals treated with the same gene delivery system (Table S3). Repetitions above this number can only be assumed if the integration sites are supported with above-average numbers of NGS reads.

We also demonstrated that hot regions are of great importance, since some gene delivery systems, such as the PB transposon, integrates into them extremely often (Table S3). A small fraction of the identified hot regions is shared between the two transposon systems studied here, but we found two orders of magnitude more hot regions of the PB system. For the PB system, hot regions have already been described by others, but with much lower resolution and not in gene therapeutically relevant cell types.[37] We assume that the expression of host genes may be altered by close vector integrations. Therefore, the large number of integrations found in the hot regions justified the need to assess whether the altered expression of genes found in these regions could pose the risk of cancer development in hepatocytes. Compared with the CGC database, PB hot regions identified in the mouse liver indicate a minor risk of carcinogenesis (Table S3). We believe that it would be advisable to subject all gene therapy protocols relying on randomly integrating vectors to similar investigations.

Both the one order of magnitude more recurrent integrations and the two orders of magnitude more hot regions observed in the case of the PB gene delivery system indicate that SB integrations are more evenly distributed on the genomic scale.

We found that the genome-wide integration spectrum of the PB system shows a stronger bias toward genes when compared with SB (Figure 2E), which is attributed to the frequent integrations of the PB transposon into the 5′ regulatory regions of genes (Figure 2F). The bias of the PB system toward promoters is similar to that often observed for viral gene delivery and has been established previously.[12,21,23] Although we confirmed that PB system-generated integrations are attracted toward promoters, we could not verify the previously reported bimodal distribution of PB integration events around TSSs.[23] According to our results, this phenomenon is largely due to the primary distribution of potential TTAA target sites around the TSSs as it is only seen in the absence of correction with a random TTAA control sample (Figure 2H; supplemental material 1: Figure S5A). Yoshida et al. also applied such a correction,[23] but it was not effective enough because the number of elements in both the group of integrations included in the analysis and the control group of randomly collected TTAA motifs was smaller than necessary. In our case, the number of elements in these groups was orders of magnitude larger, allowing us to draw a clear conclusion. It is also of note that the number of PB integrations landing around the TSSs does not increase significantly when comparing the promoter dataset (mus_musculus.GRCm39.Regulatory_Build.regulatory_features.20201021.gff, containing 25,131 elements) to the TSS dataset (5′ starting positions of genes from the dataset mus_musculus.GRCm39.GRCm39.104.gtf, containing 55,416 elements) with twice the number of elements (supplemental material 1: Figure S5A). Considering the overlap between the promoter and TSS datasets, and the fact that the TSS dataset contains a large number of predictions, it can be concluded that a significant proportion of the predicted TSSs do not attract PB integrations, which questions the validity of the TSS predictions.

We found a significant positive correlation of PB transposon integrations with histone marks of the promoter region of actively transcribed genes (Figures 2I–2K; supplemental material 1: Figure S5B) and a negative correlation with the H3K36Me3 modification accumulating in the transcribed regions of such genes (Figure 2L). This demonstrates that, although PB integrations are biased toward genes, most of the gene body is less attractive to them, while the 5′ regulatory regions are strongly favored. These preferences of the PB system, similarly to our findings, appeared in the studies of others.[21,23,38] No correlation with any histone modification was confirmed for the SB system, which also suggests that the distribution of SB integrations is more random on the genomic scale.

Our analyses showed that increasing the monitoring period from 2 to 7 months did not substantially increase the number of reads supporting vector integrations in the vicinity of any gene in the case of any of the transposon systems tested. Notably, for the PB system, no significant read number increase in the long-term monitored group was observed even for those 72 genes that overlapped with identified PB hot regions and had a CGC record. Only for the SB system could we identify 4 genes—none is recorded in the CGC database—in the vicinity of which we detected a moderate NGS read number increase. For all 4 genes, we identified 1 integration with a read support of barely over 300 reads in the group of long-term monitored animals (Table S3). These were only 6–8 times higher values than the average

of about 50 reads/integration experienced across the SBE-seq libraries (supplemental material 1: Figure S4A; Table S3). In the presence of a malignantly transformed hepatocyte colony, we would presumably need to find much higher read counts. These results indicate that the genomic integration patterns of none of the transposon systems investigated result in insertional oncogenesis during the current study period despite the intense cell division characteristic of the treatment of TT1 mice. Detection of insertional oncogenesis in mice is presumably problematic due to the short lifespan of the species. Nevertheless, if any of the transposon systems used here had integration hotspots near aggressive liver-specific proto-oncogenes—similar to those observed for retroviral vectors in the hematological compartment[1]—the phenomenon would presumably be detectable. This is also supported by the fact that direct delivery of aggressive drivers to the liver, such as the RasG12V variant, causes very rapid tumor development within 1–2 months.[39,40]

Previously it was shown that, upon NTBC withdrawal, a subpopulation of *Fah* KO hepatocytes can escape the selection process by activating the survival Akt pathway.[41] Presumably, this phenomenon is responsible for the emergence of tumors from non-corrected Fah-deficient cells reported during retroviral gene therapy treatment of TT1 mice.[42] The fact that we did not experience the appearance of these background tumors may be due to two possible reasons. On the one hand, the monitoring time was shorter in our case (here age 2–7 months versus up to 1 year[42]). On the other hand, we used hyperactive transposase helpers providing very robust gene delivery, which may result in faster elimination of non-corrected hepatocytes from the treated livers.

We also revealed that PB and SB transposon integrations are about four times more frequently clustered in large regions than their corresponding controls and that PB integrations showed the highest rate of such clustering (Figure 3A). Surplus genomic regions with integration clustering emerging in real samples fall in the 0–2 Mbp size range, comparable with local hopping windows (supplemental material 1: Figure S6). The region increment observed in the real samples compared with the controls shows an exponentially decreasing tendency with length (supplemental material 1: Figure S6), which may correspond to the availability of chromosomal target sites accessible to the locally generated transposition complex during genomic transposon remobilization. This suggests that the main determinant of the size of the local hopping window is the availability of local chromatin in the vicinity of transposon donor sites, and that it may exhibit weak transposon system specificity with respect to window size.

