

# A resurrected mammalian *hAT* transposable element and a closely related insect element are highly active in human cell culture

Xianghong Li<sup>a</sup>, Hosam Ewis<sup>a</sup>, Robert H. Hice<sup>b</sup>, Nirav Malani<sup>c</sup>, Nicole Parker<sup>a</sup>, Liqin Zhou<sup>a</sup>, Cédric Feschotte<sup>d</sup>, Frederic D. Bushman<sup>c</sup>, Peter W. Atkinson<sup>b</sup>, and Nancy L. Craig<sup>a,1</sup>

<sup>a</sup>The Howard Hughes Medical Institute and Department of Molecular Biology and Genetics, Johns Hopkins University School of Medicine, Baltimore, MD 21205; <sup>b</sup>Department of Entomology and Institute of Integrative Genomic Biology, University of California, Riverside, CA 92521; <sup>c</sup>Department of Microbiology, Perelman School of Medicine at University of Pennsylvania, Philadelphia, PA 19104; and <sup>d</sup>Department of Biology, University of Texas, Arlington, TX 76019

Contributed by Nancy L. Craig, January 10, 2012 (sent for review November 28, 2011)

**Chromosome structure and function are influenced by transposable elements, which are mobile DNA segments that can move from place to place. *hAT* elements are a superfamily of DNA cut and paste elements that move by excision and integration. We have characterized two *hAT* elements, *TcBuster* and *Space Invaders (SPIN)*, that are members of a recently described subfamily of *hAT* elements called *Buster* elements. We show that *TcBuster*, from the red flour beetle *Tribolium castaneum*, is highly active in human cells. *SPIN* elements are currently inactive elements that were recently highly active in multiple vertebrate genomes, and the high level of sequence similarity across widely diverged species and patchy phylogenetic distribution suggest that they may have moved between genomes by horizontal transfer. We have generated an intact version of this element, *SPIN<sub>ON</sub>*, which is highly active in human cells. In vitro analysis of *TcBuster* and *SPIN<sub>ON</sub>* shows that no proteins other than transposase are essential for recombination, a property that may contribute to the ability of *SPIN* to successfully invade multiple organisms. We also analyze the target site preferences of de novo insertions in the human genome of *TcBuster* and *SPIN<sub>ON</sub>* and compare them with the preferences of *Sleeping Beauty* and *piggyBac*, showing that each superfamily has a distinctive pattern of insertion. The high-frequency transposition of both *TcBuster* and *SPIN<sub>ON</sub>* suggests that these transposon systems offer powerful tools for genome engineering. Finally, we describe a *Saccharomyces cerevisiae* assay for *TcBuster* that will provide a means for isolation of hyperactive and other interesting classes of transposase mutants.**

*hAT* element | target site selection | gene therapy | insertional mutagenesis | transgenesis

**T**ransposable elements are discrete DNAs that can move within and occasionally, between genomes. A revelation of the genomic age is that a considerable fraction of many genomes derives from these elements (1), including about one-half of our own genome (2) and >85% of the maize genome (3). Transposable elements can have considerable impact on genome function because of their ability to move and thus, rearrange DNA; additionally, they often contain regulatory elements such as enhancers, promoters, or terminators. Thus, they are potent agents of natural genome engineering. Transposable elements have also been harnessed by researchers and used in many model organisms for insertional mutagenesis and transgenesis, including gene therapy (4). The most commonly exploited transposable elements are cut and paste DNA transposons, which are mobilized through excision and reinsertion of a DNA intermediate (5). The most popular systems currently used in vertebrate cells are derived from three different DNA transposon superfamilies: *Sleeping Beauty*, a *Tc1/mariner* element resurrected from fish (6), *Tol2*, a fish member of the *hAT* superfamily (7), and *piggyBac*, an insect member of the *piggyBac* superfamily (8). Important considerations in using these elements as tools are their transposition frequency and target site selection. Having elements from distinct families that are not cross-mobi-

lizable is advantageous, because it allows for the use of multiple systems in the same cell.

The effects of transposable elements on genomes are not restricted to cells that are related by vertical descent (i.e., from mother to daughter). Transposable elements can also spread between the genomes of different cells. Bacteriophage  $\mu$  and retroviruses, which integrate into their host's genomes by transposition, are elements that naturally move between cells. Horizontal transfer between species is very common in bacteria (9, 10), but it is much less frequently observed in eukaryotes and particularly, metazoans, where the sequestration of the germ line provides a significant barrier to horizontal transmission (11). Multiple examples of horizontal transmission have been suggested, however, by the observation of nearly identical transposons, which display a patchy taxonomic distribution and phylogenies that are incongruent with those phylogenies established from the analysis of chromosomal genes (11). There have been several recent descriptions of how currently inactive DNA transposons have likely undergone horizontal transmission between tetrapods. The impressive amplification of these elements in some species indicates that they were highly active at some point. The work by Pace et al. (12) recently described a currently inactive family of *hAT* transposable elements called *Space Invaders (SPIN)* that has recently (15–40 Mya) successfully invaded a number of tetrapod genomes. More recently, *SPIN* was identified in invertebrates as well (13) and also, reptiles and lizards (14, 15). Two properties of *SPIN* suggest that it might be harnessed as a powerful tool for genome engineering. First, these elements have been able to invade a very broad range of animals, including an insect, a snail, a frog, numerous squamates (lizards and snakes), and several mammalian species (12–15). Second, *SPIN* has achieved enormous copy numbers in some of these species, averaging many thousands in vertebrates (up to ~100,000 copies per haploid genome in the tenrec, an afrotherian mammal) (12), making it the most successful DNA transposon family ever reported. These characteristics suggested that *SPIN* might have an intrinsically high level of activity in a broad range of species, including mammals.

Author contributions: X.L., H.E., R.H.H., N.M., L.Z., C.F., F.D.B., P.W.A., and N.L.C. designed research; X.L., H.E., R.H.H., N.M., N.P., L.Z., and F.D.B. performed research; P.W.A. contributed new reagents/analytic tools; X.L., H.E., R.H.H., N.M., N.P., L.Z., F.D.B., P.W.A., and N.L.C. analyzed data; and X.L., H.E., R.H.H., N.M., L.Z., C.F., F.D.B., P.W.A., and N.L.C. wrote the paper.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession nos. [J5684848](https://doi.org/10.1093/seqs/j5684848)–[J5799249](https://doi.org/10.1093/seqs/j5799249)).

<sup>1</sup>To whom correspondence should be addressed. E-mail: [ncraig@jhmi.edu](mailto:ncraig@jhmi.edu).

See Author Summary on page 1988 (volume 110, number 6).

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1121543109/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1121543109/-DCSupplemental).

Bioinformatic analysis of *hAT* elements leads us to divide the *hAT* superfamily into several major subfamilies (16), two of which are the *Ac* subfamily and the *Buster* subfamily, which are named for transposase-related *Buster* proteins and include the now-inactive *Charlie* elements identified in the human genome in the work by Smit (17). We have recently described two *hAT* elements from insects, *AeBuster1* from the mosquito *Aedes aegypti* and *TcBuster* from the red flour beetle *Tribolium castaneum*, and we have shown that both are active elements using transposition in *Drosophila melanogaster* embryos as an assay (16). *SPIN* elements are also members of the *Buster* subfamily of *hAT* transposons. The sequences of these elements are aligned in Fig. S1.

We show here that a consensus derivative of currently inactive mammalian *SPIN* elements and the currently active insect *TcBuster* are both highly active in human cells. That *SPIN* can be resurrected, as has *Sleeping Beauty*, is significant, because *SPIN*s derive from tetrapods (including some mammalian species) (12), whereas *Sleeping Beauty* was derived from inactive versions of fish transposons (18). *SPIN* elements have now been identified in other insects, squamate reptiles, snails, and planaria (14) and should actually be considered an animal transposon; it is possibly the most successful and widespread DNA transposon.

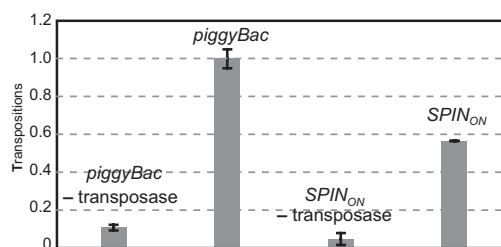
The high activity of *TcBuster* and *SPIN* elements will provide useful tools for mammalian genome engineering. We also show that these *Buster* elements have different target site selection patterns with respect to mammalian genome features than *piggyBac* and *Sleeping Beauty*, another attribute that will contribute to their usefulness for genome engineering.

Finally, we describe both integration and excision assays for *TcBuster* in *Saccharomyces cerevisiae* that will be useful for isolation of hyperactive mutants for genome engineering and structure function analysis of this interesting class of transposable elements. These assays, combined with the activity of these elements in mammalian cells, establish an efficient pipeline for the screening of hyperactive transposase mutants.

