piggyBac transposase tools for genome engineering

Xianghong Li,a,b, Erin R. Burnightc, Ashley L. Cooneyd, Nirav Malani,e, Troy Brady,e, Jeffry D. Sanderf,g, Janice Staberf, Sarah J. Wheelanh, J. Keith Joungf,g, Paul B. McCray, Jr.c,d, Frederic D. Bushmane, Patrick L. Sinnid,1, and Nancy L. Craiga,b,1

aHoward Hughes Medical Institute and bDepartment of Molecular Biology and Genetics, The Johns Hopkins University School of Medicine, Baltimore, MD 21205; +Genetics Program, Carver College of Medicine, The University of Iowa, Iowa City, IA 52242; 1Department of Pediatrics, Carver College of Medicine, The University of Iowa, Iowa City, IA 52242; 2Department of Microbiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104; 3Molecular Pathology Unit, Center for Computational and Integrative Biology, and Center for Cancer Research, Massachusetts General Hospital, Boston, MA 02129; 4Department of Pathology, Harvard Medical School, Boston, MA 02115; and 5Division of Biostatistics and Bioinformatics, Department of Oncology, The Johns Hopkins University School of Medicine, Baltimore, MD 21205

Edited by Richard A. Young, Massachusetts Institute of Technology, Cambridge, MA, and approved May 1, 2013 (received for review December 30, 2012)

The transposon piggyBac is being used increasingly for genetic studies. Here, we describe modified versions of piggyBac transposase that have potentially wide-ranging applications, such as reversible transgenesis and modified targeting of insertions. piggyBac is distinguished by its ability to excise precisely, restoring the donor site to its pretransposon state. This characteristic makes piggyBac useful for reversible transgenesis, a potentially valuable feature when generating induced pluripotent stem cells without permanent alterations to genomic sequence. To avoid further genome modification following piggyBac excision by reintegration, we generated an excision competent/integration defective (Exc+Int−) transposase. Our findings also suggest the position of a target DNA–transposase interaction. Another goal of genome engineering is to develop reagents that can guide transgenes to preferred genomic regions. Others have shown that piggyBac transposase can be active when fused to a heterologous DNA-binding domain. An Exc−Int transposase, the intrinsic targeting of which is defective, might also be a useful intermediate in generating a transposase whose integration activity could be rescued and redirected by fusion to a site-specific DNA-binding domain. We show that fusion to two designed zinc finger proteins rescued the Int− phenotype. Successful guided transgene integration into genomic DNA would have broad applications to gene therapy and molecular genetics. Thus, an Exc−Int transposase is a potentially useful reagent for genome engineering and provides insight into the mechanism of transposase–target DNA interaction.

GULOP | ROSA26 | protein-DNA interaction | induced pluripotent stem cell production

DNA “cut-and-paste” transposable elements are important tools for genome engineering, such as insertional mutagenesis and transgenesis. Research with the DNA transposon Sleeping Beauty, a “resurrected” transposon, has pioneered the use of DNA transposons in mammalian cells (1, 2). piggyBac is also a DNA transposon and a promising alternative to Sleeping Beauty. piggyBac, originally isolated from the cabbage looper moth Trichoplusia ni genome (3), has a large cargo size (4), is highly active in many cell types, and mediates long-term expression in mammalian cells in vivo (5–10). piggyBac is also distinguished by its ability to excise precisely (11), thus restoring the donor site to its pretransposon insertion sequence.

Because it can excise precisely, piggyBac is especially useful if a transgene is only transiently required. Transient integration and expression of transcription factors are important approaches to generate transgene-free induced pluripotent stem cells (iPSCs) (12, 13) as well as directed differentiation of specific cell types for both research and clinical use. Removal of the transgene is key for potential therapeutic applications of iPSCs. piggyBac has been used as a vector for reversible integration; however, reintegration of the transposon catalyzed by piggyBac (PB) transposase occurs in 40–50% of cells (14). To generate iPSCs without any genetic change, a PB mutant, which can promote only excision (Exc+) and not integration (Int−), would be a useful tool. Here we performed site-directed mutagenesis of the PB catalytic domain and isolated an excision competent/integration defective (Exc+Int−) PB. We also find that introduction of the Exc−Int− mutations into a hyperactive version of PB (5, 8, 15) yields an Exc−Int− transposase whose excision frequency is five–to six-fold higher than that of wild-type PB. We speculate that the changed amino acids in the Exc−Int− mutant are likely positions of interaction between the transposase and the target DNA.

Modification of nonviral vector systems to alter native integration patterns in a predictable manner would provide molecular biological tools with relevance to diverse applications, including therapeutics for genetic disease and ex vivo stem cell manipulation. piggyBac inserts at target sites with the sequence TTAA but displays little selectivity for particular regions of the genome other than a modest preference for regions of DNase I sensitivity (16, 17). A useful modification of PB would be to be able to guide integrations to safe harbor sites. Others have shown that PB is distinguished by its ability to remain active when fused to a DNA binding domain and that such fusions can bias insertion toward cognate sites (10, 16, 18). Notably, Wilson and colleagues (19) fused a zinc finger protein (ZFP) targeted to the upstream promoter region of the cell-cycle checkpoint kinase 2 protein-coding gene CHEK2 to the native insect-derived piggyBac (iPB) transposase. The ZFP–iPB fusions exhibited comparable transposition activity to unmodified transposase and bound the target site in their study. However, no

Significance

DNA transposons that translocate by excision from a donor site and insertion into a target site are often used for genome engineering by insertional mutagenesis and transgenesis. The piggyBac element is especially useful because it can excise precisely from an insertion site, restoring the site to its pretransposon state. Precise excision is particularly useful when transient transposon is needed, for example, in the transient introduction of transcription factors for induced pluripotent stem cell production. We have used mutagenesis to generate an Excision+ Integration− transposase that allows piggyBac excision without potentially harmful reintegration. These mutations likely lie in a target DNA-binding domain.


Conflict of interest statement: J.K.I. has a financial interest in Transposagen Biopharmaceuticals. J.K.I.’s interests were reviewed and are managed by Massachusetts General Hospital and Partners HealthCare in accordance with their conflict of interest policies. This article is a PNAS Direct Submission.

Fees freely available online through the PNAS open access option.

1To whom correspondence may be addressed. E-mail: ncraig@jhmi.edu or patrick-sinn@uiowa.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1305987110/-/DCSupplemental.

www.pnas.org/cgi/doi/10.1073/pnas.1305987110
integration events were recovered at the native site in HEK293 cells, presumably due to the paucity of TTAA sites within the target region. Indeed, this hypothesis is supported with subsequent experiments demonstrating a significant increase in integration recovery near a CHEK2 target site engineered with 12 TTAA sites. Stringent target-site selection in the genome has not been demonstrated with any transposon system, including piggyBac or Sleeping Beauty (20).