The clustering of integration sites measured by our bioinformatics method is presumably not due to local hopping alone. Other undiscovered features of the transposon systems may lead to similar clusterization. However, local hopping is certainly part of the explanation, as confirmed by the still significant activity of PB transposase at day 7 post-injection. In the case of SB transposase, the activity at day 7 post-injection may no longer be sufficient to induce significant local hopping. However, it is possible that the local hopping component is significant for the SB system as well, but it occurs in the period before the time point tested in our experiments, i.e., during the first few days following hydrodynamic injection. Further experimental work will be needed to confirm this.

Prolonged transpositional activity increases the likelihood of remobilization of genomic transposon copies, which leads, on the one hand, to the clustered distribution of transposons[43] and, on the other hand, to the formation of chromosomal DSBs with small overhangs derived from the element's ITRs.[44] DSBs are particularly toxic DNA lesions that must be repaired to preserve chromosomal integrity, but attempts by cells to repair them can sometimes lead to extensive rearrangements. Potentially, the main source of large chromosomal deletions at the sites of DSBs is the action of the single-strand-annealing DNA repair pathway, but occasionally other pathways may also contribute to the formation of similar rearrangements.[45] Therefore, alternative DNA repair events may result in an excision site structure that is different from the expected transposon footprint generated by the dominant NHEJ DNA repair pathway (supplemental material 1: Figure S2).[46] For evaluating side effect profiles it is also important to keep in mind that the microstructure of DSBs generated by the two gene delivery systems used here is different and that they might not trigger the same proportion of distinct DNA repair events, causing or not causing genome rearrangements.

Such so-called imprecise excision events of the PB transposon were already shown to be able to generate deletions in *Drosophila melanogaster*.[47] Accordingly, this potentially carcinogenic side effect may also occur in gene therapy protocols when PB transposon-based gene delivery is applied. There is no direct evidence of this yet, but clinical trials of first-in-human administration of CAR T cells produced by the PB transposon system have reported the development of CAR T cell lymphomas.[15,16] In-depth genetic studies of these product-derived lymphomas showed high transgene copy number, but no insertions into typical oncogenes. Altered gene copy numbers unrelated to the vector insertion sites were also revealed and global changes in transcription correlated with such alterations.[16] Given these findings, insertional oncogenesis has not been confirmed as a mode of tumor induction. From the findings so far, it has been concluded that the over-dosing of components of the PB system might be dangerous as it can lead to very high therapeutic gene copy numbers, but further work is needed to understand the exact tumor induction mechanism at work.[16,48] It is important to note that CAR T cell malignancy has not previously been documented following administration of CAR T cells produced using SB transposon vector in a very similar clinical trial setting.[49,50]

The mechanism we propose here may potentially explain both the observed genetic changes and tumor induction during the first clinical trial of PB-based CAR T cell administration.[16] The observed copy-number losses may simply be the result of generated deletions. The occurrence of copy-number gains could possibly be explained by prolonged PB transposase activity as well if the possibility of "hybrid element transposition"[51] is taken into account. During this process

the transposon ITRs may come from separate elements into the transposition reaction rather than from a single element (the 5′ ITR of one element pairs with the 3′ ITR of another one from a different genomic position). This may result in different chromosomal rearrangements and, with the involvement of the appropriate DNA repair pathways, copy-number gains may also be generated. Subsequently, events with driver properties could be selected and propagated during tumorigenesis.

The period of active transpositions may be reduced by administration of the transposase helper in protein form. This has recently become possible with both transposon systems used here.[52,53] However, our qPCR data demonstrate that transposase expression was rapidly shut down in the liver, and the observed prolonged transpositional activity is due to the intrinsic properties of the PB transposase protein itself. Similar rapidly declining kinetics of transposase expression was also found by others during *in vivo* liver gene delivery experiments.[54] According to our results, the transposase protein is detectable in both cases for more than 1 week, but SB is rapidly inactivated. The difference in the enzymatically active time windows may be explained by the different inactivation mechanisms, which is also indicated by the different nuclear distribution patterns that occur during overexpression of the two transposases.[55] During the first clinical trial of PB-based CAR T cells, the transposase helper was provided as an mRNA to avoid unwanted genomic insertions of the transposase CDS.[16] Presumably, the mRNA-based transposase administration used could not sufficiently alleviate the issue of prolonged transposase activity that we have highlighted here. This also suggests that this phenomenon may be related to an intrinsic property of the PB transposase protein, which cannot be resolved by mRNA administration.

Prolonged activity of the transposase helper represents a potential new risk factor for DNA transposons, the underlying hazards of which are not yet fully recognized. Although carcinogenic side effects have not been observed here, perhaps due to the short generation time of the mouse model used, they are more likely to occur—even as a consequence of the generated chromosomal rearrangements—during human gene therapy interventions, as this may have been the case during the clinical trial of PB-based CAR T cells.[16] This draws attention to the importance of squeezing the gene delivery activity of the transposon systems into a narrower time window.

MATERIALS AND METHODS

Creation of the *Fah* KO mutant strain

For blastocyst injection C57BL/6N ESC clones with a reporter-tagged deletion allele of *Fah* were provided by NorCOMM. Host blastocysts were produced by superovulation of BALB/cCrl females by intraperitoneal injection with 5.0 IU of equine chorionic gonadotropin (Folligon; Intervet) and, 48 h later, with 5.0 IU of human chorionic gonadotropin (Chorulon; Intervet) followed by mating with males of the same strain. Morula stages were harvested from isolated oviducts at day 2.5 dpc and cultured in M16 medium overnight in an incubator at 37°C and 5% CO_2 to produce host blastocysts. About 10–15 ESCs were injected per blastocyst. The injected embryos were cultured for 2–3 h to recover and then transferred into the right uterus horn of 2.5 dpc pseudopregnant RjOrl:Swiss surrogate mothers as described earlier in detail.[56,57] The offspring were selected based on their chimeric coat color. High-percentage male chimeras (>80%) were bred with C57BL/6NCrl females and the offspring were selected by coat color and genotyped by PCR for the β-Geo selection marker using the primers given (supplemental material 1: Table S2) and by Southern analysis. Mutant mice were designated as C57BL/6N-*Fah*$^{tm1(NCOM)Mfgc/Biat}$ and the line is archived in the European Mouse Mutant Archive under EM:10787.