## Results

**Resurrection of a *SPIN* Transposon.** A consensus of *SPIN* sequences from mammals, including tenrec, bat, and opossum, and another tetrapod, frog, did not yield an intact ORF, because several nucleotide positions remained unspecified. However, by considering the superconsensus and the DNA sequences of individual *SPIN* elements determined in the work by Pace et al. (12), we generated several intact *SPIN* ORFs (Tables S1 and S2); expression of these ORFs was not toxic in human cells. We assayed integration promoted by these transposases in human cells by cotransfection of a transposon donor plasmid containing a mini *SPIN* element in which 250 bp of the L end and 250 bp of the R end flank a drug resistance gene and a helper plasmid expressing the *SPIN* transposase under control of the CMV promoter. After transfection, cells were grown nonselectively for 2 d, and then, transposon integration was selected for by application of the appropriate antibiotic followed by counting of vital cells stained with methylene blue. Some of the *SPIN* transposases were active, whereas others were inactive (Fig. S2). The *SPIN* derivative *SPIN<sub>ON</sub>* was highly active (Fig. 1), and its integration frequency is about 60% of the highly active *piggyBac* element. The *SPIN<sub>ON</sub>* transposon is an active DNA transposon generated from sequences found in mammals, and therefore, its behavior in mammalian cells is of particular interest, especially given its ability to successfully invade many different organisms.

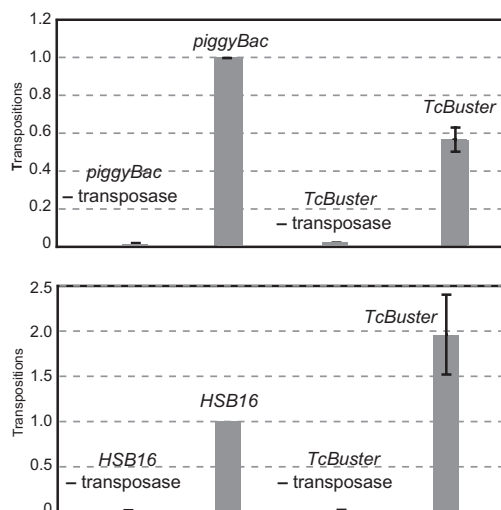
***TcBuster* is Highly Active in Mammalian Cells.** We previously showed that the insect element *TcBuster* is active in *D. melanogaster* and *Ae. aegypti* embryos (16). We now report that it can transpose at relatively high frequency in HeLa cells (Fig. 2, Upper). We assayed chromosomal integration of a mini *TcBuster* element containing 328 bp of the L end and 145 bp of the R end flanking



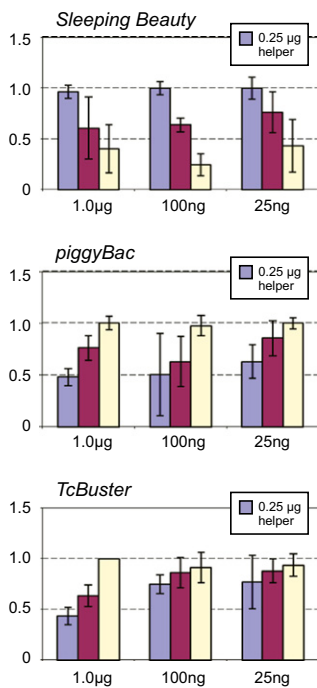
**Fig. 1.** *SPIN<sub>ON</sub>* transposition in mammalian cells. The transposition of *piggyBac* and *SPIN* elements containing blasticidin resistance genes from transfected plasmids to the HeLa genome was measured in the presence and absence of transposase expressed from a CMV promoter by selection for blasticidin-resistant cells. The frequency of *SPIN* transposition compared with the frequency of *piggyBac* transposition is set as 1.0.

a drug resistance gene from a donor plasmid DNA after cotransfection of HeLa cells with a helper plasmid expressing the *TcBuster* transposase, which was not toxic to human cells, using the selection strategy described above for *SPIN<sub>ON</sub>*. Notably, the frequency of *TcBuster* transposition is greater than the frequency of a hyperactive version of *Sleeping Beauty*, *HyperSB16* (*HSB*) (19) (Fig. 2, Lower), and it is ~60% of the frequency of *piggyBac*, which has been found to be highly active in multiple cell types (8). Thus, both *SPIN<sub>ON</sub>* and *TcBuster* are highly active, transposing at frequencies close to the frequency of *piggyBac*.

The transposition frequency of *Sleeping Beauty* in human cells is limited by the phenomenon of overproduction inhibition (that is, transposition frequency actually decreases when transposase expression is increased excessively) (20). To explore overproduction inhibition, we tested *Sleeping Beauty*, *piggyBac*, and *TcBuster* under the same ranges of donor and helper plasmid concentrations and found that, like *piggyBac*, *TcBuster* does not exhibit overproduction inhibition under the conditions tested (Fig. 3). This inhibition increases the use of *TcBuster* as a gene vector, because increased integration frequencies can be achieved simply by increasing the amount of transposase supplied. We also tested whether *TcBuster* transposition could be in-



**Fig. 2.** *TcBuster* transposition in mammalian cells. The transposition of *TcBuster*, hyperactive *Sleeping Beauty*, and *piggyBac* elements containing antibiotic resistance genes from transfected plasmids to the HeLa genome was measured in the presence and absence of transposase by selection for antibiotic-resistant cells; the frequency of transposition of *TcBuster* transposon is compared with the frequency of a hyperactive *Sleeping Beauty* (*HSB16*) (19) and *piggyBac*.

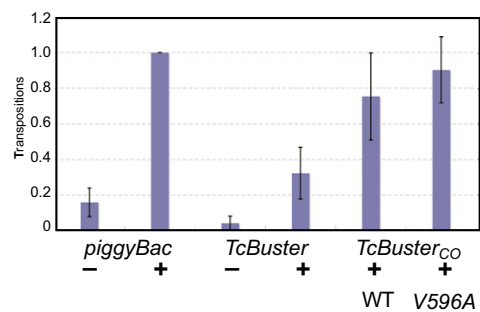


**Fig. 3.** *TcBuster* transposition in mammalian cells. The transposition of *TcBuster*, hyperactive *Sleeping Beauty*, and *piggyBac* elements containing antibiotic resistance genes from transfected plasmids to the HeLa genome was measured by selection for antibiotic-resistant cells; the frequency of transposition of *TcBuster* transposition was compared with the frequency of a hyperactive *Sleeping Beauty* (*HSB16*) and *piggyBac* using different amounts of transposase expression plasmid and the transposon-containing donor plasmid.

created by using a transposase gene codon-optimized mammalian expression, *TcBuster<sub>CO</sub>* (Table S1), compared with the natural insect gene, and we found that transposition in human cells was increased about twofold (Fig. 4). The hyperactive mutant *TcBuster<sub>CO</sub>* V596A is considered in more detail below.

***SPIN<sub>ON</sub>* and *TcBuster<sub>CO</sub>* Prefer Similar Target Sites in Particular Chromosomal Regions.** To characterize *SPIN<sub>ON</sub>* and *TcBuster<sub>CO</sub>* target site selection with respect to multiple genomic features and epigenetic marks, we isolated and sequenced large numbers of de novo transposon-end human genome junctions. After selection for insertions in HeLa cells, we used ligation-mediated PCR to capture transposon–chromosomal junctions from *TcBuster<sub>CO</sub>* and *SPIN<sub>ON</sub>* integrants, sequenced them using the 454 (Roche) platform, and determined their position in the human genome. For comparison, we also analyzed *Sleeping Beauty* and *piggyBac* insertions (Table S3). We analyzed 4,490 *Sleeping Beauty* insertion sites, 13,494 *piggyBac* insertion sites, 6,390 *TcBuster<sub>CO</sub>* insertion sites, and 8,333 *SPIN<sub>ON</sub>* insertion sites (Table S4). The numbers of insertion sites analyzed here are larger than in previous studies of transposon integration in human cells (21–25).

We used these large collections of insertion junctions to characterize the sequences in and immediately around the element target sites duplications more fully. Fig. S3A shows web logos generated by each transposon, and the percentages of insertions at particular kinds of target sites are shown in Fig. S3B; the actual number of insertions at particular sites is shown in Table S4. The target site for *Sleeping Beauty* insertion is TA, and we found that 98.4% of the *Sleeping Beauty* insertions occurred into a TA. Insertions into CA and TG accounted for 0.5% of the insertions. We found a modest preference for sequences outside the target site duplication (A at –4 and T at +4) (Fig. S3A).