In this study we analyzed the effects of targeting by the Exc+Int+ PB upon fusion to several ZFPs designed to recognize putative safe harbor regions. Importantly, we observed that N-terminal fusion of heterologous ZFPs rescued the integration defect. No significant targeting to the ZFP-binding sites, however, was observed. We conclude that an Exc+Int+ piggyBac transposasen provides a tool for use in genome engineering.

**Results**

**Design and In Vivo Analysis of Exc+Int− Mutants.** Like many transposase and integrase enzymes (21), the catalytic core of PB contains three aspartic acids that are necessary for catalytic activity: D268, D346, and D447 (22, 23) (Fig. 1). Interspersed near these required residues are 15 arginines and lysines that are highly conserved among many piggyBac proteins (22, 23) (Fig. S1). We hypothesized that a subset of these positively charged amino acids are involved in the transposase–target DNA interaction. To identify the contributions of these amino acids for transposition, we performed scanning alanine mutagenesis of the basic residues in the core domain, as well as of two additional residues (S351 and K356) (Fig. 1), which are comparable to positions in HIV-1 integrase where mutations alter HIV target joining (24).

The catalytic domain of a transposase has two functions: excision from the donor DNA and integration into the target DNA. As such, the alanine substitutions introduced into the catalytic domain may alter excision, integration, both, or neither. To discern the phenotype of the mutants, we assayed their excision activity by PCR amplification of repaired donor backbone in HEK293 cells by following cotransfection of a donor plasmid, which carries a piggyBac transposon and an excision resistant transposase. Subsequent blasticidin resistance of the HeLa cell line (8). Briefly, Tol2 transposition was used to stably introduce a gene cassette carrying a cycle 3 GFP gene (Invitrogen) into the genome. Following transient transfection of a plasmid expressing PB, the piggyBac transposon is precisely excised, the resulting GFP signal can be measured by FACS analysis.

### Analysis of PB Mutation Candidates

To further analyze the Exc+Int− mutants, we assayed their transposition activities in a cell-free assay using transposase purified from Escherichia coli (22). piggyBac transposon, starting with excision, initiates with a nick at the 3′ end of the transposon, followed by the attack of the newly exposed end on the complementary strand, four nucleotides within the donor DNA that releases the flanking donor DNA and a TTAA hairpin-containing transposon end. This hairpin is then opened by the transposase, re-exposing the 3′ OH transposon end that attacks the target DNA and leaving TTAA attached to the 5′ transposon end.

To analyze double-strand breaks at the ends of the transposon, we end-labeled a linear substrate DNA containing a piggyBac element. Cleavage at the transposon ends releases the labeled DNA fragments that flank the transposon in the donor site (Fig. 3A). This assay reveals that mutant PBs R245A, R275A/R277A, R341A, and R372A/K375A promote double-strand breaks at the transposon ends at levels similar to wild-type PB; i.e., they are Exc+ (Fig. 3B).

In contrast, the excision activity of the transposase mutants R388A and S351E was markedly reduced.

The hairpin opening and target-joining activity of the purified Exc+Int− candidates was determined by analysis of an end-labeled piggyBac end hairpin (Fig. 3C). Upon incubation with transposase, the hairpin is opened and the transposon ends can join to a target plasmid DNA. We find that mutants R245A, R275A/R277A, R341A, and R372A/K375A open terminal hairpins with activity comparable to wild-type PB (Fig. 3D). Mutant transposases S351E and R388A, however, did not exhibit hairpin opening and target-joining activity. Only one mutant among the Exc− and hairpin opening-defective transposases, R372A/K375A, was defective in target joining. These results suggest that the PB mutant R372A/K375A (PB−E2280) was the best candidate for focused studies that separate the excision and integration transposase functions.

**Analysis of PB**

To quantitatively characterize PB−E2280 transposition in cells, we assayed the excision activity in a genetically manipulated HEK293 cell line (8). Briefly, Tol2 transposition was used to stably introduce a gene cassette carrying a cycle 3 GFP gene (Invitrogen) interrupted by a piggyBac transposon, termed GFP::PB, into the genome. Following transient transfection of a plasmid expressing PB, the piggyBac transposon is precisely excised, the resulting GFP gene is restored, and the resulting GFP signal can be measured by FACs analysis. We found that the excision activity of PB−E2280 is about 30–50% of that of wild-type PB (Fig. 4). This observation prompted us to search for “suppressor” (i.e., excision-hyperactive) mutations that would increase the excision activity of PB−E2280 but maintain its Int− phenotype.

### Analysis of PB−E2280

Catalytic core amino acids selected for mutagenesis. The catalytic core of PB contains 3 requisite aspartic acids (blue). Fifteen conserved arginines and lysines (red) and two conserved amino acids important for HIV integration (green), which are targets of mutagenesis, are indicated.

**Fig. 1.** Catalytic core amino acids selected for mutagenesis. The catalytic core of PB contains 3 requisite aspartic acids (blue). Fifteen conserved arginines and lysines (red) and two conserved amino acids important for HIV integration (green), which are targets of mutagenesis, are indicated.
We screened ~6,000 colonies from three separate mutagenesis libraries and isolated about 30 hyperactive candidates that showed increased fluorescence relative to PB\(^{R372A/K375A}\), some of which were indeed hyperactive upon retesting (L15P, D19N, E45G, S103P\(^*\), R189K, M194V\(^*\), N384T, M413V, D560A, M589V, S17G/K102E, S31P/T164A, H33Y/N571S\(^*\), E44K/K334R, C97R/T242I; Fig. 4A). A subset of these mutations (indicated by *) was also isolated in our previously reported screen for increased excision hyperactivities of PB in yeast (15). We then measured excision activity of the PB\(^{R372A/K375A}\) mutants in mammalian HEK293 GFP::PB cells using the GFP excision assay (Fig. 4B). We found the highest excision activity with the single amino acid changes M194V or D450N. Notably, M194V was previously identified in our yeast screen for hyperactive mutants of wild-type PB although it was not hyperactive in mammalian cells (15). D450 is near D447, one of the requisite catalytic core amino acids (22, 23), and is highly conserved among piggyBac proteins (Fig. S1).

**Characterization of PB\(^{M194V/R372A/K375A}\) and PB\(^{R372A/K375A/D450N}\) in Vitro.**

To confirm the increased excision activity and diminished integration activity of PB\(^{M194V/R372A/K375A}\) and PB\(^{R372A/K375A/D450N}\), we purified transposase protein from E. coli and examined the double-strand break excision activity (Fig. 5A) and hairpin opening/target-joining activity (Fig. 5B) as described above. Both PB\(^{M194V/R372A/K375A}\) and PB\(^{R372A/K375A/D450N}\) had excision (Fig. 5A) and hairpin opening (Fig. 5B) activity at levels similar to or higher than wild type or the hyperactive mutant iP7. iP7 differs from wild type at seven amino acid positions (5, 8, 15). Both mutants PB\(^{M194V/R372A/K375A}\) and PB\(^{R372A/K375A/D450N}\) had diminished target joining, suggesting that the M194V or D450N mutations did not revert to the integration-defective phenotype (Fig. 5B). These data support the view that PB\(^{M194V/R372A/K375A}\) and PB\(^{R372A/K375A/D450N}\) are hyperactive for excision but are integration-defective.