Animal care and maintenance

Mice were bred and maintained in the Central Animal House at the Biological Research Center (Szeged, Hungary). The specific pathogen-free status was confirmed quarterly according to FELASA (Federation for Laboratory Animal Science Associations) recommendations.[58] All animal experiments were conducted according to the protocols approved by the Institutional Animal Care and Use Committee at the Biological Research Center. *Fah*$^{-/-}$ mice were treated with 8 mg/L Orfadin (NTBC) (Swedish Orphan Biovitrum, Stockholm, Sweden) in drinking water. After hydrodynamic injection with therapeutic constructs NTBC was withdrawn. C57BL/6NTac wild-type mice were obtained from Taconic Biosciences (Ejby, Denmark). Animal experimental procedures related to the creation of the C57BL/6N-*Fah*$^{tm1(NCOM)Mfgc/Biat}$ mutant line were discussed and approved in the Institute of Laboratory Animal Science, University of Veterinary Medicine (Vienna, Austria) by the institutional ethics committee and granted by the national authority under license number BMWF-68.205/0258-II/3b/2011. All other animal experiments were approved and performed in accordance with the guidelines of the Institutional Animal Care and Use Committee at the Biological Research Center (Szeged, Hungary) under the supervision of the Governmental Office for Csongrád County, Directorate of Food Chain Safety and Animal Health. The approval number is XVI./801/2018.

Southern blot

The probe for Southern blot was amplified from mouse genomic DNA using the primers given (supplemental material 1: Table S2). ^{32}P-Labeled probes were generated using the DecaLabel DNA Labeling kit (Thermo Fisher Scientific). Genomic DNA was extracted from tail biopsies and 20 μg was digested with SspI, electrophoresed through 0.8% agarose gel, and transferred to a HybondN membrane (GE Healthcare, Amersham, UK). After UV crosslinking, the membrane was hybridized overnight with ^{32}P-labeled probes at 65°C in 20× salt sodium citrate hybridization buffer. The membrane was exposed to a phosphorimager screen overnight.

Northern blot

Total RNA from livers, kidneys, and spleens of wild-type and *Fah*$^{-/-}$ mice was isolated with TRIzol Reagent (Invitrogen). RNA (20 μg) was separated on a 1% agarose-2.2 M formaldehyde gel, then blotted to a Hybond-N membrane (GE Healthcare). Filters were hybridized with ^{32}P-labeled cDNA probe specific for *Fah*.

Plasmid construction

The pT2HB-Fah and pXLBacII-Fah therapeutic transposon plasmids were constructed from the pT2HB (pT2/HB was a gift from Perry Hackett [Addgene plasmid no. 26557; http://n2t.net/addgene:26557; RRID:Addgene_26557]) and pXLBacII[59] plasmids, respectively, by inserting the PCR-amplified full-length mouse *Fah* CDS driven by the SV40 promoter between the transposon's ITRs. pcGlobin2-SB100 was constructed as described.[6] pcGlobin2-hyPBase was generated by inserting CDS of the hyPBase[7] into the pcGlobin2 plasmid. The hyPBase CDS source plasmid, pCMV-hyPBase, was provided by the Wellcome Trust Sanger Institute, Hinxton, Cambridge. N-Terminally HA-tagged SB100 and hyPBase were generated by standard PCR mutagenesis.

Nucleic acid isolation

For SBE-seq and transgene copy number assessment by qPCR, whole livers of gene therapeutically treated animals were lysed in 150 mL lysis buffer (100 mM Tris-HCl [pH 8], 5 mM EDTA [pH 8], 200 mM NaCl, 0.2% SDS) and incubated overnight at 50°C in the presence of 300 μg/mL Proteinase K (VWR Chemicals). One milliliter of the lysate was used for standard phenol/chloroform extraction. DNA precipitation was done with three volumes of absolute ethanol in the presence of 1/10 volume sodium acetate (3 M). Pelleted DNA was washed with 70% ethanol, air-dried, and dissolved in nuclease-free water (Sigma-Aldrich).

Genomic DNA from livers for the detection of transposase excision products by qPCR was extracted using the NucleoSpin Tissue Kit (Macherey-Nagel). The DNA concentration was determined by Qubit 3.0 Fluorometer (Thermo Fisher Scientific).

Total RNA for qRT-PCR was isolated from the livers of gene therapeutically treated mice using the NucleoSpin RNA Mini Kit (Macherey-Nagel).

Hydrodynamic tail vein injection

Plasmids for hydrodynamic tail vein injection were prepared using the NucleoBond Xtra Maxi Plus EF Kit (Macherey-Nagel). Before injection, we diluted plasmid DNA in Ringer's solution. For hydrodynamic injection, a volume equivalent to 10% of mouse body weight was administered via the lateral tail vein in 5–8 s into 6- to 8-week-old mice.[60,61] The amount of plasmid DNA was 50 μg for each of the therapeutic constructs mixed with 4 μg of the appropriate transposase helper plasmid. Gene therapeutically treated animals were sacrificed 2 and 7 months post injection. For delayed transposon introduction experiments only the plasmid encoding the transposase was injected on day 0 and then the plasmid containing the therapeutic construct was introduced on day 7.

ALT measurement from blood

Blood samples from wild-type, $Fah^{-/-}$, and gene therapeutically treated $Fah^{-/-}$ mice were collected by cardiac puncture at the time of sacrifice. After clot formation samples were centrifuged at 1,000 × *g* at 4°C for 10 min and 70 μL of the serum was immediately used for ALT level measurement on a Samsung PT10V instrument using the Liver Test11V cartridge (Samsung).

Transgene copy-number assessment, detection of transposase excision products, and determination of transposase mRNA level by qPCR

The assessment of *Fah* transgene copy number and the detection of transposase excision products were done by qPCR using the primers given (supplemental material 1: Table S2) with PerfeCTa SYBR Green SuperMix (Quantabio) on a Rotor-Gene Q instrument (QIAGEN). All reactions were carried out in triplicate using 30 ng gDNA for copy number determination and 500 ng for excision product detection. Cycling conditions were 95°C for 7 min followed by 35 cycles of 20 s at 95°C, 20 s at 64°C, and 20 s at 72°C. PCR efficiencies were analyzed with Rotor-Gene Q software (QIAGEN). Results were normalized to measurements of the *Olfr16* gene for copy-number determination and to *Omp* for excision product detection.