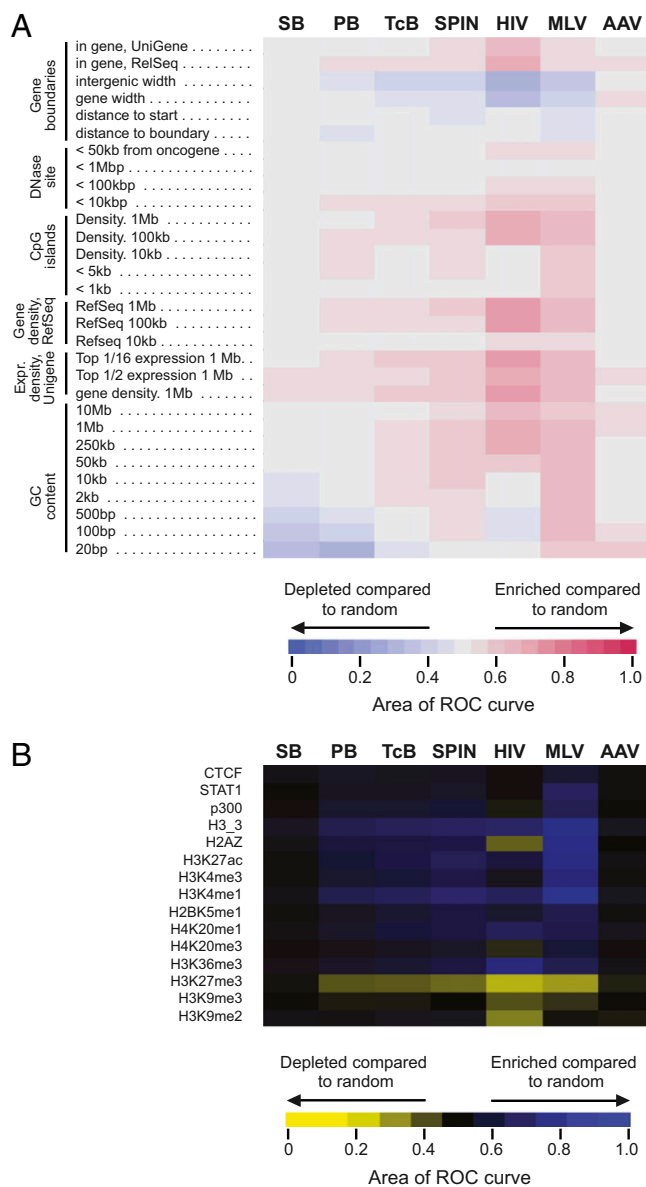
**Fig. 4.** Transposition of *TcBuster* variants. The frequency of transposition of several forms of *TcBuster* is compared with the frequency of *piggyBac*. *TcBuster* is the ORF from *Tribolium*, *TcBuster<sub>CO</sub>* is a version codon-optimized for expression in mammalian cells, and *TcBuster<sub>CO</sub>* V596A contains an amino acid change based on a hyperactive mutant of the closely related *AeBuster1*.

The target site for *piggyBac* is TTAA. We found that 97.6% of the target sequences used were TTAA. The majority of the non-TTAA target sites are pairs of symmetrically related sequences (CTAA and TTAG and ATAA and TTAT) (Fig. S3B) in which the nonconsensus positions are where the 3'OH transposon end attacks the target DNA on insertion (Fig. S3C). There is little local sequence selectivity beyond the TTAA *piggyBac* target site duplication (Fig. S3A).

With the *hAT* *Buster* elements, there is a strong preference for TA in the middle of the 8-bp target site duplication rather than adjacent to the positions of breakage and joining (Fig. S3C) (16); we find here that 93.6% of *TcBuster<sub>CO</sub>* insertions and 95.3% *SPIN* insertions contain this TA. For both elements, TG and CA in the middle of the 8-bp target site duplication are the other major type of insertion site (Fig. S3B). With *SPIN<sub>ON</sub>*, there is also some preference for TC and GA flanking this central TA; this preference is much less obvious with *TcBuster<sub>CO</sub>*. With both *SPIN<sub>ON</sub>* and *TcBuster<sub>CO</sub>*, there are modest preferences for particular sequences flanking the target site duplication, the most significant being a preference for A at –7 and T at +7 (Fig. S3A).

We have also analyzed the genome-wide insertion distributions of *TcBuster<sub>CO</sub>*, *SPIN<sub>ON</sub>*, *Sleeping Beauty*, and *piggyBac* insertions, comparing them with genomic features such as annotated transcription units and epigenetic marks (Fig. 5). For comparison, we also generated in silico a set of matched random control sites, which are random positions in the genome but matched to integration sites for the proximity to restriction enzyme recognition sites used in recovery of the experimental sites. Use of the matching procedure reduces biases in the analysis arising because of use of restriction enzymes cleavage of genomic DNA in the recovery process. Favoring or disfavoring of transposon integration relative to genomic features is summarized in heat maps using the receiver operator characteristic (ROC) area method for comparison (Fig. 5) (26). For comparison, these heat maps also summarize data on previously obtained patterns for HIV, MLV, and adeno-associated virus (AAV) insertion in HeLa cells (27–29).

The most striking overall feature of the *TcBuster<sub>CO</sub>*, *SPIN<sub>ON</sub>*, *Sleeping Beauty*, and *piggyBac* insertion profiles is that they show little insertion bias compared with HIV and MLV. In the human genome, many types of features are correlated with each other (for example, transcription units are located in G/C-rich regions that are also rich in CpG islands and favored sites for cleavage by DNase I). *TcBuster<sub>CO</sub>*, *SPIN<sub>ON</sub>*, and *piggyBac* all showed weak favoring of integration in genomic regions enriched in transcription units, CpG islands, and preferential cleavage sites for DNase I. *Sleeping Beauty* showed a different pattern, with weaker favoring of integration near these features. All four transposons showed favored integration near genes that were active in HeLa cells based on comparison with transcriptional profiling data for HeLa



**Fig. 5.** Distribution of *Sleeping Beauty*, *piggyBac*, *TcBuster<sub>CO</sub>*, and *SPIN<sub>ON</sub>* insertions on the human genome with respect to the indicated genomic features. (A) Integration frequency near selected genomic features. (B) Integration frequency near bound proteins and modified histones mapped using the ChIP-Seq method. In both A and B, integration site datasets for *Sleeping Beauty* (SB), *piggyBac* (PB), *TcBuster<sub>CO</sub>* (TCB), and *SPIN<sub>ON</sub>* (SPIN) 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 ROC area method described in the work by Berry et al. (26). In A, blue indicates that insertions are depleted compared with random. Red indicates features where insertions are enriched compared with random. Gray indicates that the distribution is random. In B, yellow and blue are used to indicate depletion or enrichment. A detailed explanation of the variables studied can be found in the work by Ocwieja et al. (58) or at [http://microb230.med.upenn.edu/assets/doc/HeatMapGuide\\_v12\\_formatted.doc](http://microb230.med.upenn.edu/assets/doc/HeatMapGuide_v12_formatted.doc).

cells (Fig. 5A, labeled Expr. Density), a pattern that is quite pronounced for HIV (30).

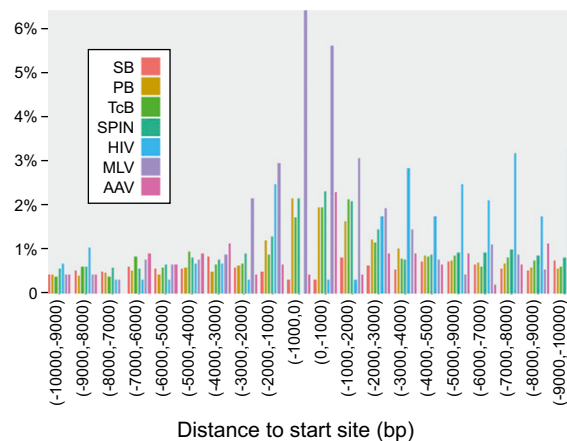
*piggyBac* and *Sleeping Beauty* favored integration in short chromosomal regions rich in A/T (Fig. 5A). This trend was most notable at short interval sizes (20 bp surrounding the target site)

and for the *piggyBac* transposon, which targets TTAA at the point of joining, potentially explaining the bias. *Sleeping Beauty* also showed a strong target sequence preference for integration in A/T-rich regions, but the interval size was larger, suggesting possible contribution of additional factors such as nucleosome wrapping (31).

We have also analyzed the distribution of *hAT* Baster elements, *Sleeping Beauty* and *piggyBac* insertions with respect to epigenetic marks such as histone modification, and the binding of certain regulatory proteins mapped previously in HeLa cells (32) (Fig. 5B). *Sleeping Beauty* displays no significant difference from random insertion for the large number of epigenetic features that we have evaluated, resembling the pattern for AAV but differing from the other transposons, HIV, and MLV. *piggyBac*, *TcBuster<sub>CO</sub>*, and *SPIN<sub>ON</sub>* showed preferences for integration near chromatin marks characteristic of active transcription units (e.g., H3K27 acetylation and H3K4 monomethylation) and disfavored integration near a mark characteristic of inactive chromatin (H3K27 trimethylation).