**Excision and Integration Activities of PB\(^{M194V/R372A/K375A}\) and PB\(^{R372A/K375A/D450N}\) in Mammalian Cells.** Using PB\(^{M194V/R372A/K375A}\) and PB\(^{R372A/K375A/D450N}\), we performed GFP::PB excision assays in HEK293 cells, screening for increased GFP as described above. Both PB\(^{M194V/R372A/K375A}\) and PB\(^{R372A/K375A/D450N}\) had excision activities greater than PB\(^{R372A/K375A}\) and similar to wild-type PB (Fig. 6A). Combining the M194V and D450N mutations as in PB\(^{M194V/R372A/K375A/D450N}\) did not lead to any further increase in activity; rather, the frequency was decreased to below that of PB\(^{R372A/K375A}\).

When the M194V/R372A/K375A and R372A/K375A/D450N mutations were introduced into the hyperactive iP7 mutant background, excision activities for all mutants increased, and the effect of the hyperactive M194V and D450N mutations was more pronounced (approximately five- to sixfold) (Fig. 6B). Combination of the M194V and D450N mutations in the iP7\(^{R372A/K375A}\) backbone, however, resulted in excision activities decreasing to wild-type PB levels.

To confirm that the M194V and D450N mutations did not revert the integration defect conferred by the R372A/K375A mutations in the iP7 background, we analyzed their activity using a colony formation assay in HeLa cells as described above. We observed that iP7\(^{M194V/R372A/K375A}\) and iP7\(^{R372A/K375A/D450N}\) had very low integration activity, similar to that of the –Tps control (Fig. 6B). Thus, they are iP7 Exc\(^{Int}\) mutants that are hyperactive for excision but defective for integration, i.e., iP7 Exc\(^{Int}\) mutants. Paradoxically, the M194V and D450N mutations increase the excision activity of PB\(^{R372A/K375A}\) and iP7\(^{R372A/K375A}\) yet decrease the excision activity of PB and iP7 in mammalian cells (Fig. S3).

**Screening in Saccharomyces cerevisiae for PB\(^{R372A/K375A}\) Mutants That Are Hyperactive for Excision.** We previously reported that piggyBac transposes efficiently in S. cerevisiae (22). We used this genetically tractable system to isolate excision-hyperactive mutants of PB\(^{R372A/K375A}\). Similarly to the HEK293 GFP::PB assay described above, we used a fluorescence assay in yeast, in which PB promotes the precise excision of a piggyBac transposon from a fluorescent reporter gene mCherry, termed mCherry::PB (8). In this system, colonies are screened for increased mCherry expression to reflect increased excision. As expected, the –Tps negative control did not result in increased fluorescence, whereas mCherry-positive yeast were readily observed in the presence of wild-type PB (Fig. 4A). Consistent with the observations in mammalian cells, the excision promoted by PB\(^{R372A/K375A}\) in S. cerevisiae was reduced compared with wild-type PB (Fig. 4A). To identify mutations that would increase excision activity, we generated a PB\(^{R372A/K375A}\) mutant library by error-prone PCR mutagenesis and introduced it into the mCherry::PB strain by homologous recombination with a gapped pGAL5 PB plasmid. After transformation, single colonies were pin-replicated to SC (synthetic complete)—Trp–Ura+galactose plates and grown for 3 d, and mCherry fluorescence was analyzed using a Typhoon scanner or a fluorescence microscope.
Exc<sup>+</sup>Int<sup>−</sup> piggyBac Transposases Promote Precise Excision. One reason for isolating Exc<sup>+</sup>Int<sup>−</sup> PBs was to avoid element reintegration during transient transgenesis. Another goal for transient transgenesis is restoration of the transposon donor site to its original pretransposon sequence. We evaluated precise excision by the Exc<sup>+</sup>Int<sup>−</sup> transposases in yeast by determining the frequency at which the transposases promoted imprecise excision from the CAN1 gene, the arginine transporter. If cells are CAN1<sup>+</sup>, they are sensitive to the arginine analog canavanine. Inactivation of CAN1 results in canavanine resistance (CanR). We found that, following transposon excision, the imprecise excision frequencies of the Exc<sup>+</sup>Int<sup>−</sup> mutants scored by measuring the frequency of CanR upon excision in yeast are less than 0.3%, similar to wild-type PB (Table S1).

Fusion of iPB7 and iPB7 Exc<sup>+</sup>hyperInt<sup>−</sup> Mutants to Designer Zinc Finger Proteins That Bind to Putative Safe Harbor Regions of the Human Genome. Another motivation for generating the Exc<sup>+</sup>Int<sup>−</sup> PBs was to ask if their integration activity could be restored and redirected to selected genomic regions by fusion of designed ZFPs with engineered specificities to the transposase. We chose to fuse the Exc<sup>+</sup>hyperInt<sup>−</sup> iPB7s to ZFPs designed to target binding sites in putative safe harbor loci, the human ROSA26 gene, or the human L-gulono-γ-lactone oxidase (GULOP) pseudogene. Construction and analysis of the binding activity of the ROSA26 and GULOP ZFPs is described in Figs. S4–S6 in SI Materials and Methods.

We made fusion proteins linking the ROSA26 or GULOP ZFPs with a 15-amino-acid linker to the N terminus of iPB7, the Exc<sup>+</sup>hyperInt<sup>−</sup> transposase iPB7<sup>R372A/K375A/D450N</sup>, a transposase lacking the K375A mutation that was recovered during plasmid construction (Fig. 7). We used colony formation assays to test the integration activity of our chimeric transposases. Cotransfection into HeLa cells of the GULOP–iPB7 and the ROSA26–iPB7 fusion constructs with a transposon plasmid containing a puromycin resistance selectable marker resulted in about

---

**Fig. 3.** In vitro cleavage, hairpin opening, and target-joining assays of mutant piggyBac transposases. (A) Linearized transposon donor plasmid DNA was end-labeled (*) at both the 3′ end of the left (L) and right (R) end flanking sides. Cleavage at the appropriate piggyBac transposon terminal repeat (triangles) sites results in products of the indicated sizes: left flank cleaved (1.1 kb), right flank cleaved (630 bp), or both flanks cleaved. (B) Linearized donor plasmid DNA was incubated with the indicated wild-type or mutant PB for increasing amounts of time (1, 5, 25, and 125 min). The cleavage events of the L-end, R-end, or both ends are discerned by electrophoresis on an agarose gel. (C) Schematic representation of the hairpin opening and target-joining assay. A 74-nt hairpin DNA containing 35 bp of the piggyBac left end including the 13-bp terminal inverted repeat (triangle) and the TTAA hairpin was end-labeled (*) at the 5′-end. Appropriate cleavage of the 4-bp terminal hairpin results in a 35-nt opened hairpin product on a denaturing gel. Incubation of the opened hairpin with PB and a circular DNA target 2.7-kb plasmid results in single-end or double-end joining as indicated. (D) End-labeled hairpin DNA was incubated with the indicated wild-type or mutant PB for increasing amounts of time (1, 3, 10, and 20 min), and products were discerned by electrophoresis on a denaturing gel.
80–100% of the number of drug-resistant colonies compared with those of the unmodified iP7 control. Fusion of the GULOP ZFP to the Exc$^+$Int$^-$ transposase iP7R372A/K375A/D450N did not restore integration activity.