For the determination of transposase mRNA level, 2 μg of total RNA isolated from livers was treated with PerfeCTa DNase I (Quantabio) and reverse transcribed using a RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific). qRT-PCR was performed with the primers given (supplemental material 1: Table S2) under the conditions described above and the amount of mRNA for each transposase was normalized to that of *Rpl27*.

Immunoblotting

Age-matched C57BL/6Ntac mice were hydrodynamically injected with 4 μg of either pHA-SB100 or pHA-hyPBase, sacrificed 3, 7, or 14 days post injection, and livers were removed. The median lobe was dounce homogenized in 2 mL RIPA buffer supplemented with PMSF. Cleared samples were sonicated and supplemented with equal volumes of RIPA and stored at −80°C. Protein concentrations were calculated using the Pierce BCA Protein Assay kit (Thermo Fisher Scientific). A total of 80 μg of protein was separated on 8% (HA-SB100 samples) or 10% (HA-hyPBase samples) SDS-PAGE gels. The separated bands were transferred onto 0.2 μm nitrocellulose (Amersham). Membranes were incubated with peroxidase conjugated anti-HA (Roche, 12013819001, 1:1,000) and peroxidase conjugated anti-GAPDH (Invitrogen, MA5-15738-HRP, 1:10,000) antibodies. Filters were developed by using ECL Prime western blotting Detection Reagent (Amersham).

Immunohistochemistry

Mice were sacrificed 2 months after gene therapeutic treatment. Livers were removed and fixed in 4% formalin overnight and embedded in paraffin. From fixed tissues, 5 μm sections were cut and incubated at 56°C overnight. Immunohistochemistry was performed using the EnVision FLEX Mini Kit (DAKO). Antigen retrieval was done in a PT Link machine (DAKO). Primary antibody rabbit polyclonal anti-FAH (Thermo Fisher Scientific, PA5-42049, 1:100) was incubated for 120 min. Secondary antibody polyclonal goat anti-rabbit-HRP (DAKO, P0448) was incubated for 30 min. Visualization was done using an EnVision FLEX DAB+ Chromogen System

(DAKO, GV825). After hematoxylin counterstaining for 5 min, slides were mounted and scanned with a Pannoramic Digital Slide Scanner (3D Histech).

Sequencing of transposon insertion sites using SBE-seq

Genomic DNA (5 μg) from mouse liver was enzymatically fragmented, end-repaired, and dA-tailed using the QIAseq FX DNA Library Kit (QIAGEN). DNA was cleaned with AmpPure XP beads following each step. Next, a linker with a special structure, disfavoring the formation of linker-linker PCR products (supplemental material 1: Table S2), was ligated to the genomic DNA fragments. Then the 400–1,500 bp range of ligated fragments were selected using a Pippin Prep instrument (Sage Science). Next, on this selected fragment pool two nested PCRs were carried out, where one member of the PCR primer pairs was transposon-specific and the other was linker specific. The first PCR was performed for 16 cycles using a biotinylated transposon-specific primer (supplemental material 1: Table S2). Then the biotinylated PCR products were purified with streptavidin-coated paramagnetic beads. The second PCR was performed on the purified PCR products for another 16 cycles using transposon- and linker-specific primers carrying the Illumina indexing-specific paired-end adapters to create the final NGS library (supplemental material 1: Table S2). The libraries were cleaned with AmpPure XP beads. Each step of creating the NGS libraries was verified by analyzing sample aliquots on a TapeStation4200 instrument (Agilent) using D1000 screen tapes. The final concentration of each library was determined by a Qubit fluorometer, and libraries were spiked with 10% PhiX control library before proceeding with the Illumina run. On the Illumina instruments, custom-made Read2 and Index Read primers were used (supplemental material 1: Table S2) instead of the standard Illumina primers.

Identification and bioinformatics analysis of vector integration sites

The sequencing of SBE-seq libraries was performed using an Illumina MiSeq instrument. Paired-end reads from Illumina FASTQ files were assembled using PEAR[62] and filtered for the presence of PB and SB ITR sequences using the Smith-Watermann algorithm. Following the trimming of Illumina adapters and ITR sequences using Cutadapt,[63] reads extending at least 15 bp were considered for further analysis. We used the Burrows-Wheeler algorithm (BWA)[64] to map trimmed reads to the *Mus musculus* genome version GRCm39. To assure the quality of mapping we utilized a custom method to identify genomic regions where mismapped reads accumulate due to the imperfections of reference genome assemblies. We mapped reads from two different whole-genome sequencing experiments (SRR7278720 and SRR7278736) using BWA applying the mem -a parameter. Genomic regions with 100× the average read coverage were identified. Mapped reads from our experiments residing in these identified high read coverage regions were excluded from further analysis. The integration sites were determined as those genomic coordinates where at least three reads had their start point, and integration site sequences were identified as the first two and four nucleotides for SB and PB transposons, respectively.

Using a custom script, hot regions were collected from each liver as regions with an insertion density of greater than one per kilobase, with at least five insertions in total, and a less than 1 kb maximum distance between the adjacent insertions. To accept a hot region, these criteria had to be independently met in at least two animals in the given experimental panel over the same genomic region. Using the BEDTools[65] program we combined overlapping hot regions from different samples and annotated them. Adjacent hot regions were merged if a gap smaller than 1 kbp separated them. Regions overlapping with the Sfi1 locus were removed due to the inaccuracy of the reference genome here, as the actual copy number of Sfi1 is higher than the copy number 1 in the reference genome.[66] Genes located in the 5 kbp environment of the hot regions were also annotated according to their involvement in cancer-related processes. Genes implicated in cancer were identified by using the CGC (https://cancer.sanger.ac.uk/census) section of the COSMIC database (version 91), which contains an expert-curated collection of genes causally implicated in cancer. The transposon integration hotspot region list entries were supplemented by the tier 1 and tier 2 classification information of CGC.

