

A single amino acid switch converts the *Sleeping Beauty* transposase into an efficient unidirectional excisionase with utility in stem cell reprogramming

Lisa Kesselring¹, Csaba Miskey¹, Cecilia Zuliani², Irma Querques², Vladimir Kapitonov¹, Andrea Laukó³, Anita Fehér³, Antonio Palazzo⁴, Tanja Diem¹, Janna Lustig¹, Attila Sebe¹, Yongming Wang⁵, András Dinnyés³, Zsuzsanna Izsvák⁵, Orsolya Barabas² and Zoltán Ivics^{1,*}

¹Transposition and Genome Engineering, Division of Medical Biotechnology, Paul Ehrlich Institute, Langen, Germany, ²Structural and Computational Biology Unit, European Molecular Biology Laboratory, Heidelberg 69117, Germany, ³BioTalentum Ltd, Gödöllő, 2100 Gödöllő, Hungary, ⁴Department of Biology, University of Bari 'Aldo Moro', Italy and ⁵Mobile DNA, Max Delbrück Center for Molecular Medicine, Berlin, Germany

Received December 18, 2018; Revised November 07, 2019; Editorial Decision November 12, 2019; Accepted November 22, 2019

ABSTRACT

The *Sleeping Beauty* (SB) transposon is an advanced tool for genetic engineering and a useful model to investigate cut-and-paste DNA transposition in vertebrate cells. Here, we identify novel SB transposase mutants that display efficient and canonical excision but practically unmeasurable genomic re-integration. Based on phylogenetic analyses, we establish compensating amino acid replacements that fully rescue the integration defect of these mutants, suggesting epistasis between these amino acid residues. We further show that the transposons excised by the exc^+/int^- transposase mutants form extrachromosomal circles that cannot undergo a further round of transposition, thereby representing dead-end products of the excision reaction. Finally, we demonstrate the utility of the exc^+/int^- transposase in cassette removal for the generation of reprogramming factor-free induced pluripotent stem cells. Lack of genomic integration and formation of transposon circles following excision is reminiscent of signal sequence removal during V(D)J recombination, and implies that cut-and-paste DNA transposition can be converted to a unidirectional process by a single amino acid change.

INTRODUCTION

DNA transposable elements, or transposons for short, are mobile genetic elements capable of moving from one ge-

netic location to another in the genome. Most DNA transposons are mobilized by a cut-and-paste transposition reaction, that minimally requires a transposase protein and the terminal inverted repeat (TIR) sequences of the transposon. During transposition, the transposase (i) interacts with its binding sites in the TIRs, (ii) promotes the assembly of a synaptic complex, also called paired-end complex (PEC), (iii) catalyzes excision of the element out of its donor site and (iv) integrates the excised transposon at a new location in target DNA. The majority of known transposases, similarly to retroviral and retrotransposon integrases and the RAG1 V(D)J recombinase, contain a highly conserved aspartate-aspartate-glutamate (DDE) amino acid triad in their RNaseH-type catalytic domains (1–4). These amino acids play an essential role by coordinating Mg^{2+} ions required for the catalytic steps (DNA cleavage and joining) of transposition (5,6).

The key biochemical step of all transposon excision reactions executed by DDE enzymes is the release of 3'-OH groups at each transposon end, which are then used in the strand transfer reaction during integration (7,8). First, a single DNA strand is nicked by transposase-catalyzed hydrolysis of the phosphodiester bond in the DNA backbone (7). During cut-and-paste transposition, nicking is followed by cleavage of the complementary DNA strand resulting in a double-strand break (DSB) that liberates the transposon from the donor DNA (Supplementary Figure S1). To catalyze second strand cleavage, DDE enzymes developed versatile strategies (9–11). Most DDE transposases use a single active site to cleave both DNA strands at one transposon end via a DNA hairpin intermediate [reviewed in (11)] either on the transposon end (12–15) or on the flanking donor

*To whom correspondence should be addressed. Tel: +49 6103 77 6000; Fax: +49 6103 77 1280; Email: zoltan.ivics@pei.de
Present address: Laboratory of Translational Nanotechnology, IRCCS - Istituto Tumori 'Giovanni Paolo II', Bari, Italy

DNA (16–20). Members of the Tc1/*mariner* family do not transpose via a hairpin intermediate (21,22), indicating that double-strand cleavage is the result of two sequential hydrolysis reactions by the transposase (23). The second step of the transposition reaction is the transfer of the free 3'-OH groups on the transposon ends to the target DNA molecule by transesterification. Similarly to the initial DNA cleavage, strand transfer is executed by a nucleophilic attack. In this case, the 3'-OH groups of the transposon serve as nucleophiles, directly coupling the element to the target without previous target DNA cleavage (Supplementary Figure S1).