Notably, however, we did see restoration of integration activity with a ZFP fusion in two cases. The integration activity of iP7R372A/K375A/D450N, a transposase lacking the K375A mutation was less than 10% of iP7. Fusion of both the ROSA26 and GULOP ZFPs to this integration-defective transposase mutant iP7R372A/K375A/D450N did restore integration activity (Fig. 7). These results support the hypothesis that engineered ZFPs fused to the iP7R372A/K375A/D450N can restore the DNA target-binding function. Comparison of the integration patterns of the ZFP fusion and the ZFP fusion transposases (Table S2 and Fig. S7) did not, however, reveal discernible differences.

**Discussion**

Several properties make piggyBac a particularly useful tool for genetic manipulation. It is active in many cell types, including mammalian cells (9, 14, 25), and is distinguished by its ability to excise precisely (11), restoring the donor site to its preinsertion sequence. In addition, PB and iP7 remain active when fused to heterologous DNA-binding domains. We have manipulated the target recognition and joining properties of the PB to create a Exc$^+$Int$^-$ PB that probes the piggyBac transposition mechanism and is a useful tool for genome engineering.

Here, we have isolated mutants that will be valuable for reversible transgenesis. Several highly conserved acidic residues (D268, D346, and D467) in PB are essential for DNA breakage and joining (22, 23). We previously suggested that they lie on an RNase H-like fold as is found in other transposases and retroviral integrases (21). Such active-site domains must also contain amino acids that interact with the DNA substrates of transposition. As a strategy to identify amino acids that interact with the target DNA, we substituted highly conserved basic amino acids in the catalytic domain to alanine. We also mutated two other amino acids that correspond to residues in HIV integrase that influence target joining (24). Most of these mutations had no effect on transposition or resulted in the loss of both excision and integration activity. However, one mutant, PB$^{R372A/K375A}$, had only slightly decreased excision activity but was highly defective in integration in mammalian cells. These amino acids are good candidates for interacting with target DNA, and future studies of their mechanism will provide insights into transposase structure and function.

We used two strategies to improve the modestly decreased excision activity of the integration-defective PB$^{R372A/K375A}$ mutant. First, we used a screen in yeast to identify two additional mutations, M194V and D450N, that each increased the excision.

---

**Fig. 4.** Mutagenesis screen for hyperactive excision variants of PB$^{R372A/K375A}$ in yeast and mammalian cells. (A) Error-prone PCR was used to generate PB$^{R372A/K375A}$ transposase variants. The piggyBac transposon donor plasmid, mCherry::PB-TrKanMX::URA3, and the PB$^{R372A/K375A}$ transposase variant library were introduced to yeast and grown on selective media. mCherry fluorescence intensity indicates relative precise excision frequency. The corresponding genotypes and phenotypes of the potential hyperactive excision variants are indicated. (B) Promising variants identified in the yeast-based excision assay were retested in a mammalian cell-based excision assay. The indicated PB$^{R372A/K375A}$ variants were transiently transfected into HEK293 GFP::PB cells. The precise excision of the piggyBac transposon from the genome results in a repaired GFP expression cassette. The frequency of excision is indicated by GFP fluorescence intensity and determined by FACS analysis. No transposase (–Tps), wild-type PB, and PB$^{R372A/K375A}$ transposase were included as controls in both assays.
activity of PB<sup>R372A/K375A</sup> to levels above the excision frequency of wild-type PB. We also tested the effect of the M194V/R372A/K375A and R372A/K375A/D450N mutations on a hyperactive version of the transposase (iPB7) (5, 8, 15) that contains seven changed amino acids. We found that iPB7<sup>M194V/R372A/K375A</sup> and iPB7<sup>R372A/K375A/D450N</sup> had increased excision activity but were still integration-defective.

The mechanism by which the M194V and D450N mutations suppress the excision defect imposed by R352A/K375A remains to be determined. M194V is located in the conserved C terminus of PB, upstream of the catalytic domain, the function of which is not yet known (23). In a screen for hyperactive PB mutants in yeast, we also isolated a M194V mutant but found that it is not hyperactive in mammalian cells (15). D450 is highly conserved among piggyBac (22, 23) and is closely positioned to D447, which is essential for DNA breakage and joining and is part of the RNase H-based motif of conserved acidic amino acids.

One reason to isolate the Exc<sup>+</sup>hyperInt<sup>−</sup> PB was to facilitate reversible transgenesis without harmful transposon reintegration. Another was to facilitate construction of transposases that direct insertion to chosen specific sites. Studies show that the Sleeping Beauty transposase is not very amenable to this strategy; for example, addition of a DNA-binding domain either ablates or greatly diminishes Sleeping Beauty transposase activity (10). Hyperactive Sleeping Beauty enzymes can be active as a fusion ZFP transposase; however, precise genomic site-directed integration has not been achieved using such methods (20). PB is still highly active when fused to a heterologous DNA-binding domain at its amino terminus (10). We speculated that decreasing the intrinsic targeting activity would make target-site selection more dependent upon a heterologous DNA-binding domain in a fusion protein.