To examine overlaps with genomic features we acquired gene annotation (version 104) and regulatory feature data from Ensembl (http://www.ensembl.org/Mus_musculus/Info/Index). TSSs were collected from gene annotation data as the first genomic positions of gene features. The ratio of transposon integrations overlapping genomic features as well as the density of transposon integrations in the 5 kb vicinity of TSSs and histone modifications were calculated using BEDTools intersect. Density plots were drawn with R (version 4.1.1) using the ggplot2 package (with geom_smooth function, version 3.3.5). The control dataset was generated by collecting all TA and TTAA (canonical integration sites, for SB and PB, respectively) genomic positions using a custom script and then randomly selecting a control dataset with the same number of elements as the real vector integration datasets under investigation using GNU Coreutils shuf command.

For investigation of potential local hopping we used a custom script to assess vector insertion density in 0.5 Mb windows starting from each integration site. We identified those regions where the density of integrations was significantly greater than predicted by the binomial model, taking into account the number of integrations on the given chromosome and the length of the given region relative to the total length of the chromosome. With a significance setting of 5%, only regions with significantly more integration than expected were accepted as local hopping-positive regions. The windows were placed in both directions starting from all integrations. Integrations located in a local hopping-positive region had to fulfill the test requirement continuously. The first and last integrations from which overlapping windows yet fulfilled the test criteria were defined as the 5′ and 3′ endpoints of the local hopping-positive region. We only accepted regions where at least 5 consecutive integrations have met the criteria without interruption. Each control sample consisted of 12 integration site sets, similar to the real samples, corresponding to the 12-12

transposon-treated animals. Each of the 12 sets contained a number of randomly selected PB (TTAA) or SB (TA) integration sites exactly equal to the number of integrations identified in the corresponding PB- or SB-treated animals.

To get rid of the distortion effect of hot regions, the number of integrations in the areas of the previously defined hot regions was artificially reduced to one in each animal; therefore, hot regions were virtually treated as a single extended integration site. During PB local hopping analysis we used histone modification-biased control data, meaning that we collected the same quantity (29%) of TTAA genomic sequences as observed for true PB integrations randomly from genomic regions that exhibited chromatin modifications showing a strong correlation with the PB-generated integration events (merged peaks of H3K27Ac and H3K4Me1-3 ChIP-seq experiments). All applied custom scripts were written in Java (version 17) or Python (version 3.7).

Analysis of ChIP-seq data

The National Center for Biotechnology Information (NCBI) Sequence Read Archive served as the source of raw histone modification ChIP-seq experiment data (Table S4). The primary analysis of the downloaded raw data was carried out using an analysis pipeline developed by the University of Debrecen.[67] Read mapping was performed using BWA to the reference genome version GRCm39. The model-based analysis of ChIP-seq (MACS2)[68] was used for peak calling (applying the –broad parameter). Peaks overlapping ENCODE blacklisted regions[69] were excluded with the BEDTools program.

Data visualization and statistics

Statistical calculations were carried out in the Python and R (http://www.R-project.org) environments. Statistics were calculated using Fisher's exact, Pearson's chi-squared, and one-way ANOVA tests. The threshold for significance was $p < 0.05$. Genome browser-compatible files were made using BEDTools and makeUCSCfile. We used the GraphPad Prism version 8.4.3 for Windows, GraphPad Software (San Diego, CA; www.graphpad.com) in the data-visualization phases. For the creation of sequence logo representations, nucleotide sequences of all uniquely identified transposon integration sites were collected with a custom script using the genome version GRCm39. The probability matrices of non-canonical integration sites (not TA and TTAA, for SB and PB, respectively) were calculated using Biopython (version 1.76). Sequence logos were drawn with Logomaker (version 0.8).

DATA AVAILABILITY

The datasets generated and analyzed during the current study are available in the NCBI Sequence Read Archive (SRA) repository. PB and SB transposon integration site data generated by SBE-seq were deposited in the SRA database, under bioproject ID PRJNA605684. The processed bed files containing the identified unique and filtered genomic vector integration sites per animal and in merged format from 1 to 12 animals treated with either PB or SB vectors were deposited in GEO database and are accessible through GEO Series accession number GSE212895.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.omtm.2023.03.003.

ACKNOWLEDGMENTS

This work was supported by the Momentum Program of the Hungarian Academy of Sciences (LP2015-5/2015), the National Research, Development and Innovation Office (Hungary) grant GINOP-2.3.2-15-2016-00024, European Union's Horizon 2020 research and innovation program under grant agreement no. 739593 and the National Research, Development and Innovation Office (PharmaLab, RRF-2.3.1-21-2022-00015). A.G.K. was supported by the Szeged Scientists Academy under the sponsorship of the Hungarian Ministry of Innovation and Technology (FEIF/646-4/2021-ITM_SZERZ). R.K. was supported by the ÚNKP-22-3-SZTE-261 New National Excellence Program of the Ministry for Culture and Innovation from the source of the National Research, Development and Innovation Fund. E.C. was supported by the Higher Education Institutional Excellence Programme (NKFIH-1150-6/2019) of the Ministry for Innovation and Technology in Hungary: FIKP_20428-3_2018_FELITSTRAT. The authors thank Tibor Rauch for helpful discussions, Balázs Papp for support in data analysis, and Erzsébet Fejes for linguistic support.

AUTHOR CONTRIBUTIONS

L.M. contributed to the conception and design. L.M. and A.N. contributed to the writing, review, and revision of the manuscript with input from all authors. L.M. contributed to the analysis and interpretation of data with help from F.S., L.G.N., I.N., and L.H. R.K., L.H., and A.N. performed the animal experiments. T.R. contributed to the creation of the KO mouse strain used. A.N. and A.G.K. participated in the construction and purification of plasmids. G.I. and D.L. carried out the NGS experiments. G.I., B.T., E.C., G.J., B.M.V., Z.H., B.B., A.K., L.P., and E.B. performed bioinformatics analysis. A.B.D., A.G.K., and P.G. carried out qPCR and qRT-PCR experiments. G.P.-B. and K.S.A.A. performed the immunohistochemistry. A.B. and R.K. performed the western blot experiment. L.M. contributed to the study supervision. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Hacein-Bey-Abina, S., Garrigue, A., Wang, G.P., Soulier, J., Lim, A., Morillon, E., Clappier, E., Caccavelli, L., Delabesse, E., Beldjord, K., et al. (2008). Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. J. Clin. Invest. *118*, 3132–3142. https://doi.org/10.1172/JCI35700.