Sleeping Beauty (SB) is a synthetic transposon that was constructed based on sequences of transpositionally inactive elements isolated from fish genomes (24). SB is a Tc1/*mariner* superfamily transposon and follows a classical cut-and-paste transposition reaction. It supports a full spectrum of genetic engineering applications/methods [reviewed in (25)] including the generation of transgenic cell lines, induced pluripotent stem cell (iPSC) reprogramming (26–31), phenotype-driven insertional mutagenesis screens in the area of cancer biology [reviewed in (32–34)], germline gene transfer in experimental animals (35–41) and somatic gene therapy both *ex vivo* and *in vivo* [reviewed in (25,42–48)].

In most of the genetic engineering applications highlighted above, permanent insertion of a transgene cassette is required for long-term or even permanent expression of a gene of interest. However, certain applications would benefit from ‘transient transgenesis’, where presence and expression of a gene of interest is only transiently required. One such paradigmatic application is the generation of iPSCs with reprogramming transcription factors, where presence of these factors is only required during reprogramming but dispensable or even undesired once iPSCs are established. Transient delivery of reprogramming factors can be accomplished by non-integrating vector systems (49) or by genomic integration of expression cassettes followed by their excision, so as to result in genetically ‘clean’ but phenotypically altered cells. Indeed, reprogramming factor-free iPSCs have been generated by applying components of the FLP- and Cre-recombinase systems to either delete or exchange the genomically integrated reprogramming factors (26,50). One particular feature of transposon-based vectors is that transposon excision is not always followed by re-integration into a new genomic location. Thus, transposase-mediated excision offers an opportunity for removal of the transgenes after completion of reprogramming. Transposition-mediated generation of mouse and human iPSCs cells and subsequent removal of the reprogramming factors from the pluripotent cells by transient re-expression of the transposase have already been achieved with the *piggyBac* (PB) system (51,52). One caveat that still remains is the possibility of the transposon to jump into a new location during the factor removal process. Indeed, it was estimated that ~75% of SB transposon excision events are followed by chromosomal integration (53). A way to solve this problem would be to develop a transposase that allows cutting but is deficient in pasting.

Evidence for the excision and integration reactions being separated in space and time came from mutational analysis of the Tn10 transposase, which identified mutants that

were proficient in transposon excision but deficient in transposon integration (exc^+/int^- phenotype) (54). Excised but not reintegrated transposon molecules were also observed during transposition of Tc1, Tc3 or *Minos* elements (55–57). Recent mutational screening also identified PB transposase mutants with an exc^+/int^- phenotype (58). Previous alanine-scanning mutagenesis across the SB transposase identified a mutant, K248A, which retained transposon excision activity but severely impaired re-integration (59). Unfortunately, excision efficiency of this mutant is greatly reduced, and it generates significantly more sequence heterogeneity at the excision sites than wild-type SB transposase, which compromise its further use for precision genetic engineering.

Here we systematically explored position 248 in the SB transposase to gain further insight into the transposition mechanism and to create novel mutants for genetic engineering. The position of K248 in a target-capture structural model of the SB transposase suggests that this amino acid is involved in interaction with target DNA. Consistently, we identify novel SB transposase mutants with a phenotype of efficient and precise excision and no measurable genomic re-integration. In the absence of integration, transposon excision by these transposase mutants results in extrachromosomal circles that cannot undergo a further round of transposition, thereby terminating the transposition reaction. We demonstrate the utility of the exc^+/int^- transposase in ‘transient transgenesis’ for the generation of reprogramming factor-free induced pluripotent stem cells.