Thus, we fused the Exc<sup>+</sup>hyperInt<sup>−</sup> PB to two designed ZFPs that bind to the human GULOP and ROSA26 region that we hypothesized would be safe harbors for PB integration. Despite the significant increase in the frequency of integration with GULOP-

Fig. 5. In vitro cleavage, hairpin-opening, and target-joining assays of Exc<sup>+</sup>hyperInt<sup>−</sup> mutant transposases. (A) Linearized donor plasmid DNA was end-labeled and incubated with the indicated transposase mutant for increasing amounts of time (1, 5, 25, or 125 min) as shown. The cleavage events of the L-end, R-end, or both ends are discerned by electrophoresis on an agarose gel. The single cleavage products—transposon plus the right flank (Tn+Rf) and transposon plus the left flank (Tn+Lf)—are indicated. (B) A 84-nt hairpin DNA containing 40 bp of the piggyBac transposon left end including the 13-bp terminal inverted repeat (triangles) and the TTAA hairpin was end-labeled at the 5′ end. Labeled hairpin was incubated with the indicated PB mutant and a circular DNA target for increasing amounts of time (1, 5, 25, and 125 min), and the products were discerned by electrophoresis on a denaturing gel. Appropriate cleavage of the 4-bp terminal hairpin results in a 40-nt opened hairpin fragment and target joining to pUC19 (2.7 kb). Wild-type PB, PB<sup>R372A/K375A</sup>, and iPB7 were included as controls in both assays.
allowing reversible transgenesis (12, 15, 26). Thus, piggyBac can be exceptionally useful in generating iPSCs without a permanent DNA sequence change. The genes for the necessary transcription factors for iPSC generation are integrated into the target genome using a piggyBac transposon as the vector that can serve as a long-term source of transcription factors. Once iPSC cell transformation has occurred, however, the piggyBac vector can then be re-exposed to transposase and excised by its natural “precise excision” pathway, restoring the genome to its pretransposon state. One potential hazard, however, is that the excised transposon may re-integrate. Wild-type PB mediates reintegration about 40–60% of the time after excision (14) and thus is a significant hazard. Such reintegration, however, will not occur with Exc+Int transposase, proving a safer pathway for genome engineering.

Combining the technologies of ZFP engineering and transposase enzymatic biology presents an opportunity to create tools for genomic engineering. These tools have the potential to improve the safety and utility of delivery vehicles for therapeutic genes, both in vivo and ex vivo.

**Materials and Methods**

* piggyBac Transposase Mammalian Expression Vector Constructs. The piggyBac transposase helper plasmid pXL-PB-El for mammalian expression contains a Flag-tag fused to the N terminus of PB in pcDNA3.1/myc-HisA (InvivoGen). The piggyBac ORF was amplified using the primer N-Flag-PM-5f (GGGAGATTCCTGCGACCATGGACTAACAAGGAGCAGTAGACAAA atggtagctttcagga-tgg (EcoRI, Italic; GCCACC, Kozak sequence; Flag tag, upper case; piggyBac, lower case)) and primer PB-4r (GGGGGCGGGCtacagaaactttggcacatatcataattatg (Ntot, italic; piggyBac, lower case)), digested with EcoRI and Ntot, and cloned into EcoRI- and Ntot-digested pcDNA3.1/myc-HisA vector. The initial Exc+Int mutant candidates were obtained by site-directed mutagenesis. All hyperactive excision candidates obtained by screening in yeast plasmids were recloned into pcDNA3.1/myc-HisA for testing in mammalian cells. The PB ORFs were PCR amplified with primers PB-f GAGGTACCAGAGCTCAATAGTTCTTGCGGTG and pbexr2 (GAGAGTGCACCATATATGCGGTG), and the products were digested with KpnI and NotI and cloned into EcoRI- and NotI-digested pcDNA3.1/myc-HisA vector containing a transposon used for excision and integration assays of the Exc+Int candidates, as well as the PB integration libraries, has been previously described (13, 17).

**Mammalian Transposon Donor Plasmid.** The transposon donor plasmid pXL-PB-D-GFPBlasticidinR (Bsd) containing a transposon used for excision and integration assays of the Exc+Int candidates, as well as the PB integration libraries, has been previously described (13, 17).

**Mammalian Cell Culture.** HEK293 and HeLa cells were cultured in DMEM supplemented with 5% (vol/vol) FBS.

**Donor Repair Excision Assay in Mammalian Cells.** HEK293 cells (2×10^6) were transfected with 8 μg transposon donor plasmid pXL-PB-D-GFP/Bsd and 2 μg piggyBac transposase helper plasmid pXL-PB-El or the Exc+Int candidate mutants with FuGENE 6 transfection reagent (Roche Applied Science). Three days later, cells were harvested and the helper, donor, and repaired donor plasmids were recovered by the Hirt method (27). The plasmid DNA was used as a PCR template with two primers against the DNA sequence flanking the transposon in the donor plasmids pBEXl2 (GGAAAGGGAGGGCCAGCAG) and pBEX2 (GAGAGTGCACCATATAGTTCTTGCGGTG), and the products were displayed on a 1% agarose gel in 1× trisborate/EDTA buffer (TBE).

**Mammalian Colony Formation Assay for Integration.** HeLa cells (3–5×10^5) were transfected with 1.6 μg transposon donor plasmid pXL-PB-D-GFP/Bsd and 0.4 μg pXL-PB-El helper plasmid and Exc+Int mutant transposase plasmids with FuGENE 6 transfection reagent (Roche Applied Science). Two days later, cells were trypsinized and diluted, and growth continued in DMEM containing 3.5 μg/mL of blasticidin for 18–21 d. The medium was changed every 2–3 d. Surviving colonies were fixed with 4% (wt/vol) paraformaldehyde, stained with 0.2% methylene blue, and counted.

**piggyBac Transposase Purification.** PB wild type and selected Exc+Int excision transposase proteins were expressed in E. coli BL21 codon plusDE3 RIL (Stratagene) using the pET21b vector (Novagen) with a C-terminal His-tag and purified by chromatography on a Ni2+ column as previously described (22).

**In Vitro Cleavage of Linear Donor Plasmid DNA.** The donor plasmid pXL-PB-D-GFP/Bsd was linearized by digestion with Psp64I and 3’ end-labeled with
element was piggyBac or a ORF, leading to colony piggyBac insertion. Precise excision of the puromycin, stained with methylene blue, counted, and normalized to the unmodified IPB7 control.

TTAA were synthesized (Integrated DNA Technologies), PAGE gel-puriﬁed and labeled at the 5′ end with \([\gamma-\text{32P}]\text{ATP. Ten-microliter reactions were performed as previously described (22), stopped at times indicated in theﬁgure legends by addition of 500 mM EDTA to 40 mM. Samples were spin-dried, resuspended in 2x (vol/vol) STOP solution (United States Biologicals), boiled for 5 min, chilled on ice, and displayed on a 1% acrylamide–7 M urea–1x TBE denaturing acrylamide gel. The gel was dried, exposed to a PhosphorImager screen, and analyzed by Imagequant software.

Excision Assay in Yeast Using a Fluorescent Reporter. This assay was previously described by Burnight et al. (8). Brieﬂy, a GFP gene containing the piggyBac transposon GFP::PB was introduced into the genome of HEK293 cells. No GFP is detectable because of the piggyBac insertion in GFP. Excision occurs in the presence of transposase introduced by transfection with wild-type or various mutant plasmids and FuGENE 6 transfection reagent. Four days later, cells were harvested and FACs-analyzed with FACS Calibur (BD Biosciences) to measure the percentage of GFP+ cells in the entire cell population. Data were analyzed with FlowJo 8.5.3. GFP+ cells were counted against total cells, and the number of GFP+ cells from wild-type PB was used to normalize the readout of various mutant PBs.