MLV strongly favors integration near gene 5' ends (28) (Fig. 5A, distance to start; note that the tile is blue on the heat map, because favored distances are shorter than random), and the heat maps in Fig. 5A suggested that this finding is true of *TcBuster<sub>CO</sub>*, *SPIN<sub>ON</sub>*, and *piggyBac* as well. To investigate, integration site frequency was plotted surrounding transcription start sites (Fig. 6). *TcBuster<sub>CO</sub>*, *SPIN<sub>ON</sub>*, and *piggyBac* each showed increased transposition frequency near transcription start sites, and this frequency achieved significance ( $P = 0.007$  and  $P = 0.0003$ , respectively, for comparison with matched random controls) for *SPIN<sub>ON</sub>* and *piggyBac*, although the magnitude of the effect was lower than for MLV. HIV disfavored integration at transcription start sites, which has been reported previously, and unexpectedly, *Sleeping Beauty* also did.

***TcBuster<sub>CO</sub>* Can Transpose in *S. cerevisiae*.** We, and others, have shown that *Hermes* (33) and *Ac* (34), another *hAT* *Ac* subfamily element, can transpose in the yeast *S. cerevisiae*. We have found that *TcBuster<sub>CO</sub>* can also both excise from a donor site and integrate into a new target site in yeast. Expression of *TcBuster<sub>CO</sub>* in yeast is not toxic but we were unable, however, to establish a *SPIN<sub>ON</sub>* system in yeast because of lethality in the presence of *SPIN<sub>ON</sub>* transposase and a *SPIN* transposon (see below).



**Fig. 6.** Distribution of *Sleeping Beauty*, *piggyBac*, *TcBuster<sub>CO</sub>*, and *SPIN<sub>ON</sub>* insertions on the human genome with respect to transcription start sites. Integration sites near transcription start sites were compiled onto a common transcription start site, and the proportions were mapped. The x axis shows the distance from the transcription start site, and the y axis shows the percentage of integration sites in each interval.

To analyze *TcBuster<sub>CO</sub>* integration, we used a two-plasmid system (Fig. S4A). One plasmid expressed the *TcBuster<sub>CO</sub>* transposase under the control of the *GALS* promoter. The other donor plasmid contained a mini *TcBuster*-ClonNAT element containing 328 bp of the L end and 145 bp of the R end flanking a NatMx cassette that provides resistance to ClonNAT and a separate WT *URA3* gene on the plasmid backbone. Plasmid-free cells can be selected for by treatment with fluoroortic acid (5-FOA), because a toxic metabolite is produced in *URA3* cells. Thus, chromosomal integration of the mini *TcBuster<sub>CO</sub>* element can be assayed by selection for 5-FOA<sup>R</sup> plasmid-free cells that are also ClonNAT<sup>R</sup>. In particular, we measured the number of 5-FOA<sup>R</sup> ClonNAT<sup>R</sup> and total cells in colonies grown in the presence or absence of galactose, which induces expression of the transposase. After growth on medium containing galactose to induce transposase expression, we observe a mini *TcBuster<sub>CO</sub>* integration frequency of about  $5 \times 10^{-4}$  integrations/cell compared with background frequencies that are about 100-fold lower (i.e.,  $6.9 \times 10^{-6}$  in the presence of glucose and  $1.3 \times 10^{-6}$  in the absence of transposase) (Fig. 7). We evaluated the chromosomal location of 12 insertions by Southern blotting and sequencing, finding only a single insertion per cell. All were in different chromosomal positions, and 12 of 12 integrants had the *TcBuster* consensus nnnTAnnn target site duplication (Table S5).

To measure *TcBuster<sub>CO</sub>* excision, we used another two-plasmid system (Fig. S4B). Again, one plasmid expressed the *TcBuster<sub>CO</sub>* transposase under the control of the *GALS* promoter. The other donor plasmid contained a mini *TcBuster* element inside a *URA3* derivative such that reversion from Ura<sup>-</sup> to Ura<sup>+</sup> can be used to

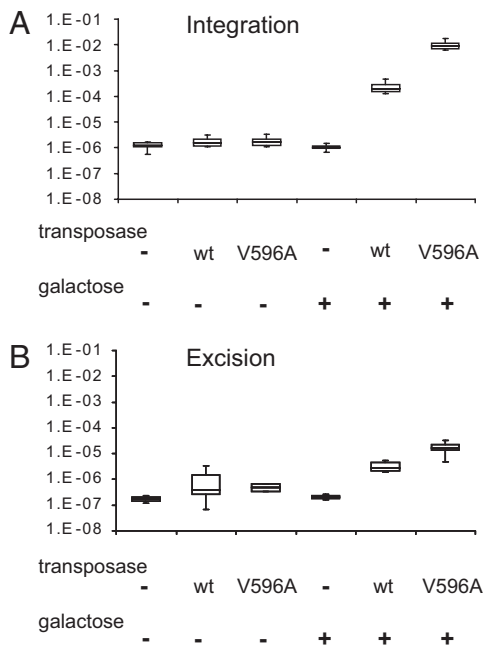
measure transposon excision as we have previously described with *piggyBac* (Fig. S4B) (35). The plasmid *URA3* gene contained the yeast actin intron. The actin intron can be readily spliced from the *URA3*::actin intron mRNA such that cells that are deleted for chromosomal *URA3* but contain *URA3*::actin intron do not require supplementation with uracil for growth (36). When a several-kilobase mini *TcBuster* transposon containing 328 bp of the L end and 145 bp of the R end flanking a ClonNAT<sup>R</sup> cassette is introduced into the actin intron of *URA3*::actin intron, however, the large actin mini *TcBuster* intron cannot be spliced from the mRNA, and cell growth requires supplementation with uracil. Excision of the transposon restores the small actin intron such that cells do not require supplementation with uracil. Thus, measuring the number of Ura<sup>+</sup> cells compared with the number of total cells is a measure of transposon excision. We measured Ura<sup>+</sup> and total cells in colonies grown in the presence or absence of galactose, which induces expression of the transposase. After growth on galactose, the frequency of *TcBuster<sub>CO</sub>* Ura<sup>+</sup> revertants is about  $2.7 \times 10^{-6}$ , significantly higher than the background level observed after growth in glucose ( $2.6 \times 10^{-7}$ ) or in the absence of transposase ( $1.2 \times 10^{-7}$ ) (Fig. 7B).

Notably, the observed frequency of *TcBuster<sub>CO</sub>* integration of  $5 \times 10^{-4}$  is about 200-fold higher than the observed frequency of excision ( $2.7 \times 10^{-6}$ ), likely because of inefficient repair of the hairpins formed in the flanking donor DNA on element excision (34, 37, 38).

We also analyzed the donor site sequence in multiple Ura<sup>+</sup> revertants. In the actin intron, the mini *TcBuster* element is flanked by imperfect target site duplications (CTTTAGGC and CTTTATAC derived from the *T. castaneum* genome), which are, in turn, flanked by XhoI sites derived from the actin intron. DNA sequence analysis of the donor site in Ura<sup>+</sup> revertants revealed that, in 16 of 20 cases, a single XhoI site was present, reconstituting the actin intron sequence. This repair could result from either homology-dependent repair using the chromosomal actin intron in *ACT1* as a template or end joining after resection of the flanking *T. castaneum* genome donor DNA. In the other four cases, a portion of the L-terminal inverted repeat was still present; this finding could reflect incomplete repair or illegitimate recombination that deleted most of the element, which we have seen in other assays lacking a *hAT* transposase.

We also asked if *SPIN<sub>ON</sub>* could transpose in yeast using both integration and excision assays but found that no cells survived galactose induction in the presence of both the transposase and the mini *SPIN* transposon element. Cell growth was not affected by the expression of *SPIN<sub>ON</sub>* transposase alone, indicating that lethality is directly related to transposition, perhaps from failure to repair the gaps that flank the newly inserted element or repair the gapped donor site. *SPIN<sub>ON</sub>* induction was also lethal in cells containing a mini *TcBuster* transposon; the sequence does have some similarity to the *SPIN* ends, but no lethality or transposition was observed with *TcBuster<sub>CO</sub>* transposase and a mini *SPIN<sub>ON</sub>* element.