2. Montini, E., Cesana, D., Schmidt, M., Sanvito, F., Bartholomae, C.C., Ranzani, M., Benedicenti, F., Sergi, L.S., Ambrosi, A., Ponzoni, M., et al. (2009). The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site

selection in a mouse model of HSC gene therapy. J. Clin. Invest. *119*, 964–975. https://doi.org/10.1172/JCI37630.

3. Deyle, D.R., and Russell, D.W. (2009). Adeno-associated virus vector integration. Curr. Opin. Mol. Ther. *11*, 442–447.

4. Donsante, A., Vogler, C., Muzyczka, N., Crawford, J.M., Barker, J., Flotte, T., Campbell-Thompson, M., Daly, T., and Sands, M.S. (2001). Observed incidence of tumorigenesis in long-term rodent studies of rAAV vectors. Gene Ther. *8*, 1343–1346. https://doi.org/10.1038/sj.gt.3301541.

5. 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. https://doi.org/10.1126/science.1142658.

6. Mátés, L., Chuah, M.K.L., Belay, E., Jerchow, B., Manoj, N., Acosta-Sanchez, A., Grzela, D.P., Schmitt, A., Becker, K., Matrai, J., et al. (2009). Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. Nat. Genet. *41*, 753–761. https://doi.org/10.1038/ng.343.

7. Yusa, K., Zhou, L., Li, M.A., Bradley, A., and Craig, N.L. (2011). A hyperactive piggyBac transposase for mammalian applications. Proc. Natl. Acad. Sci. USA *108*, 1531–1536. https://doi.org/10.1073/pnas.1008322108.

8. Tipanee, J., Chai, Y.C., VandenDriessche, T., and Chuah, M.K. (2017). Preclinical and clinical advances in transposon-based gene therapy. Biosci. Rep. *37*. https://doi.org/10.1042/BSR20160614.

9. Moldt, B., Yant, S.R., Andersen, P.R., Kay, M.A., and Mikkelsen, J.G. (2007). Cis-acting gene regulatory activities in the terminal regions of sleeping beauty DNA transposon-based vectors. Hum. Gene Ther. *18*, 1193–1204. https://doi.org/10.1089/hum.2007.099.

10. Cadiñanos, J., and Bradley, A. (2007). Generation of an inducible and optimized piggyBac transposon system. Nucleic Acids Res. *35*, e87.

11. Shi, X., Harrison, R.L., Hollister, J.R., Mohammed, A., Fraser, M.J., Jr., and Jarvis, D.L. (2007). Construction and characterization of new piggyBac vectors for constitutive or inducible expression of heterologous gene pairs and the identification of a previously unrecognized activator sequence in piggyBac. BMC Biotechnol. *7*, 5.

12. de Jong, J., Wessels, L.F.A., van Lohuizen, M., de Ridder, J., and Akhtar, W. (2014). Applications of DNA integrating elements: facing the bias bully. Mob. Genet. Elements *4*, 1–6. https://doi.org/10.4161/2159256X.2014.992694.

13. 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. https://doi.org/10.1101/gad.7.12a.2298.

14. Ivics, Z., Li, M.A., Mátés, L., Boeke, J.D., Nagy, A., Bradley, A., and Izsvák, Z. (2009). Transposon-mediated genome manipulation in vertebrates. Nat. Methods *6*, 415–422. https://doi.org/10.1038/nmeth.1332.

15. Bishop, D.C., Clancy, L.E., Simms, R., Burgess, J., Mathew, G., Moezzi, L., Street, J.A., Sutrave, G., Atkins, E., McGuire, H.M., et al. (2021). Development of CAR T-cell lymphoma in 2 of 10 patients effectively treated with piggyBac-modified CD19 CAR T cells. Blood *138*, 1504–1509. https://doi.org/10.1182/blood.2021010813.

16. Micklethwaite, K.P., Gowrishankar, K., Gloss, B.S., Li, Z., Street, J.A., Moezzi, L., Mach, M.A., Sutrave, G., Clancy, L.E., Bishop, D.C., et al. (2021). Investigation of product-derived lymphoma following infusion of piggyBac-modified CD19 chimeric antigen receptor T cells. Blood *138*, 1391–1405. https://doi.org/10.1182/blood.2021010858.

17. Scott, C.R. (2006). The genetic tyrosinemias. Am. J. Med. Genet. C Semin. Med. Genet. *142C*, 121–126. https://doi.org/10.1002/ajmg.c.30092.

18. 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. https://doi.org/10.1038/ng0895-453.

19. 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. https://doi.org/10.1016/0140-6736(92)92685-9.

20. 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. https://doi.org/10.1038/ng0396-266.

21. de Jong, J., Akhtar, W., Badhai, J., Rust, A.G., Rad, R., Hilkens, J., Berns, A., van Lohuizen, M., Wessels, L.F.A., and de Ridder, J. (2014). Chromatin landscapes of retroviral and transposon integration profiles. PLoS Genet. *10*, e1004250. https://doi.org/10.1371/journal.pgen.1004250.

22. Sondka, Z., Bamford, S., Cole, C.G., Ward, S.A., Dunham, I., and Forbes, S.A. (2018). The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. Nat. Rev. Cancer *18*, 696–705. https://doi.org/10.1038/s41568-018-0060-1.

23. Yoshida, J., Akagi, K., Misawa, R., Kokubu, C., Takeda, J., and Horie, K. (2017). Chromatin states shape insertion profiles of the piggyBac, Tol2 and Sleeping Beauty transposons and murine leukemia virus. Sci. Rep. *7*, 43613. https://doi.org/10.1038/srep43613.

24. Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q., and Zhao, K. (2008). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat. Genet. *40*, 897–903. https://doi.org/10.1038/ng.154.

25. Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823–837. https://doi.org/10.1016/j.cell.2007.05.009.

26. Lehnertz, B., Ueda, Y., Derijck, A.A., Braunschweig, U., Perez-Burgos, L., Kubicek, S., Chen, T., Li, E., Jenuwein, T., and Peters, A.H.F.M. (2003). Suv39h-mediated histone H3 lysine 9 methylation directs DNA methylation to major satellite repeats at pericentric heterochromatin. Curr. Biol. *13*, 1192–1200. https://doi.org/10.1016/s0960-9822(03)00432-9.