MATERIALS AND METHODS

Phylogenetic sequence analyses

A diverse set of 76 transposase amino acid sequences encoded by Tc1 transposons was collected from transposons described in the literature, reported in Repbase and identified by us in GenBank. We used FGENESH+ (60) to predict the exon-intron structure of the Tc1 transposase encoded by transposons reported in Repbase. The multiple alignment of the transposase amino acid sequences was obtained by using MAFFT (61). The multiple alignment was edited manually by deletion of short variable N- and C-terminal regions that do not constitute the Tc1 catalytic core. The set of collected transposases is reported as a supplemental file Tc1.fa, the full-length multiple alignment is reported as a supplemental file Tc1.mfa, the manually edited multiple alignment used for the obtaining the phylogenetic tree is reported as the Tc1_core1.mfa supplemental file. The phylogenetic tree was built by using PhyML (62) (automatic model selection, Akaike information criterion, SPR improvement with 5 random starting trees, aLRT branch support). The same tree structure was supported by different modifications of the PhyML parameters.

Structural modeling

The full-length Tn10 transposase structure was modeled using the crystal structure of the Tn5 transposase [Protein Data Bank ID 1MUS; (63)] via Phyre2 (64,65) in intensive mode. The SB target capture complex (TCC) model was assembled based on the SB PEC homology model described

in (66); the transposase catalytic domain was replaced with the crystal structure (67), and target DNA was modelled by docking the integration target substrate from the PFV intasome structure [as in (67)]. Finally, the assembled model was refined in HADDOCK (68). Structural models were visualized using the PyMOL Molecular Graphics System (Version 1.5.0.4, Schrödinger, LLC).

Site-directed mutagenesis of the SB100X transposase

All mutations were generated using the Q5 polymerase (NEB, Ipswich, MA, USA) and the plasmid pCMV(CAT)T7-SB100X (44). 5'-phosphorylated primers for the particular positions were created with 'NEBaseChanger™' (<http://nebasechanger.neb.com/>) and were designed with 5'-ends annealing back-to-back. Primers were synthesized with a 5'-phosphate to enable a downstream intramolecular ligation reaction and were ordered from Eurofins (Eurofins/MWG, Luxembourg). The primer sequences are listed in Supplemental Methods. PCR cycling conditions were set according to the manufacturer's instructions. The annealing temperatures of the mutagenic primers were calculated with the 'NEB Tm calculator™' software (<https://www.neb.com/tools-and-resources/interactive-tools/tm-calculator>). The PCR products were purified with QIAquick PCR Purification Kit (QIAGEN, Venlo, Holland), eluted in 30 µl elution buffer and digested with 2 µl *DpnI* (NEB, Ipswich, MA, USA) for 2 h at 37°C, followed by 20 min heat-inactivation at 80°C. The linear, double-stranded PCR products were circularized by ligation with T4 DNA Ligase (NEB, Ipswich, MA, USA) overnight at room temperature. The circularized PCR products were transformed into chemically competent *E. coli* (Invitrogen/Life Technologies, Carlsbad, CA, USA), grown in Luria-Bertani (LB) medium for 1 h, and selected for chloramphenicol resistance by plating on LB agar plates containing 25 µg/ml chloramphenicol. To confirm the presence of the desired mutations, DNAs from several single colonies were purified using the QIAprep spin miniprep kit (QIAGEN, Venlo, Holland) and were sequenced by GATC Biotech (Konstanz, Germany).

Western blotting

One day prior to transfection, 2×10^5 HeLa cells were seeded onto six well plates. 1.5 µl of TransIT-LT1 transfection reagent (Mirus Bio LLC, Madison, WI, USA) was used to transfect 500 ng of DNA (each transfection reaction was filled up to 500 ng with the plasmid pUC19). 50 ng of transposase plasmid [mutant SB100X transposase or wild-type SB100X (pCMV(CAT)T7-SB100X)] or green fluorescent protein (GFP) expression plasmid (pmaxGFP™; Lonza, Basel, Switzerland) were transfected. 48 h post transfection, HeLa cells were lysed in RIPA buffer (150 mM NaCl, 1.0% Triton X-100, 1.0% Na-deoxycholate, 0.1% SDS, 25 mM Tris, pH 8.0) supplemented with protease inhibitor cocktail (Complete Mini, Roche, Basel, Switzerland). Protein was extracted using the Bioruptor® Plus (Diagenode, Denville, NJ, USA), 10 cycles, high power, 30 s ON/30 s OFF at 4°C. Total protein was quantified using BCA Protein Assay