***TcBuster<sub>CO</sub>* Transposase Mutation Results in Increased Transposition in Yeast and Mammalian Cells.** A yeast assay system for transposition is particularly useful, because it provides powerful genetic tools to look for interesting transposase mutants (for example, hyperactive transposase mutants). In other work, we have used the yeast *URA3*::actin intron excision assay to isolate hyperactive mutants of *TcBuster*'s close relative *AeBuster1*, finding that *AeBuster1* V597A increases *AeBuster1* integration about 200-fold and excision about 30-fold. *TcBuster<sub>CO</sub>* V596 is equivalent to *AeBuster1* V597, and we have examined the effect of *TcBuster<sub>CO</sub>* V596A on transposition in both yeast and mammalian cells. In yeast, *TcBuster<sub>CO</sub>* V596A increases integration about 50-fold and excision about fivefold (Fig. 7). We also observe increased integration with *TcBuster<sub>CO</sub>* V596A in mamma-



**Fig. 7.** *TcBuster<sub>CO</sub>* transposition in yeast. (A) Integration assays that measure the transposition of a *TcBuster* transposon from a plasmid into the host genome by assaying plasmid-free cells for the presence of a transposon-encoded marker are shown. The transposase is expressed from a plasmid *GALS* promoter and thus, is induced in the presence of galactose. *TcBuster<sub>CO</sub>* V596A has an amino acid change comparable with the hyperactive *AeBuster1* V597A mutant. (B) Excision assays that measure the excision of a *TcBuster* transposon from a donor site in a plasmid *URA3* gene by following reversion from Ura<sup>-</sup> to Ura<sup>+</sup> are shown. The transposase is expressed from a plasmid *GALS* promoter and thus, is induced in the presence of galactose. *TcBuster<sub>CO</sub>* V596A has an amino acid change comparable with the hyperactive *AeBuster1* V597A mutant.

lian cells (Fig. 4). Thus, isolation of hyperactive transposases in yeast can provide improved tools for genome engineering in mammalian cells.

We sequenced 12 insertions in the *S. cerevisiae* genome generated with *TcBuster<sub>CO</sub>* V596A, and we found that all inserted into different sites and made 8-bp target site duplications (Table S5).

**Purified *TcBuster<sub>CO</sub>* and *SPIN<sub>ON</sub>* Transposases Can Promote Transposition *In Vitro*.** We have also examined the activity of purified *TcBuster<sub>CO</sub>* and *SPIN<sub>ON</sub>* transposases *in vitro*. We tagged the *SPIN<sub>ON</sub>* and *TcBuster<sub>CO</sub>* ORFs with affinity tags, expressed them in *S. cerevisiae*, and purified them by affinity chromatography, finding that both are active *in vitro* in the absence of other proteins, which may account for their apparently high activity in multiple organisms.

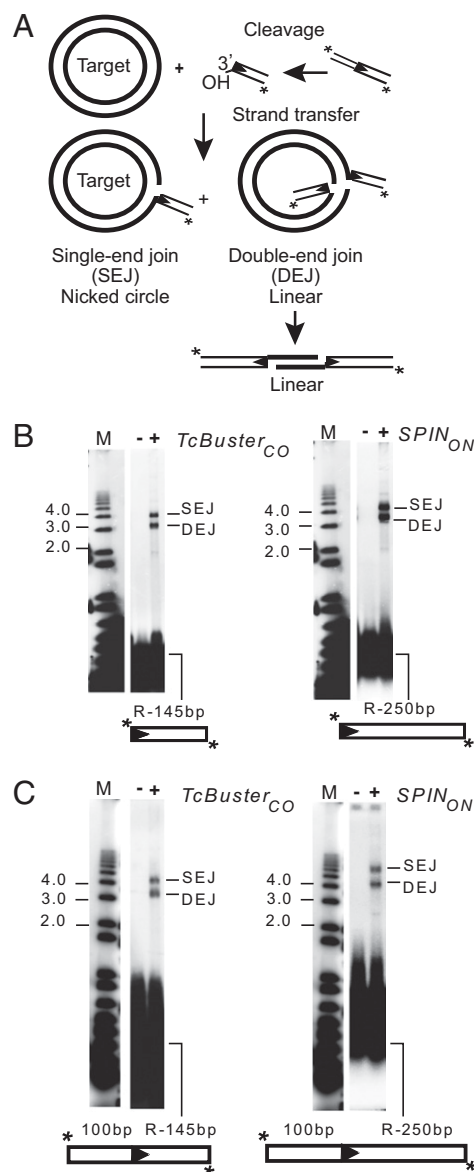
We used two assays to evaluate transposition *in vitro* (Fig. 8A). In one assay, we used a precleaved transposon-end oligonucleotide in which the 3'OH transposon end was already exposed to evaluate the target joining. In the other assay, we used a transposon-end oligonucleotide in which the end was flanked by donor sequences, allowing us to characterize both donor excision and target joining. Two products are observed in these reactions: one product is a nicked circle form of the target plasmid DNA in which a single transposon-end oligonucleotide is covalently linked to one strand of the target DNA, generating a nicked circle species called a single end join (SEJ), and the other product is a linear plasmid form in which two transposon-end oligonucleotides were joined to the target DNA at the same position to form a double end join (DEJ). In the presence of  $Mn^{2+}$ , both purified *TcBuster<sub>CO</sub>* and *SPIN<sub>ON</sub>* transposases can promote target joining using a precleaved substrate (Fig. 8B) and can also cleave flanking donor DNA from the transposon end, which can then join to target DNA (Fig. 8C). *TcBuster<sub>CO</sub>* can also use  $Mg^{2+}$  as a metal cofactor (Fig. S5).

***Buster* Transposons Excise from the Donor Through a Hairpin Intermediate, and 3'OH Transposon Ends Join to the Target DNA.** We have previously shown (39) that the *hAT* element *Hermes* excises from the donor DNA through a hairpin reaction that begins with a nick 1 nt into the donor DNA adjacent to the 5' transposon end, exposing a 3'OH in the top strand of the donor DNA. This 3'OH then attacks its complementary strand, generating a hairpin on the flanking donor DNA and releasing the transposon end with an exposed 3'OH. This 3'OH then directly attacks the target DNA as in other reactions mediated by RNase H-family transposases (40, 41).

To evaluate the *Buster* element double-strand cleavage reaction that disconnects the transposon end from the flanking donor DNA and target joining, we incubated *TcBuster<sub>CO</sub>* transposase with an oligonucleotide substrate in which a transposon end was flanked by nontransposon sequences and examined recombination over time using both  $Mg^{2+}$  and  $Mn^{2+}$  as cofactors (Fig. 9). When the reactions were displayed on a native agarose gel, we observe increasing amounts of both SEJ and DEJ products over time with both cofactors. We also analyzed these reactions by displaying them on denaturing gels, and we observed the formation of a 64-nt transposon top strand resulting from the initial nicking reaction and a 219-bp nucleotide hairpin species (Fig. 9A).

To verify that the terminal 3'OH end of the *Buster* transposons was covalently linked to the substrate DNA, we used *TcBuster<sub>CO</sub>* transposase and a precleaved transposon-end oligonucleotide that was labeled at the 5' end of the strand containing the terminal 3'OH as a substrate; we displayed the reaction products on a denaturing agarose gel, observing a single species the length of the plasmid plus the length of the transposon, consistent with 3' transposon-end joining to a single target strand (Fig. 9C).

Thus, *TcBuster<sub>CO</sub>* transposase from the *Buster* subfamily of transposons breaks and joins DNA as does *Hermes*, the *Ac*

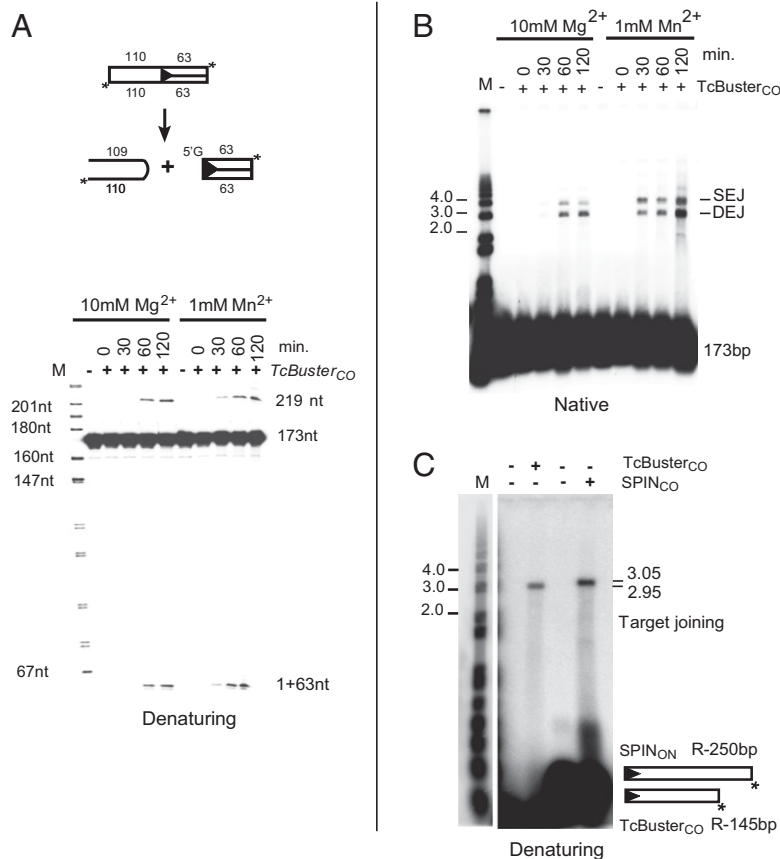


**Fig. 8.** *hAT* *Buster* element transposition *in vitro*. The products of *in vitro* transposition reactions using end-labeled oligonucleotides and circular plasmid targets are displayed on agarose gels. (A) Substrates, steps, and products of transposition. In SEJs, a single transposon end oligonucleotide joins to the target plasmid, giving a nicked circle; in DEJs, two transposon end oligonucleotides have joined to the target plasmid, giving a linear species. (B) Target joining and coupled cleavage using *TcBuster<sub>CO</sub>* transposase and end oligonucleotides as indicated are shown. (C) Target joining and coupled cleavage *SPIN<sub>ON</sub>* and end oligonucleotides as indicated are shown.

subfamily *hAT* transposase, using the same sorts of phosphoryl transfer reactions as other RNase H-family transposases and required no additional proteins as obligate cofactors.