Construction of can1::PBTn-KanMX/URA3. A piggyBac transposon carrying KanMX and URA3 was constructed by PCR ampliﬁcation of PBTn-KanMX from mCherry::PB-TnKanMX with Ura3+ (5-FOAS) because of the mCherry fluorescent protein because of the piggyBac transposon insertion. Precise excision of the piggyBac element restores the mCherry ORF, leading to colony ﬂuorescence that can be detected by a Typhoon scanner or a ﬂuorescence microscope (Leica; M165FC). In this work, the piggyBac element was PB-TnKanMX. The yeast strain was BY4727 (28) MATA his3Δ1200 leu2Δ10Δ1 lys2Δ10 met5Δ0 trp1Δ63 ura3Δ0. The wild-type and R372A/K375A mutants of piggyBac transposase ORFs were cloned into the XmaI and XhoI sites of the p414GALS vector (29).

Isolation of Hyperactive Derivatives of the Reduced Excision Activity of pbR372A/K375A. We mutagenized the Pb R372A/K375A ORF by error-prone PCR in the presence of Mn2+ with primers from the p414GALS promoter and CYC1 terminator region, using p414GALS Pb R372A/K375A as template. The resulting library was introduced into a linearized p414GALS TRP1 vector digested with XmaI and XhoI by homologous recombination in yeast BY4727 containing the mCherry::PB-TnKanMX URA3 donor plasmid. Transformant colonies were selected on synthetic complete (SC) plates lacking tryptophan and uracil (SC-Trp-Ura).

To identify Exc−/Int− mutants, more than 6,000 individual transformants were picked in 96-well SC-Trp−Ura + 2% (vol/vol) galactose plates. We scored transposition by looking for increased ﬂuorescence over 1–3 d, and 30 hyperactive candidates were found in the primary screen. Yeast plasmid DNAs were isolated, and the mutant Pb R372A/K375A ORFs were sequenced and reassembled, yielding M194V and D450N as the best suppressors.

Mammalian GFP Excision Assay. This system was previously described by Burnight et al. (8). Briefly, a GFP gene containing the piggyBac transposon GFP::PB was introduced into the genome of HEK293 cells. No GFP is detectable because of the piggyBac insertion in GFP. Excision occurs in the presence of transposase introduced by transfection with wild-type or various mutant plasmids and FuGENE 6 transfection reagent. Four days later, cells were harvested and FACs-analyzed with FACS Calibur (BD Biosciences) to measure the percentage of GFP+ cells in the entire cell population. Data were analyzed with FlowJo 8.5.3. GFP+ cells were counted against total cells, and the number of GFP+ cells from wild-type PB was used to normalize the readout of various mutant PBs.

Assay of Imprecise piggyBac Excision from can1::PBTn-KanMX/URA3, can1::PBTn-KanMX/URA3 is Can+ because of the piggyBac insertion in the arginine transporter CAN1 and is Ura− (5-FOAS) because of URA3 in the piggyBac.
transposon, p414GAL5-piggyBac transposase plasmids expressing PB, pR372A/pEFGGGGSGGGGSGGGGSQF, pR372A/pR372A-pEFGGGGSGGGGSGGGGSQF, and pIB7 transposase were transformed into can1Δ::PBR-Kan/URA3 by selection on SC-Trp-Arg+Can+G418 plates and then restreaked on SC-Trp-Arg+Can+ 2% (wt/vol) galactose to induce transposase expression. Colonies were resuspended in water and plated on SC-Trp-Arg+Can+S-FOA to measure the number of cells that underwent imprecise excision—i.e., were CanR without element reintegration—and on SC-Trp-S-FOA to measure cells that underwent excision without transposon reintegration. The frequency of imprecise excision is calculated as the ratio of the number of cells that underwent imprecise excision without element reintegration to the number of cells that underwent excision without transposon reintegration. It should be noted that this underestimates the imprecise excision frequency for transposases that are IntR because imprecise excision events in cells in which reintegration occurs (about 40–60% of excisions) are not considered.

**Designed ZFP–Transposase Fusion Plasmid Constructs.** The ZFP–pIB7 chimeric transposase plasmids were constructed by inserting each of the ZFP-linker cDNAs into the ecoRI/XbaI sites upstream of the pIB7 sequence (5, 15). The ZFPs and transposases are separated by the amino acid linker sequence EFGGGGGSGGGGSGGGGSQF. The described ZFP–pIB7 mutants were constructed by mutating different combinations of three amino acids (R372A, K375A, and/or D450N) via site-directed mutagenesis (Quickchange Site-directed Mutagenesis Kit, Stratagene).

**Mammalian Transposition Assays Using a Puromycin-Resistant Transposon with the ROSA26 and GULOP ZFP Fusions to piggyBac.** A total of 400 ng of pXLT-Bac PB-GFP/Puro transposon plasmid (8) alone or with 400 ng of the indicated transposase plasmids was transfected into HeLa cells using Lipofectamine 2000 (Invitrogen) according to the manufacturer’s instructions and cultured under puromycin selection (0.5 μg/mL) for 2 wk. Following selection, puromycin-resistant colonies were fixed with 4% (wt/vol) paraformaldehyde, stained with methylene blue, and counted.

**ACKNOWLEDGMENTS.** We thank Dr. Rupak Mitra and Dr. Courtney Busch for many discussions about piggyBac; Benjamin Brett, Dr. Todd Scheetz, and Dr. Tom Bair for their expert consultation; and Maureen R. Regan for technical assistance. This work was supported by a grant from the Maryand Stem Cell Research Fund (to N.L.C.). N.L.C. is an investigator of the Howard Hughes Medical Institute. This work was supported by Grants AI 052845 and AI088020 (to F.D.B.). J.D.S. was supported by National Institutes of Health (NIH) Grant T32 CA09216. J.K.J. was supported by NIH Director’s Pioneer Award DP1 GM105378 and by NIH Grant R01 GM088040. This work was supported by the NIH Grant R01 HL-105821 (to P.L.S. and J.K.J.) and partially supported by the University of Iowa Center for Gene Therapy (NIH P30 DK-54759).

Supporting Information

Li et al. 10.1073/pnas.1305987110

SI Materials and Methods

Construction and Testing of Zinc Finger Proteins. Several criteria have been put forth to define genomic safe harbors (1, 2). Ideally, a safe harbor should be distant from the 5′ end of a gene, and especially distant from any oncogene. Gene addition should be outside a transcriptional unit, including microRNAs, and outside ultraconserved regions of the human genome. The location must be accessible to allow the transposon and transposase to reach the target sequence, thereby promoting efficient integration. A transcriptionally active region would help to ensure that the DNA is accessible and may be required to ensure stable expression of the therapeutic inserted transgene. We chose human ROSA26 and gulono-γ-lactone oxidase (GULOP) loci as two candidate safe harbors. We reasoned that, because mouse Rosa26 is a target for many site-specific insertions of foreign DNA with no known adverse effects, the human ROSA26 (3) also represents a safe harbor candidate. GULOP is a unitary pseudogene that is far distant from neighboring transcriptional units. In most nonhuman mammals, GULOP synthesizes the precursor of L-ascorbic acid (vitamin C); however, in humans, the majority of the gene has been deleted, and within the remaining sequence several anomalous nucleotide changes have occurred (4, 5). None of the genes flanking GULOP or ROSA26 are known tumor suppressors or oncogenes. Neither candidate encodes a protein product, although ROSA26 encodes a noncoding RNA.