Kit (Pierce, Rockford, IL, USA). Proteins (10 µg per lane) were loaded onto 10% polyacrylamide gels and subjected to sodium dodecyl sulfate–polyacrylamide gel electrophoresis. Gels were transferred to nitrocellulose membrane (Hybond ECL, Amersham Bioscience, Little Chalfont, UK) and immunoblotting was performed according to standard procedures. Proteins were detected with goat monoclonal anti-SB transposase antibody (R&D Systems, Minneapolis, USA) at dilution 1:5000, or mouse monoclonal anti-Vinculin antibody (Abcam, Cambridge, UK) at dilution 1:3000, and chemiluminescence using ECL Prime Western Blotting Detection Kit (Amersham Bioscience). Signals were captured on a film (Hyperfilm ECL High performance chemiluminescence film, Amersham Bioscience, Little Chalfont, UK).

FACS-based excision assay

1.5×10^5 HeLa cells were seeded one day prior to transfection. HeLa cells were transfected with 2 µg of the transposon donor plasmid pCMV(CAT)-GFP//T2Neo (67) and 200 ng of a mutant SB100X transposase, SB100X or a catalytically inactive transposase mutant (E279D, hereafter called D3). As a GFP control we transfected 1 µg of an SB transposon expression plasmid for the green fluorescent protein (GFP) (pT2.CAGGS.AmGFP transposon; contains the GFP gene driven by the CAGGS promoter). Six microliters of TransIT-LT1 transfection reagent (Mirus Bio LLC, Madison, WI, USA) were used per transfection reaction. Three days post-transfection cells were trypsinized, washed with PBS, fixed in 1% Paraformaldehyde (PFA) in PBS and FACS analyzed with BD LSR II flow cytometer (BD Biosciences, Franklin Lakes, NJ, USA). Data were analyzed with FCS Express 4 Flow Cytometry (De Novo Software, Glendale, CA). The % GFP positive cells in cultures transfected with SB100X transposase mutants/pCMV(CAT)-GFP//T2Neo were normalized to that in cells transfected with wild-type SB100X/pCMV(CAT)-GFP//T2Neo.

Transposition assay

For transposition assay in HeLa cells, 2×10^5 cells were seeded onto six-well plates one day prior to transfection. Three microliters of TransIT-LT1 transfection reagent (Mirus Bio LLC, Madison, WI USA) was used to transfect 1 µg of DNA (each transfection reaction was filled up to 1 µg with the plasmid pUC19). To obtain predominantly single-copy transposon insertions (69), 10 ng of transposon donor plasmid (pT2B/puro) were co-transfected with 5 ng of transposase plasmid (mutant transposase, inactive D3 transposase or SB100X). Forty-eight hours after transfection cells were trypsinized and 2.5–5% of the cells were replated to 10 cm plates and selected for transposon integration using 3 µg/ml puromycin (InvivoGen, San Diego, CA, USA). After 3 weeks of selection, cell colonies were fixed with 10% (vol/vol) formaldehyde in phosphate-buffered saline (PBS), stained with methylene blue in PBS, and counted. For *in vitro* comparisons of relative transposition efficiencies at least three independent experiments were performed.

Single-copy transposon remobilization assay

For remobilization assays in HeLa cells, duplicates of 3×10^5 cells were seeded onto six well plates one day prior to transfection. Six microliters of TransIT-LT1 transfection reagent (Mirus Bio LLC, Madison, WI, USA) was used to transfect 1 μg of DNA (each transfection reaction was filled up to 1 μg with the plasmid pUC19). To mobilize the SB transposon, 500 ng of plasmids expressing the mutant transposases K248S or K248T were transfected. As controls 500 ng of pCMV(CAT)T7-SB100X or 500 ng plasmid expressing the D3 catalytically inactive SB transposase were used. Forty-eight hours after transfection cells were trypsinized, replated to 10 cm plates and selected for transposon excision using 3 $\mu\text{g}/\text{ml}$ puromycin (InvivoGen, San Diego, CA, USA) or selected for transposon integration after excision using 3 $\mu\text{g}/\text{ml}$ puromycin and 600 $\mu\text{g}/\text{ml}$ G418. After 3 weeks of selection, cell colonies were fixed with 10% (vol/vol) formaldehyde in phosphate-buffered saline (PBS), stained with methylene blue in PBS, and counted. For *in vitro* comparisons of relative transposition efficiencies at least three independent experiments were performed.