## Discussion

The experiments described here provide detailed analysis of two members of the newly bioinformatically defined *Buster* subfamily of *hAT* transposable elements (16) (Fig. S1). One of these elements, *TcBuster* from the red flour beetle *T. castaneum*, is likely currently active in this host; we show here that it is active in mammalian cells, yeast, and *in vitro*, and we have shown previously that it can be active in *Drosophila* (16). By contrast, there are no known currently active *SPIN* elements; most of the cur-



**Fig. 9.** Mechanism of *Buster* element transposition. (A) *TcBuster* transposition occurs through a hairpin intermediate. The products of transposition reactions over time use *TcBuster*<sub>CO</sub> transposase, a 3' end-labeled oligo in which the transposon end is flanked by donor DNA and divalent metal are displayed (A) on a denaturing acrylamide gel, (B) a native agarose gel, and (C) a denaturing agarose gel. SEJ and DEJ products as described in Fig. 8 are observed on the native gel. On the denaturing gel, a 64-bp fragment containing the top strand of the transposon (63 nt) and 1 nt from the flanking donor DNA (reflecting that the initiating nick occurs 1 nt from the 5' transposon end into the flanking donor DNA) and flanking donor 219-nt DNA hairpin species containing 110 nt of the bottom strand and 109 nt of top strand increase over time. (B) Coupled cleavage and strand transfer. The products of the reactions in A are displayed on a native agarose gel. (C) The 3' end of the transposon is attached covalently to the target DNA. The products of reactions using either *TcBuster*<sub>CO</sub> or *SPIN*<sub>ON</sub> transposases as indicated and their cognate transposon ends labeled on the 5' end of the bottom strand are displayed on a denaturing agarose gel. The covalent linkage of the labeled transposon end and target DNA in the reaction product shows that the 3'OH transposon end joined to the target DNA.

rently identified elements contain stop codons in the transposase ORF. *SPIN* elements were identified in the work by Pace et al. (12) in a number of tetrapod genomes, including those genomes of several mammals and more recently, in two invertebrates (13, 15). These elements were highly active in these genomes ~20–40 Mya, amplifying to high copy number. Intriguingly, the patchy phylogenetic distribution of closely related *SPIN* elements, combined with their extreme level of sequence similarity, suggests that they entered these different hosts by horizontal transmission. We have made an active *SPIN* element that we have called *SPIN*<sub>ON</sub> by synthesizing a consensus version.

We have shown that both *TcBuster* and *SPIN*<sub>ON</sub> are highly active in mammalian cells, and we have compared them to two other transposons that are widely used in mammalian cell genome engineering: *Sleeping Beauty*, which has been resurrected from bony fish, and *piggyBac*, which was discovered in a baculovirus of the cabbage looper moth (18, 42). Both *SPIN*<sub>ON</sub> and *TcBuster*<sub>CO</sub> have transposition frequencies in HeLa cells higher than the hyperactive *HSB16* mutant of the *Sleeping Beauty* transposase (19), and they approach the activity of the *piggyBac* transposase in these cells (Figs. 1–4). We did not see signs of overproduction inhibition for the *TcBuster* transposase or *piggyBac* transposase under our experimental conditions. Future work will be required to compare *TcBuster* and *SPIN*<sub>ON</sub> with the 100×

*Sleeping Beauty* transposase and the hyperactive *piggyBac* transposase in different cell types (43).

**Target Site Selection at the Nucleotide Level.** We have analyzed *hAT* *Buster* element target site selection in mammalian cells and for comparison, target selection by *Sleeping Beauty* and *piggyBac*. For all elements, we used ligation-mediated PCR to recover large numbers of de novo mammalian insertion sites and analyzed them by next generation sequencing; the number of target sites ranged from 4,500 to 13,500 among the elements, which is larger than other transposon datasets from mammalian cells studied to date.

A previous study indicated that nnnTAnnn is the preferred target site of *Buster* elements (16). We have found that the requirement for the central TA dinucleotide is, in fact, quite stringent, being present in 94.6% of the 6,390 *TcBuster* insertion sites and 95.3% of 8,333 *SPIN*<sub>ON</sub> mammalian insertions (Fig. S3 and Table S4). The preference for this *Buster* target feature is very similar to the TA (98.4%) and TTAA (97.6%) target site requirements that we have found for the *Sleeping Beauty* and *piggyBac* elements, respectively (44, 45). The *Buster* target site duplication TAs are in the center of the 8-bp target site duplication (nnnTAnnn), whereas the *Mariner/Tc1* TA (*Sleeping Beauty*) and *piggyBac* TTAA comprise the entire target site duplication. We know of no other element having such a stringent

requirement for a particular sequence in an internal portion of the target site duplication (Fig. S3). The central TA is a defining feature of elements of the *Buster* subfamily of elements; most members of the *hAT Ac* subfamily prefer to insert into nTnnnnAn sites (16). The stringency of the *Buster* elements for nnnTAnnn provides an excellent system to study the mechanism of target site selection at the molecular level. Our other work on the *Ac* subfamily element *Hermes*, including the establishment of an in vitro transposition system (39) and analysis of *Hermes* structure at the crystallographic level (41), will provide useful complementary tools for future mechanistic studies of the *hAT Buster* elements.

Our analysis of large numbers of de novo insertion sites for *piggyBac*, *TcBuster*, and *SPIN* also revealed preferred secondary insertion sites. With *piggyBac*, the most common secondary sites contained a C or A at the position of transposon-end joining as opposed to the standard T in a TTAA target site. With the *Buster* elements, secondary sites had TG or CA changes at the central TA of the nnnTAnnn target site duplication rather than changes at the positions of end joining (Fig. S3 and Table S4). It will be interesting to analyze the effects of these changes on transposon excision.

**Target Site Selection at the Genome-Wide Level.** We have also analyzed the pattern of *Sleeping Beauty*, *piggyBac*, *TcBuster<sub>CO</sub>*, and *SPIN<sub>ON</sub>* insertions with respect to genomic-wide features, such as genes and gene regulatory sites, and epigenetic features, such as posttranslationally modified histones (Fig. 5). Three of the transposons show a weak favoring of integration in gene-rich, transcriptionally active regions (*piggyBac*, *TcBuster<sub>CO</sub>*, and *SPIN<sub>ON</sub>*). Genomic features associated with active genes, and gene-rich regions were also associated, including CpG islands, DNase I cleavage sites, histone acetylation sites, and activating histone methylation marks. The H3K27 trimethylation mark, which is negatively associated with transcription, was negatively associated with integration of these three transposons. *Sleeping Beauty* was the most divergent, showing little preference for or against integration near these genomic features, which was seen in a previous study (26). *Sleeping Beauty* most resembled AAV, which becomes integrated at DNA double-strand breaks (46). Evidently, the placement of spontaneous DNA double-strand breaks is relatively insensitive to the chromosomal features studied here, although why *Sleeping Beauty* integration should show a similar distribution is an interesting topic for additional study.

The three transposons that favored integration in gene-rich regions (*piggyBac*, *TcBuster<sub>CO</sub>*, and *SPIN<sub>ON</sub>*) also tended to favor integration near gene 5' ends, paralleling MLV. *Sleeping Beauty*, in contrast, resembled HIV in disfavoring integration near transcription start sites. For HIV, integration targeting is known to be largely due to tethering by binding to the cellular LEDGF/p75 protein (38, 47, 48), and therefore, the lack of integration near transcription start sites may be caused by the lack of the tethering factor. Whether MLV binds a positively acting tethering factor at transcription start sites is unknown but is, at present, a leading model. It may be that *piggyBac*, *TcBuster<sub>CO</sub>*, and *SPIN<sub>ON</sub>* bind to cellular factors at transcription start sites that promote nearby integration. Our previous studies of the distribution of *Hermes* insertions in the *S. cerevisiae* genome suggest that a major mechanism for targeting these insertions to transcription start sites is the lack of nucleosome binding in these regions (33).