To identify regions within these genes that are rich in piggyBac target sequence sites TTAAs, we developed a scoring algorithm that analyzed TTA density for indicated regions (Fig. S4). For each TTA, the number of adjacent sites was determined within a given window. A 128-bp window on either side of each site was used; thus the score denotes the TTA density within a 256-bp sliding window.

Six-finger zinc finger arrays were assembled using two-finger zinc finger units as previously described (6). Two-finger units, each expected to specify 6 bp of DNA, were chosen from three-finger zinc finger proteins (ZFPs) engineered by the oligomerized pool engineering method or used to practice the context-dependent assembly method (7, 8). Using these two-finger units, we assembled six-finger arrays targeted to TTA-rich regions within the ROSA26 and GULOP sites (Fig. S4).

Bacterial Two-Hybrid and Mammalian One-Hybrid Assays. The GULOP and human ROSA26 zinc finger proteins were assayed for activity using a bacterial two-hybrid–based reporter system (7, 8) (Fig. S5). β-Galactosidase assays for assessing the DNA-binding activities of zinc finger proteins in a bacterial two-hybrid assay were performed as described previously (8). Mammalian one-hybrid assays were performed as described previously (9) (Fig. S5). Briefly, the activation plasmids were constructed by inserting cDNA encoding a C-terminal fusion of the herpes simplex virus protein 16 activation domain and each of the engineered ZFPs into the BamHI/XhoI-digested pCAGGs backbone. The ZFP target reporter plasmid was constructed by annealing oligos containing four copies in tandem of the ZFP target sequence and cloning the annealed oligos upstream of a minimal human thymidine kinase promoter driving firefly luciferase in the pTATA vector (a kind gift from James Darnell, Laboratory of Molecular Cell Biology, The Rockefeller University, New York, NY). HeLa cells were transfected with 0.4 μg each of ZFP activator and target reporter plasmids using Lipofectamine 2000 (Invitrogen) as directed by the manufacturer. Transiently transfected cells were harvested in 1× Passive Lysis Buffer (Promega) after 48 h. Twenty-μl lysates were assayed using the Luciferase Reporter Assay System (Promega) according to the manufacturer’s instructions. Based on the results of these assays, we selected the ZFPs termed ROSA3b and GULOP1b for further use in this study. The ROSA26 target site is GATGCTGTAGGGATGCA (58% GC) and the GULOP target site is TGGGATGCAAGCCAGATG (58% GC). The DNA sequences of the ZFPs are shown (Fig. S6).

Integration-Site Recovery for Illumina HiSeq2000 Sequencing. Integration sites were recovered as described (10). Briefly, HeLa cells (5 × 10⁶) were transfected with 10 μg pXL-BacII PB-GFP/Puro transposon plasmid and 2 μg of each transposase plasmid, and then integrants were selected with puromycin (0.5 μg/mL) for 3 wk. Genomic DNA from three separate transfections was extracted from the integration library using the DNeasy tissue kit (Qiagen). Pooled DNA (2 μg) was digested overnight with Apol or BstYI at 50 °C and 60 °C, respectively; DNA fragments were purified with the QIAquick PCR purification kit (Qiagen) and ligated to Apol and BstYI linkers overnight at 16 °C. Nested PCR was carried out under stringent conditions using the transposon end-specific primers AACCCCTCGATATACAGACCGATAAACACATGCAGTCAATTTCAGCG (primary) and AATGACCCGGCACCGAGTTCCAACAGGACCTTTCACCCTTCAGAGCGCCTTCCGAACCCCTCAACTCTTTCCCTTCCCTACAGGAGACCTTGCACTGTCACCAATAGTATATAAGGCCTTCTTTCGACCGATCT(secondary; XXXX denotes bar code; underlined sequence indicates Illumina cluster-generation sequence) and linker-specific primers CGTACCAACGAGCAGAGAAGCGGAGGACCAACAGTATGCAATTTTACGCAACAGCGACCGACCCAGAT (primary) and CAAGCAGAAGACGGCATACGAGCTCTTCGGTTCTTCCCTACAGGACCTTTCACCCTTCAGAGCGCCTTCCGAACCCCTCAACTCTTTCCCTTCCCTACAGGAGACCTTGCACTGTCACCAATAGTATATAAGGCCTTCTTTCGACCGATCT(secondary). DNA barcodes were included in the second-round PCR primers to track sample origin. The PCR products were gel-purified, pooled, and sequenced using the Illumina HiSeq2000 sequencing platform.

Reads from each flow cell lane were trimmed according to the barcodes and linkers expected, using a custom R wrapper for the BioStrings trimLRPatterns function (11) and allowing no mismatches in the barcode and up to two mismatches in the linker sequence. Trimmed reads were aligned to the hg18 human genome build using Bowtie (12), allowing two mismatches in each alignment and requiring the alignment to be unique.

Insertion-site coordinates were sorted and collapsed; multiple reads often mapped to a single site. Furthermore, many sites with large numbers of reads were immediately flanked by a few sites with one or two reads. Upon examination, these nearly always prove to be slight alignment errors. Thus, insertion counts in this configuration are collapsed into the site with the most counts, using a simple Perl script that scans for insertions mapping to adjacent positions. This leaves a set of sites, each associated with a number of mapped insertions. As we do not know whether multiple recovered insertions are real or are PCR artifacts, we proceed with the analysis using only the sites. For a subset of the sites, we have recovered insertions in both orientations (on the + and the − strand). These are necessarily independent events, and these “bidirectional” sites are noted separately.

For genome-wide feature correlation analysis, we could not include all sites, due to computational limitations. Thus, we included all bidirectional sites for each of the experiments in HeLa cells, as well as a randomly chosen subset of sites whose insertion counts were in the third quartile of the insertion counts for all sites, reasoning that these should be strong sites, yet representative of the insertion landscape for each experiment. After this...
process we had subsets of roughly 2,500–3,000 sites for each of the experiments. Initial sites, insertion counts, and bidirectional status (0 if not bidirectional, 1 otherwise) are provided as supplemental -s; the files labeled “sites_analyzed” are those that were included in the genome-wide analysis and the others contain the full list of sites for each element. R and Perl scripts are available upon request.