Transposon footprint analysis

These analyses were done essentially as described earlier (22). 1×10^6 cultured HeLa cells were co-transfected with 5 μg of a puromycin resistance gene-containing donor plasmid (pT2B/puro) or with the pCMV(CAT)-GFP//T2Neo plasmid and 500 ng of either a mutant transposase expression plasmid, a plasmid expressing the SB100X transposase (positive control) or a plasmid expressing the D3 catalytically inactive transposase mutant (negative control). Two days post-transfection, cells were harvested, low molecular weight DNA was isolated and digested with *StyI* overnight at 37°C (to get rid of the vector backbones). 2 μl of digested plasmids were added to a 50 μl PCR reaction mixture containing 1 \times Thermo buffer, 200 μM dNTPs, 0.2 μM forward and reverse primer each and 1.25 units of *Taq* Polymerase (NEB). A series of PCR primers has been used for nested excision products detecting PCR reactions. In the first PCR round SB excision products were amplified using primers PUC2 and PUC5 (for pT2B/puro). The PCR conditions were as follows: initial denaturation of 5 min at 95°C, followed by 30 cycles of 30 s denaturation at 95°C, 30 s annealing at 63°C, 15 s elongation at 72°C and a final 5 min elongation at 72°C. One μl of a 1:50 dilution of the first PCR product was used for the second round PCR using nested primers PUC3 and PUC4. The composition of the PCR mixture was otherwise the same. Nested PCR was performed with 5 min initial denaturation at 95°C, followed by 30 cycles of 30 s denaturation at 95°C, 30 s annealing at 65°C, 15 s elongation at 72°C and a final 5 min elongation at 72°C. For footprint analysis using pCMV(CAT)-GFP//T2Neo first-round PCRs were done with primers GFP-fw1 and GFP-rev1 and second round with GFP-fw2 and GFP-rev2 with initial denaturation 94°C – 1 min fol-

lowed by 30 cycles of 94°C – 30 s, 60°C – 30 s and 68°C – 15 s and final elongation 68°C – 3 min.

For footprint analysis following transposon excision from a genomic locus, 1.5×10^6 transgenic HepG2 and HeLa cells harboring single-copy SB transposons within the puromycin N-acetyl-transferase ORF (*puro*) were plated to 10 cm cell culture plates, and transfected with 3 μg expression plasmids encoding the SB100X, K248S and K248T transposases. Genomic DNA of the transfected cells was isolated with the Quick-DNA Miniprep Kit (Zymo Research) from cells with or without 1 $\mu\text{g}/\text{ml}$ puromycin selection 5 days post-transfection. 500 ng of DNA was used for amplifying the genomic excision sites using Q5 Polymerase (NEB) with GC enhancer with the PCR conditions: 98°C 30 s; 30 cycles of 98°C 10 s, 70°C 30 s; 72°C 2 min with the primers puroRev0.5 and puroFw1. The PCR products were purified with the Clean and Concentrator Kit (Zymo Research), eluted in 20 μl and 1 μl elute was used for the nested PCR reaction with primers puroRev1.5 and puroFw1.5 using the PCR program: 98°C 30 s; 30 cycles of 98°C 10 s, 68°C 30 s; 72°C 2 min. Amplification products were separated by loading 10 μl of the second PCR reaction on a 1.5% agarose gel using Midori Green staining. The PCR products were cut out from the agarose gel, purified by column (Zymo Research, Irvine, CA, USA) and cloned by pGEM-T-Easy Vector Kit (Promega, Fitchburg, WI, USA) or by the pJet PCR Cloning Kit (Thermo Fisher Scientific). Colonies were picked and assayed by colony PCR amplifying the ligated excision products. Colonies were resuspended in 10 μl dH₂O each and added to a 25 μl PCR reaction mixture containing 1 \times Standard buffer, 200 μM dNTPs, 1 mM MgCl₂, 0.2 μM T7 primer and SP6 primer and 1.25 units of *Taq* Polymerase (NEB). The PCR conditions were as followed: initial denaturation of 5 min at 94°C, followed by 25 cycles of 30 s denaturation at 94°C, 30 s annealing at 55°C, 30 s elongation at 72°C and a final 5 min elongation at 72°C. PCR with primers pJet1.2Fw and pJet1.2Rev were done using the Q5 polymerase and GC enhancer with the following conditions: 98°C 30 s; 25 cycles of 98°C 10 s, 66°C 10 s, 72°C 30 s; 72°C 2 min. Amplification products were separated by loading the whole PCR reaction on a 1.5% agarose gel using Midori Green staining. The PCR products were cut out from the agarose gel and purified by column (Zymo Research, Irvine, CA, USA) and eluted in 12 μl elution buffer. The purified PCR products were sequenced using PUC3 primer (or with pJet1.2Fw).