***TcBuster<sub>CO</sub>* and *SPIN<sub>ON</sub>* Are Promising Tools for Genome Engineering.** In addition to being interesting in their own right because of their important roles in genome structure and function and as elaborate protein–DNA machines, DNA transposons have been used in many organisms as important tools for genome engineering, including insertional mutagenesis and transgenesis. Notably, they have been used for the identification of genes involved in human disease and in the treatment of some of these genetic diseases through gene

therapy (5, 6, 49). For example, their ability to move within genomes and act as mutagens in model systems has been elegantly used to identify genes potentially involved in oncogenesis in mice, leading to the identification of their human orthologs as possible candidates for additional investigation (50–52). The two transposons used for these studies were *Sleeping Beauty* and *piggyBac*, each being members of different superfamilies of transposons and having different target site preferences. Their ability to carry genes into genomes has led to the development of these transposons and the *Tol2* element from the medaka fish as gene vectors for use in human cells (4, 53). *Tol2* is a member of the *hAT* superfamily and therefore, has a target site preference different from the preferences of *piggyBac* and *Sleeping Beauty* (54). Having transposons that have different target site preferences and differ in their mechanisms of excision and integration during transposition is important to provide broad genome coverage in mutagenesis studies, and it also allows the possibility of introducing different transposons with different cargoes into genomes, which may be especially useful with human genome engineering, gene vectors for use in human gene therapy, regenerative medicine, and rodent genetics.

The development of the *TcBuster* yeast system provides a powerful means not only of studying transposition (for example, identifying host proteins involved in transposition by analyzing transposition in the yeast deletion strain collection) but importantly, for isolating hyperactive versions of the *TcBuster* transposase that will be hyperactive in mammalian cells. We have already used a similar yeast excision assay to identify hyperactive *piggyBacs* that are also hyperactive in mammalian cells (21) and isolate hyperactives of *AeBuster1*, which is closely related to *TcBuster*. We have shown here that changes in similar amino acids in *TcBuster<sub>CO</sub>* can lead to increased transposition. A hyperactive 100× *Sleeping Beauty* mutant (55, 56) has been successfully isolated by site-directed mutagenesis, and is hyperactive in HeLa cells as well as murine embryonic and adult stem cells. Such hyperactive transposases facilitate transposon use as vectors in human gene therapy, regenerative medicine, and rodent genetics (44, 55). As a consequence, the use of these *hAT* transposons should accelerate the ability to find genes involved in disease in rodent models and therefore, identify the relevant human orthologs.

The generation and testing of mutants of at least *TcBuster* will be accelerated by our ability to determine their function both in vitro and in vivo in yeast before placing into mammalian systems. This pipeline, which should be able to be extended to the *SPIN<sub>ON</sub>* transposases by making replacements, should generate a suite of transposases that will further expand the growing toolkit of transposons that can be used in mammalian systems.

## Methods

**Transposition in Mammalian Cells.** Transposition was measured in HeLa cells by cotransfecting a donor plasmid containing mini transposon-containing segments from the ends of the element flanking an antibiotic resistance gene or other marker and a helper plasmid expressing a transposase under a human CMV promoter, which was followed by selection for antibiotic-resistant cells.

**Integration Site Recovery, 454 Sequencing, and Analysis.** Mammalian integration sites were recovered as described (31). Briefly, genomic DNA was extracted from an integration transfection library using the DNeasy tissue kit (Qiagen); 2 μg genomic DNA were digested overnight with *ApoI* or *BstYI*, and then ligated to linkers overnight at 16 °C. Nested PCR was then carried out under stringent conditions using transposon end-specific primers complementary to transposon sequences and linker-specific primers complementary to the DNA linker. DNA barcodes were included in the second-round PCR primers to track sample origin. The PCR products were gel-purified, pooled, and sequenced using 454 sequencing platform. Only sequences that uniquely aligned to the human genome by BLAT (hg18, version 36.1; >98% match score) and began within 3 bp of the LTR end were used in downstream analyses.



### Bioinformatic Analysis of Target Site Duplications and Other Genomic Features.

Detailed bioinformatic methods for analysis of association with chromosomal features are described in the work by Berry et al. (26). The methods for generating heat maps based in receiver operating characteristic curves (ROC) are as described in the work by Berry et al. (26). A detailed description of the methods used to generate the genomic features heat map in Fig. 5A can be found at [http://microb230.med.upenn.edu/assets/doc/HeatMapGuide\\_v12\\_formatted.doc](http://microb230.med.upenn.edu/assets/doc/HeatMapGuide_v12_formatted.doc).

**TcBuster Integration and Excision Assays in *S. cerevisiae*.** Integration assays were performed in BY4727 (*MAT alpha his3Δ200 leu2Δ0 lys2Δ0 met150 trp1Δ63 ura3Δ0*) (57) after transformation with the Trp<sup>+</sup> pGALS *TcBuster* transposase helper plasmid and the pRS416 *URA3* mini *TcBuster*-ClonNAT donor plasmid. Integration was measured by selection for 5-FOA-resistant plasmid-free cells containing the ClonNAT-resistant mini *TcBuster* element. Excision assays were performed in BY4727 transformed with the TRP<sup>+</sup> pGALS *TcBuster<sub>CO</sub>* transposase helper plasmid and the HIS<sup>+</sup> *URA3::actin intron::TcBuster*-ClonNat excision donor plasmid. Excision was measured by selection for cells that reverted from Ura<sup>-</sup> to Ura<sup>+</sup>.

***TcBuster<sub>CO</sub>* and *SPIN<sub>ON</sub>* Transposase Expression and Purification from Yeast.** Strep-HA tags (IBA) were added to the N termini of the *TcBuster<sub>CO</sub>* and

*SPIN<sub>ON</sub>* transposases, allowing for purification from yeast extracts by affinity chromatography on a Strep Tactin (IBA) column. More than 60% of transposase protein was recovered and 90% pure as judged from protein gels.

**In Vitro Transposition Assays.** One hundred fifty nanomolar transposase was incubated with 1.5 nM radiolabeled transposon end and 10 nM pUC19 plasmid as the target DNA in 25 mM MOPS, pH 7.0, 25 mM Tris-HCl, pH 8.0, 37.5 mM NaCl, 1 mM MnCl<sub>2</sub>, 5 mM DTT, 20% (vol/vol) glycerol or DMSO, and 100 μg/mL BSA in a final volume of 20 μL at 37 °C in 1 h. Transposon end segments were either flanked by donor DNA or had exposed 3'OH transposon termini; end fragments were generated by PCR.

**ACKNOWLEDGMENTS.** We thank Dr. Sue Brown (Kansas State University, Manhattan, KS), Dr. David O'Brochta (University of Maryland, College Park, MD), and Dr. Bradley Fletcher (University of Florida, Gainesville, FL) for providing *T. castaneum* genomic DNA, *piggyBac* plasmids, and a *Sleeping Beauty* plasmid, respectively. We thank Kefang Xie for identifying the hyperactive *AeBuster1* hyperactive mutant in a screen in yeast. We also thank Patti Kodeck for her assistance with the manuscript and Helen McComas for her assistance with the figures. This work was supported by National Institutes of Health Grants GM077582 (to C.F.), AI52845 (to F.D.B.), AI45741 (to P.W.A. and N.L.C.), and GM076425 (to N.L.C.). The work was also supported by a grant from the Maryland Stem Cell Research Fund (to N.L.C.), and N.L.C. is an Investigator of the Howard Hughes Medical Institute.