Fig. S1. Protein sequence alignment of piggyBac family members. The catalytic domain of eight piggyBac transposase family members were aligned to Trichoplusia ni (Tni) (1, 2). Blue boxes indicate the requisite catalytic amino acids, red boxes indicate conserved argenines and lysines, and green boxes indicate the positions of HIV integrase mutations with known altered target joining in HIV integrase (3).


Fig. S2. Colony formation assay with PB<sup>R372A</sup> and PB<sup>K375A</sup> individual mutations. HeLa cells were transiently cotransfected with a transposon expressing blasticidin<sup>r</sup> and the indicated mutant piggyBac (PB) transposase. Cells were selected for blasticidin resistance and stained with methylene blue to identify viable cell colonies. No transposase (−Tps), wild-type transposase (PB), and PB<sup>R372A/K375A</sup> transposase were included as controls.

Li et al. www.pnas.org/cgi/content/short/1305987110

3 of 7
Fig. S3. Excision assays of individual M194V and D450N mutations on the PB or iPB7 transposase backbones. The indicated wild-type PB (Left) or insect-derived piggyBac transposase 7 (iPB7) (Right) mutants were transiently transfected into HEK293 GFP::PB cells. The frequency of excision is indicated by GFP fluorescence intensity, determined by FACS analysis, and normalized to the wild-type PB control. No transposase (−Tps) and unmodified iPB7 were included as additional controls.

Fig. S4. Schematic representation of ZFP target genomic loci. (A) The human ROSA26 locus is flanked by the THUMPD3 and SETD5 genes on chromosome 3p25.3. (B) The GULOP pseudogene is flanked by EPHX2 and CLU in chromosome 8p21.1. Gray arrows indicate approximate ZFP target sites. The TTAA density score was determined for a given 256-bp window.
**Fig. S5.** Engineered ZFP activity in cells. (A) A bacterial two-hybrid (B2H) assay was used to assay activity of engineered ZFPs. A six-finger ZFP is fused to the Gal11P fragment, shown schematically. ZFP binding to its target recruits RNA polymerase to a weak promoter driving the reporter LacZ gene in bacteria through interaction of the GAL4 domain fused to the RNAP. Eight ZFPs targeting four sites at or near the ROSA26 locus or eight ZFPs targeting four sites at or near the GULOP locus were evaluated. Bars represent LacZ activity in bacteria transformed with the ZFP library. The dashed line represents an arbitrary threshold at which B2H activity is typically effective in mammalian cells. (B) For mammalian one-hybrid (M1H) assays, activator plasmids expressing the ZFPs fused to the VP16 activation domain from Herpes Simplex Virus 1 were cotransfected with plasmids containing four copies of the target sequence upstream of a minimal promoter driving firefly luciferase, shown schematically. ZFP binding to its target sequence activates luciferase transcription. ZFP activity is reported as a function of luciferase activity. Bars represent mean fold activation in cells transfected with activator and reporter plasmids relative to luciferase activity in cells transfected with reporter alone. n = 3.

**Fig. S6.** Sequences of the GULOP and ROSA26 ZFPs. The primary sequence of the DNAs encoding the GULOP or ROSA26 ZFPs are shown.
**Fig. S7.** Distribution of iP7, GULOP-iPB7, GULOP-iPB7 R372A/D450N, ROSA26-iPB7, and ROSA26-iPB7 R372A/D450N–mediated insertions in the human genome. Integration-site datasets for ZFP–iPB7-mediated insertions are indicated by the columns, and genomic features or ChIP-Seq datasets are indicated by the rows (the latter were calculated over 10-kb windows). The departure from random distribution is indicated by colored tiles (key at bottom), and differences from random placement were scored using the Receiver Operator Characteristic (ROC) area method described previously (1). A detailed explanation of the variables studied can be found in Ocwieja et al. (2) or at [http://microb230.med.upenn.edu/assets/doc/HeatMapGuide_v12_formatted.doc](http://microb230.med.upenn.edu/assets/doc/HeatMapGuide_v12_formatted.doc).

(A) The integration frequency relative to selected genomic features is shown. Red shading indicates features where insertions are favored compared with random, whereas blue shading indicates unfavored integration events. Gray indicates random distribution. The distribution of HIV-, MLV-, and Adeno Associated Virus-mediated integrations are shown for comparison. (B) The integration frequency relative to bound proteins and modified histones was mapped using the ChIP-Seq method. Yellow and blue are used to indicate depletion or enrichment, respectively.

Table S1. Imprecise excision by excision competent hyper/integration defective mutant transposases

<table>
<thead>
<tr>
<th>Transposase</th>
<th>Imprecise repair, %</th>
</tr>
</thead>
<tbody>
<tr>
<td>PB</td>
<td>0.13*</td>
</tr>
<tr>
<td>pgR372A/K375A</td>
<td>0.14</td>
</tr>
<tr>
<td>pgM194V/R372A/K375A</td>
<td>0.21</td>
</tr>
<tr>
<td>pgR372A/K375A/D450N</td>
<td>0.27</td>
</tr>
<tr>
<td>iP87</td>
<td>0.24*</td>
</tr>
</tbody>
</table>

*Imprecise excision frequencies were determined as described in Materials and Methods. The imprecise excision frequency of Int+ transposases is underestimated by 40–60% because imprecise excisions that are accompanied by transposon reintegrations are not counted.

Table S2. Illumina sequencing of ZFP–iPB7-mediated genomic integrations

<table>
<thead>
<tr>
<th>Element</th>
<th>Reads</th>
<th>Alignments</th>
<th>Initial sites</th>
<th>Collapsed sites</th>
<th>TTAA sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>iP87</td>
<td>26,300,573</td>
<td>3,011,317</td>
<td>45,523</td>
<td>43,984</td>
<td>40,800</td>
</tr>
<tr>
<td>GLO-iPB7</td>
<td>79,825,963</td>
<td>9,897,552</td>
<td>61,914</td>
<td>58,900</td>
<td>54,803</td>
</tr>
<tr>
<td>GLO-iPB7R372A/D450N</td>
<td>94,236,814</td>
<td>11,829,337</td>
<td>74,210</td>
<td>70,393</td>
<td>66,379</td>
</tr>
<tr>
<td>ROSA-iPB7</td>
<td>85,295,144</td>
<td>10,691,398</td>
<td>62,795</td>
<td>59,609</td>
<td>55,924</td>
</tr>
<tr>
<td>ROSA-iPB7R372A/D450N</td>
<td>49,842,350</td>
<td>6,964,687</td>
<td>41,599</td>
<td>39,659</td>
<td>37,033</td>
</tr>
</tbody>
</table>

The number of total mapped integration reads and unique alignments for each ZFP–iPB7 chimera and unmodified iP87 control are indicated and were determined as described in Materials and Methods and SI Materials and Methods. Collapsed sites, TTAA+ non-TTAA insertion sites; TTAA sites, only TTAA insertion sites.