Circle formation assay

HeLa cells were seeded in 10 cm petri dishes and transfected with 500 ng of a transposase expression plasmid (SB100X, K248S, K248T) and 5 μg of the zeocin resistance gene carrying transposon plasmid pT/zeo (70). Seventy two hours post-transfection cells were washed twice with PBS, trypsinized and the cell pellet was frozen overnight. The pellet was resuspended in 250 μl resuspension buffer (50 mM TrisHCl, 10 mM EDTA, 100 $\mu\text{g}/\text{ml}$ RNase A) and lysed by addition of 250 μl lysis buffer (200 mM NaOH, 1% SDS (w/v)). Through addition of 300 μl neutralization buffer

[3.0 M potassium acetate (CH₃CO₂K) pH5.5] high molecular weight genomic DNA precipitated. The supernatant was removed and low molecular weight DNA was precipitated with 0.8 vol isopropanol, washed with 70% ethanol and dissolved in 10 µl 10 mM Tris pH 8. The extracted plasmid DNA was subjected to restriction digestion with *Pvu*II that cleaves the original pT/zeo vector in the plasmid backbone, but leaves transposon circles intact. The digested DNA was purified by column (Zymo Research, Irvine, CA, USA), and electroporated into Top10 electrocompetent cells according to manufacturer's protocol. Bacteria were plated on LB agar plates containing 20 µg/ml zeocin directly after electroporation to avoid clonal propagation of individual products in liquid culture. Twenty individual clones were analyzed for circle formation by colony PCR. The colonies were resuspended in 10 µl dH₂O and added to a 15 µl PCR reaction mixture containing 1x Standard buffer, 200 µM dNTPs, 1 mM MgCl₂, 0.2 µM SB_IRDR_R_FW primer and SB_IRDR_L_RV primer each and 1.25 units of *Taq* Polymerase (NEB). The PCR conditions were as follows: initial denaturation of 5 min at 94°C, followed by 25 cycles of 30 s denaturation at 94°C, 30 s annealing at 51°C, 45 s elongation at 72°C and a final 5 min elongation at 72°C. Amplification products were separated by loading the whole PCR reaction on a 2% agarose gel using Midori Green staining. DNA fragments were excised from the agarose gel and purified using the Zymoclean™ Gel DNA Recovery Kit. In the last step the DNA was eluted in 20 µl EB buffer (Zymo Research) and sequenced for circle formation using SB_IRDR_R_FW primer.