1. Biémont C (2010) A brief history of the status of transposable elements: From junk DNA to major players in evolution. *Genetics* 186:1085–1093.
2. Lander ES, et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921.
3. Schnable PS, et al. (2009) The B73 maize genome: Complexity, diversity, and dynamics. *Science* 326:1112–1115.
4. Izsvák Z, Hackett PB, Cooper LJ, Ivics Z (2010) Translating Sleeping Beauty transposition into cellular therapies: Victories and challenges. *Bioessays* 32:756–767.
5. Claeys Bouuaert C, Chalmers RM (2010) Gene therapy vectors: The prospects and potentials of the cut-and-paste transposons. *Genetica* 138:473–484.
6. Ivics Z, Izsvák Z (2010) The expanding universe of transposon technologies for gene and cell engineering. *Mob DNA* 1:25.
7. Kawakami K (2007) Tol2: A versatile gene transfer vector in vertebrates. *Genome Biol* 8(Suppl 1):S7.
8. Kim A, Pyykko I (2011) Size matters: Versatile use of PiggyBac transposons as a genetic manipulation tool. *Mol Cell Biochem* 354:301–309.
9. Ochman H, Lawrence JG, Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* 405:299–304.
10. Gogarten JPT, Townsend JP (2005) Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol* 3:679–687.
11. Schaaek S, Gilbert C, Feschotte C (2010) Promiscuous DNA: Horizontal transfer of transposable elements and why it matters for eukaryotic evolution. *Trends Ecol Evol* 25:537–546.
12. Pace JK, 2nd, Gilbert C, Clark MS, Feschotte C (2008) Repeated horizontal transfer of a DNA transposon in mammals and other tetrapods. *Proc Natl Acad Sci USA* 105:17023–17028.
13. Gilbert C, Schaaek S, Pace JK 2nd, Brindley PJ, Feschotte C (2010) A role for host-parasite interactions in the horizontal transfer of transposons across phyla. *Nature* 464:1347–1350.
14. Gilbert C, Hernandez SS, Flores-Benabib J, Smith EN, Feschotte C (2012) Rampant horizontal transfer of SPIN transposons in squamate reptiles. *Mol Biol Evol* 29(2):503–515.
15. Novick P, Smith J, Ray D, Boissinot S (2010) Independent and parallel lateral transfer of DNA transposons in tetrapod genomes. *Gene* 449:85–94.
16. Arensburg P, et al. (2011) Phylogenetic and functional characterization of the hAT transposon superfamily. *Genetics* 188:45–57.
17. Smit AF (1999) Interspersed repeats and other mementos of transposable elements in mammalian genomes. *Curr Opin Genet Dev* 9:657–663.
18. Ivics Z, Hackett PB, Plasterk RH, Izsvák Z (1997) Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* 91:501–510.
19. Baus J, Liu L, Heggstad AD, Sanz S, Fletcher BS (2005) Hyperactive transposase mutants of the Sleeping Beauty transposon. *Mol Ther* 12:1148–1156.
20. Hartl DL, Lozovskaya ER, Nurminski DI, Lohe AR (1997) What restricts the activity of mariner-like transposable elements. *Trends Genet* 13:197–201.
21. Yusa K, Zhou L, Li MA, Bradley A, Craig NL (2011) A hyperactive piggyBac transposase for mammalian applications. *Proc Natl Acad Sci USA* 108:1531–1536.
22. Huang X, et al. (2010) Gene transfer efficiency and genome-wide integration profiling of Sleeping Beauty, Tol2, and piggyBac transposons in human primary T cells. *Mol Ther* 18:1803–1813.
23. Galvan DL, et al. (2009) Genome-wide mapping of PiggyBac transposon integrations in primary human T cells. *J Immunother* 32:837–844.
24. Yant SR, et al. (2005) High-resolution genome-wide mapping of transposon integration in mammals. *Mol Cell Biol* 25:2085–2094.
25. Vigdal TJ, Kaufman CD, Izsvák Z, Voytas DF, Ivics Z (2002) Common physical properties of DNA affecting target site selection of sleeping beauty and other Tc1/mariner transposable elements. *J Mol Biol* 323:441–452.
26. Berry C, Hannehalli S, Leipzig J, Bushman FD (2006) Selection of target sites for mobile DNA integration in the human genome. *PLoS Comput Biol* 2:e157.
27. Lewinski MK, et al. (2006) Retroviral DNA integration: Viral and cellular determinants of target-site selection. *PLoS Pathog* 2:e60.
28. Wu X, Li Y, Crise B, Burgess SM (2003) Transcription start regions in the human genome are favored targets for MLV integration. *Science* 300:1749–1751.
29. Miller DG, et al. (2005) Large-scale analysis of adeno-associated virus vector integration sites in normal human cells. *J Virol* 79:11434–11442.
30. Schröder AR, et al. (2002) HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* 110:521–529.
31. Wang GP, Ciuffi A, Leipzig J, Berry CC, Bushman FD (2007) HIV integration site selection: Analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. *Genome Res* 17:1186–1194.
32. Meylan S, et al. (2011) A gene-rich, transcriptionally active environment and the pre-deposition of repressive marks are predictive of susceptibility to KRAB/KAP1-mediated silencing. *BMC Genomics* 12:378.
33. Gangadharan S, Mularoni L, Fain-Thornton J, Wheelan SJ, Craig NL (2010) DNA transposon Hermes inserts into DNA in nucleosome-free regions in vivo. *Proc Natl Acad Sci USA* 107:21966–21972.
34. Weil CF, Kunze R (2000) Transposition of maize Ac/Ds transposable elements in the yeast *Saccharomyces cerevisiae*. *Nat Genet* 26:187–190.
35. Mitra R, Fain-Thornton J, Craig NL (2008) piggyBac can bypass DNA synthesis during cut and paste transposition. *EMBO J* 27:1097–1109.
36. Yu X, Gabriel A (1999) Patching broken chromosomes with extranuclear cellular DNA. *Mol Cell* 4:873–881.
37. Yu J, Marshall K, Yamaguchi M, Haber JE, Weil CF (2004) Microhomology-dependent end joining and repair of transposon-induced DNA hairpins by host factors in *Saccharomyces cerevisiae*. *Mol Cell Biol* 24:1351–1364.
38. Marshall HM, et al. (2007) Role of PSIP1/LEDGF/p75 in lentiviral infectivity and integration targeting. *PLoS ONE* 2:e1340.
39. Zhou L, et al. (2004) Transposition of hAT elements links transposable elements and V (D)J recombination. *Nature* 432:995–1001.
40. Craig N (2002) *Mobile DNA II*, eds Craig N, Craigie R, Gellert M, Lambowitz A (ASM Press, Washington), pp 423–456.
41. Hickman AB, Chandler M, Dyda F (2010) Integrating prokaryotes and eukaryotes: DNA transposases in light of structure. *Crit Rev Biochem Mol Biol* 45:50–69.
42. Cary LC, et al. (1989) Transposon mutagenesis of baculoviruses: Analysis of Trichoplusia ni transposon IFP2 insertions within the FP-locus of nuclear polyhedrosis viruses. *Virology* 172:156–169.
43. Grabundzija I, et al. (2010) Comparative analysis of transposable element vector systems in human cells. *Mol Ther* 18:1200–1209.
44. Elick TA, Bauser CA, Principe NM, Fraser MJ, Jr. (1996) PCR analysis of insertion site specificity, transcription, and structural uniformity of the Lepidopteran transposable element IFP2 in the TN-368 cell genome. *Genetica* 97:127–139.
45. Liu CL, et al. (2005) Single-nucleosome mapping of histone modifications in *S. cerevisiae*. *PLoS Biol* 3:e328.
46. Miller DG, Petek LM, Russell DW (2004) Adeno-associated virus vectors integrate at chromosome breakage sites. *Nat Genet* 36:767–773.
47. Ciuffi A, et al. (2005) A role for LEDGF/p75 in targeting HIV DNA integration. *Nat Med* 11:1287–1289.
48. Shun MC, et al. (2007) LEDGF/p75 functions downstream from preintegration complex formation to effect gene-specific HIV-1 integration. *Genes Dev* 21:1767–1778.
49. Dupuy A (2010) Splendours and miseries of the benefit-risk ratio. *Ann Dermatol Venereol* 137:267–268.
50. Ding S, et al. (2005) Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell* 122:473–483.
51. Dupuy AJ, Akagi K, Largaespada DA, Copeland NG, Jenkins NA (2005) Mammalian mutagenesis using a highly mobile somatic *Sleeping Beauty* transposon system. *Nature* 436:221–226.

52. Dupuy AJ, Jenkins NA, Copeland NG (2006) Sleeping beauty: A novel cancer gene discovery tool. *Hum Mol Genet* 15:R75–R79.
53. Ivics Z, Izsvák Z (2006) Transposons for gene therapy! *Curr Gene Ther* 6:593–607.
54. Hori H, Suzuki M, Inagaki H, Oshima T, Koga A (1998) An active Ac-like transposable element in teleost fish. *J Mar Biotechnol* 6:206–207.
55. Belay E, et al. (2010) Novel hyperactive transposons for genetic modification of induced pluripotent and adult stem cells: A nonviral paradigm for coaxed differentiation. *Stem Cells* 28:1760–1771.
56. Xue XHX, et al. (2009) Stable gene transfer and expression in cord blood-derived CD34+ hematopoietic stem and progenitor cells by a hyperactive Sleeping Beauty transposon system. *Blood* 114:1319–1330.
57. Brachmann CB, et al. (1998) Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other applications. *Yeast* 14:115–132.
58. Ocwieja KE, et al. (2011) HIV integration targeting: A pathway involving Transportin-3 and the nuclear pore protein RanBP2. *PLoS Pathog* 7:e1001313.