27. Wang, W., Lin, C., Lu, D., Ning, Z., Cox, T., Melvin, D., Wang, X., Bradley, A., and Liu, P. (2008). Chromosomal transposition of PiggyBac in mouse embryonic stem cells. Proc. Natl. Acad. Sci. USA *105*, 9290–9295. https://doi.org/10.1073/pnas.0801017105.

28. Carlson, C.M., Dupuy, A.J., Fritz, S., Roberg-Perez, K.J., Fletcher, C.F., and Largaespada, D.A. (2003). Transposon mutagenesis of the mouse germline. Genetics *165*, 243–256.

29. Izsvák, Z., Stüwe, E.E., Fiedler, D., Katzer, A., Jeggo, P.A., and Ivics, Z. (2004). Healing the wounds inflicted by sleeping beauty transposition by double-strand break repair in mammalian somatic cells. Mol. Cell *13*, 279–290. https://doi.org/10.1016/s1097-2765(03)00524-0.

30. 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. https://doi.org/10.1007/s10545-008-0793-1.

31. Zhang, G., Gao, X., Song, Y.K., Vollmer, R., Stolz, D.B., Gasiorowski, J.Z., Dean, D.A., and Liu, D. (2004). Hydroporation as the mechanism of hydrodynamic delivery. Gene Ther. *11*, 675–682. https://doi.org/10.1038/sj.gt.3302210.

32. Yin, H., Xue, W., Chen, S., Bogorad, R.L., Benedetti, E., Grompe, M., Koteliansky, V., Sharp, P.A., Jacks, T., and Anderson, D.G. (2014). Genome editing with Cas9 in adult mice corrects a disease mutation and phenotype. Nat. Biotechnol. *32*, 551–553. https://doi.org/10.1038/nbt.2884.

33. 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. https://doi.org/10.1006/mthe.2002.0812.

34. Baratta, J.L., Ngo, A., Lopez, B., Kasabwalla, N., Longmuir, K.J., and Robertson, R.T. (2009). Cellular organization of normal mouse liver: a histological, quantitative immunocytochemical, and fine structural analysis. Histochem. Cell Biol. *131*, 713–726. https://doi.org/10.1007/s00418-009-0577-1.

35. Sohlenius-Sternbeck, A.K. (2006). Determination of the hepatocellularity number for human, dog, rabbit, rat and mouse livers from protein concentration measurements. Toxicol. Vitro *20*, 1582–1586. https://doi.org/10.1016/j.tiv.2006.06.003.

36. Friedrich, M.J., Rad, L., Bronner, I.F., Strong, A., Wang, W., Weber, J., Mayho, M., Ponstingl, H., Engleitner, T., Grove, C., et al. (2017). Genome-wide transposon screening and quantitative insertion site sequencing for cancer gene discovery in mice. Nat. Protoc. *12*, 289–309. https://doi.org/10.1038/nprot.2016.164.

37. Li, M.A., Pettitt, S.J., Eckert, S., Ning, Z., Rice, S., Cadiñanos, J., Yusa, K., Conte, N., and Bradley, A. (2013). The piggyBac transposon displays local and distant reintegration preferences and can cause mutations at noncanonical integration sites. Mol. Cell Biol. *33*, 1317–1330. https://doi.org/10.1128/MCB.00670-12.

38. Gogol-Döring, A., Ammar, I., Gupta, S., Bunse, M., Miskey, C., Chen, W., Uckert, W., Schulz, T.F., Izsvák, Z., and Ivics, Z. (2016). Genome-wide profiling reveals remarkable parallels between insertion site selection properties of the MLV retrovirus and the piggyBac transposon in primary human CD4(+) T cells. Mol. Ther. *24*, 592–606. https://doi.org/10.1038/mt.2016.11.

39. Kopasz, A.G., Pusztai, D.Z., Karkas, R., Hudoba, L., Abdullah, K.S.A., Imre, G., Pankotai-Bodó, G., Migh, E., Nagy, A., Kriston, A., et al. (2022). A versatile transposon-based technology to generate loss- and gain-of-function phenotypes in the mouse liver. BMC Biol. *20*, 74. https://doi.org/10.1186/s12915-022-01262-x.

40. Ju, H.L., Ahn, S.H., Kim, D.Y., Baek, S., Chung, S.I., Seong, J., Han, K.H., and Ro, S.W. (2013). Investigation of oncogenic cooperation in simple liver-specific transgenic mouse models using noninvasive in vivo imaging. PLoS One *8*, e59869. https://doi.org/10.1371/journal.pone.0059869.

41. Orejuela, D., Jorquera, R., Bergeron, A., Finegold, M.J., and Tanguay, R.M. (2008). Hepatic stress in hereditary tyrosinemia type 1 (HT1) activates the AKT survival pathway in the fah-/- knockout mice model. J. Hepatol. *48*, 308–317. https://doi.org/10.1016/j.jhep.2007.09.014.

42. Grompe, M., Overturf, K., al-Dhalimy, M., and Finegold, M. (1998). Therapeutic trials in the murine model of hereditary tyrosinaemia type I: a progress report. J. Inherit. Metab. Dis. *21*, 518–531. https://doi.org/10.1023/a:1005462804271.

43. Kokubu, C., Horie, K., Abe, K., Ikeda, R., Mizuno, S., Uno, Y., Ogiwara, S., Ohtsuka, M., Isotani, A., Okabe, M., et al. (2009). A transposon-based chromosomal engineering method to survey a large cis-regulatory landscape in mice. Nat. Genet. *41*, 946–952. https://doi.org/10.1038/ng.397.

44. Skipper, K.A., Andersen, P.R., Sharma, N., and Mikkelsen, J.G. (2013). DNA transposon-based gene vehicles - scenes from an evolutionary drive. J. Biomed. Sci. *20*, 92. https://doi.org/10.1186/1423-0127-20-92.

45. Bhargava, R., Onyango, D.O., and Stark, J.M. (2016). Regulation of single-strand annealing and its role in genome maintenance. Trends Genet. *32*, 566–575. https://doi.org/10.1016/j.tig.2016.06.007.

46. Adams, M.D., and Sekelsky, J.J. (2002). From sequence to phenotype: reverse genetics in Drosophila melanogaster. Nat. Rev. Genet. *3*, 189–198. https://doi.org/10.1038/nrg752.

47. Kim, H., Kim, K., Kim, J., Kim, S.H., and Yim, J. (2012). Mutagenesis by imprecise excision of the piggyBac transposon in Drosophila melanogaster. Biochem. Biophys. Res. Commun. *417*, 335–339. https://doi.org/10.1016/j.bbrc.2011.11.110.

48. Schambach, A., Morgan, M., and Fehse, B. (2021). Two cases of T cell lymphoma following Piggybac-mediated CAR T cell therapy. Mol. Ther. *29*, 2631–2633. https://doi.org/10.1016/j.ymthe.2021.08.013.

49. Kebriaei, P., Singh, H., Huls, M.H., Figliola, M.J., Bassett, R., Olivares, S., Jena, B., Dawson, M.J., Kumaresan, P.R., Su, S., et al. (2016). Phase I trials using Sleeping Beauty to generate CD19-specific CAR T cells. J. Clin. Invest. *126*, 3363–3376. https://doi.org/10.1172/JCI86721.

50. Srour, S.A., Singh, H., McCarty, J., de Groot, E., Huls, H., Rondon, G., Qazilbash, M., Ciurea, S., Bardelli, G., Buck, J., et al. (2020). Long-term outcomes of Sleeping Beauty-generated CD19-specific CAR T-cell therapy for relapsed-refractory B-cell lymphomas. Blood *135*, 862–865. https://doi.org/10.1182/blood.2019002920.

51. Gray, Y.H. (2000). It takes two transposons to tango: transposable-element-mediated chromosomal rearrangements. Trends Genet. *16*, 461–468. https://doi.org/10.1016/s0168-9525(00)02104-1.

52. Querques, I., Mades, A., Zuliani, C., Miskey, C., Alb, M., Grueso, E., Machwirth, M., Rausch, T., Einsele, H., Ivics, Z., et al. (2019). A highly soluble Sleeping Beauty transposase improves control of gene insertion. Nat. Biotechnol. *37*, 1502–1512. https://doi.org/10.1038/s41587-019-0291-z.

53. Cai, Y., Bak, R.O., Krogh, L.B., Staunstrup, N.H., Moldt, B., Corydon, T.J., Schrøder, L.D., and Mikkelsen, J.G. (2014). DNA transposition by protein transduction of the piggyBac transposase from lentiviral Gag precursors. Nucleic Acids Res. *42*, e28. https://doi.org/10.1093/nar/gkt1163.

54. Bell, J.B., Aronovich, E.L., Schreifels, J.M., Beadnell, T.C., and Hackett, P.B. (2010). Duration of expression and activity of Sleeping Beauty transposase in mouse liver following hydrodynamic DNA delivery. Mol. Ther. *18*, 1796–1802. https://doi.org/10.1038/mt.2010.152.

55. Bire, S., Casteret, S., Arnaoty, A., Piégu, B., Lecomte, T., and Bigot, Y. (2013). Transposase concentration controls transposition activity: myth or reality? Gene *530*, 165–171. https://doi.org/10.1016/j.gene.2013.08.039.

56. Rülicke, T. (2004). Pronuclear microinjection of mouse zygotes. Methods Mol. Biol. *254*, 165–194.

57. Rülicke, T., Haenggli, A., Rappold, K., Moehrlen, U., and Stallmach, T. (2006). No transuterine migration of fertilised ova after unilateral embryo transfer in mice. Reprod. Fertil. Dev. *18*, 885–891.

58. FELASA working group on revision of guidelines for health monitoring of rodents and rabbits, Mähler Convenor, M., Berard, M., Feinstein, R., Gallagher, A., Illgen-Wilcke, B., Pritchett-Corning, K., and Raspa, M. (2014). FELASA recommendations for the health monitoring of mouse, rat, hamster, Guinea pig and rabbit colonies in breeding and experimental units. Lab. Anim. *48*, 178–192.

59. Li, X., Harrell, R.A., Handler, A.M., Beam, T., Hennessy, K., and Fraser, M.J., Jr. (2005). piggyBac internal sequences are necessary for efficient transformation of target genomes. Insect Mol. Biol. *14*, 17–30. https://doi.org/10.1111/j.1365-2583.2004.00525.x.

60. Bell, J.B., Podetz-Pedersen, K.M., Aronovich, E.L., Belur, L.R., McIvor, R.S., and Hackett, P.B. (2007). Preferential delivery of the Sleeping Beauty transposon system to livers of mice by hydrodynamic injection. Nat. Protoc. *2*, 3153–3165. https://doi.org/10.1038/nprot.2007.471.

61. Portier, I., Vanhoorelbeke, K., Verhenne, S., Pareyn, I., Vandeputte, N., Deckmyn, H., Goldenberg, D.S., Samal, H.B., Singh, M., Ivics, Z., et al. (2018). High and long-term von Willebrand factor expression after Sleeping Beauty transposon-mediated gene therapy in a mouse model of severe von Willebrand disease. J. Thromb. Haemost. *16*, 592–604. https://doi.org/10.1111/jth.13938.

62. Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics *30*, 614–620. https://doi.org/10.1093/bioinformatics/btt593.

63. Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. *17*, 3. https://doi.org/10.14806/ej.17.1.200.

64. Li, H., and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics *25*, 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

65. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.

66. Quinlan, A.R., Clark, R.A., Sokolova, S., Leibowitz, M.L., Zhang, Y., Hurles, M.E., Mell, J.C., and Hall, I.M. (2010). Genome-wide mapping and assembly of structural variant breakpoints in the mouse genome. Genome Res. *20*, 623–635. https://doi.org/10.1101/gr.102970.109.

67. Barta, E. (2011). Command line analysis of ChIP-seq results. EMBnet J. *17*, 5. https://doi.org/10.14806/ej.17.1.209.

68. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-seq (MACS). Genome Biol. *9*, R137. https://doi.org/10.1186/gb-2008-9-9-r137.

69. Amemiya, H.M., Kundaje, A., and Boyle, A.P. (2019). The ENCODE blacklist: identification of problematic regions of the genome. Sci. Rep. *9*, 9354. https://doi.org/10.1038/s41598-019-45839-z.