Work with iPSCs

Mouse iPSCs were cultured in knockout DMEM with 15% FBS, 2 mM L-glutamine, 1x nonessential amino acids, 0.1 mM 2-mercaptoethanol, 25 µg/ml L-ascorbic acid and 10 ng/ml purified recombinant leukemia inhibitory factor (LIF), and were maintained on Geltrex™-coated plates (Thermo Fisher; 1:100 dilution in DMEM F12). For transfection, iPSCs were plated onto a six-well plate at a density of 2×10^5 per well one day prior to transfection. On the following day, the cells were transfected with 10 µg of K248T transposase expression plasmid using Lipofectamine 3000 transfection reagent (Thermo Fisher). Two days after transfection, the cells were washed with PBS, trypsinized with 400 µl trypsin and centrifuged at 800 rpm for 10 min. The supernatant was aspirated and the pellet was resuspended in mouse iPSC medium with gently pipetting in a total volume of 3 ml and transferred into pre-warmed mouse iPSC medium in a 10 cm Geltrex™-coated dish. Next day, the FIAU (Moravek Biochemicals) selection (200 ng/ml) was started. After selection for 2 weeks, one surviving colony that expressed GFP was trypsinized, picked mechanically and transferred to a Geltrex™-coated 10 cm dish. The cells were expanded for 5 days in mouse iPS medium with FIAU and genomic DNA of the parental iPSC line and the K248T 'cassette-removed' iPSC line was isolated with DNeasy Blood & Tissue Kit (Qiagen, Venlo, Netherlands). 1 µg of samples were digested with *Dpn*II for

18 h. 100 ng of the digested samples were ligated to *Dpn*II-splinkerette linkers (25 pmol) in 20 µl reactions. Five microliters of the ligated DNA were used for the first PCR with Linker Primer and T-Bal Rev with a cycle of 94°C for 3 min, followed by 10 cycles of 94°C for 30 s, ramp down at 0.5°C/s to 63°C, anneal at 63°C for 30 s and 1 min elongation at 72°C; After the first 10 cycles of PCR, a second 25 cycle amplification was performed with 94°C for 30 s, ramp down at 1°C/s to 61°C, anneal at 61°C for 30 s, 1 min elongation at 72°C and a final elongation at 72°C for 10 min. 1 µl of a 1:100 dilution of the first PCR was used for nested PCR using primers Nested and T-Bal. Nested PCR was performed with 3 min initial denaturation at 94°C, followed by 10 cycles of 94°C for 30 s, ramp down at 1°C/s to 64°C, anneal at 64°C for 30 s and 1 min 30 s elongation at 68°C; After the first 10 cycles of PCR, a second 25 cycle amplification was performed with 94°C for 30 s, ramp down at 1°C/s to 60°C, anneal at 61°C for 20 s and 1 min 30 s elongation at 68°C and a final elongation at 68°C for 10 min.

For immunocytochemistry, parental and cassette-removed iPSCs were seeded at a density of 1×10^5 cells per well onto Nunc™ Lab-Tek™ 2-well Chambered Coverglass (Thermo Fisher) coated with Geltrex™. On the following day, cells were fixed with 4% paraformaldehyde solution (Santa Cruz Biotechnology) and permeabilized with 0.1% (v/v) Triton™ X-100 (Sigma-Aldrich). Slides were washed three times with PBS and incubated with 5% (v/v) BSA (Sigma-Aldrich) for 30 min at room temperature to block unspecific binding. Slides were incubated with anti-nanog antibody (rabbit polyclonal; Bethyl Laboratories) in 1% (v/v) BSA for 2 h at room temperature. Slides were washed 4 times with 1% (v/v) BSA and incubated for 45 min in the dark with anti-rabbit Alexa Fluor® 647 antibody (diluted 1:1000 in 1% (v/v) BSA; Thermo Fisher) and DAPI. Cells were imaged with a Zeiss LSM 780 confocal microscope (using a 20x objective) in the ALMF core facility at EMBL Heidelberg.

RESULTS

The K248 group of Tc1 transposons

Earlier alanine-scanning mutagenesis found that the K248A mutation in the SB transposase diminished transposon integration (59). To explore the sequence diversity at this position, we performed phylogenetic analysis and a multiple sequence alignment of Tc1/*mariner* transposases. This revealed that transposases containing lysines at positions corresponding to SB's K248 form a separate group (called here the K248 group) (Figure 1A). In addition to the diagnostic K248, all K248 group transposase sequences, including those encoded by transposons in animals, land plants, fungi and protists, also contain universally conserved D in position 246. The K248 group can be subdivided into four subgroups that we named SB (vertebrates), S (cnidarians, insects, hexapods, tardigrades and planarians), Tc1 (insects and vertebrates) and FPP (from fungi, plants and protists) (Figure 1A).

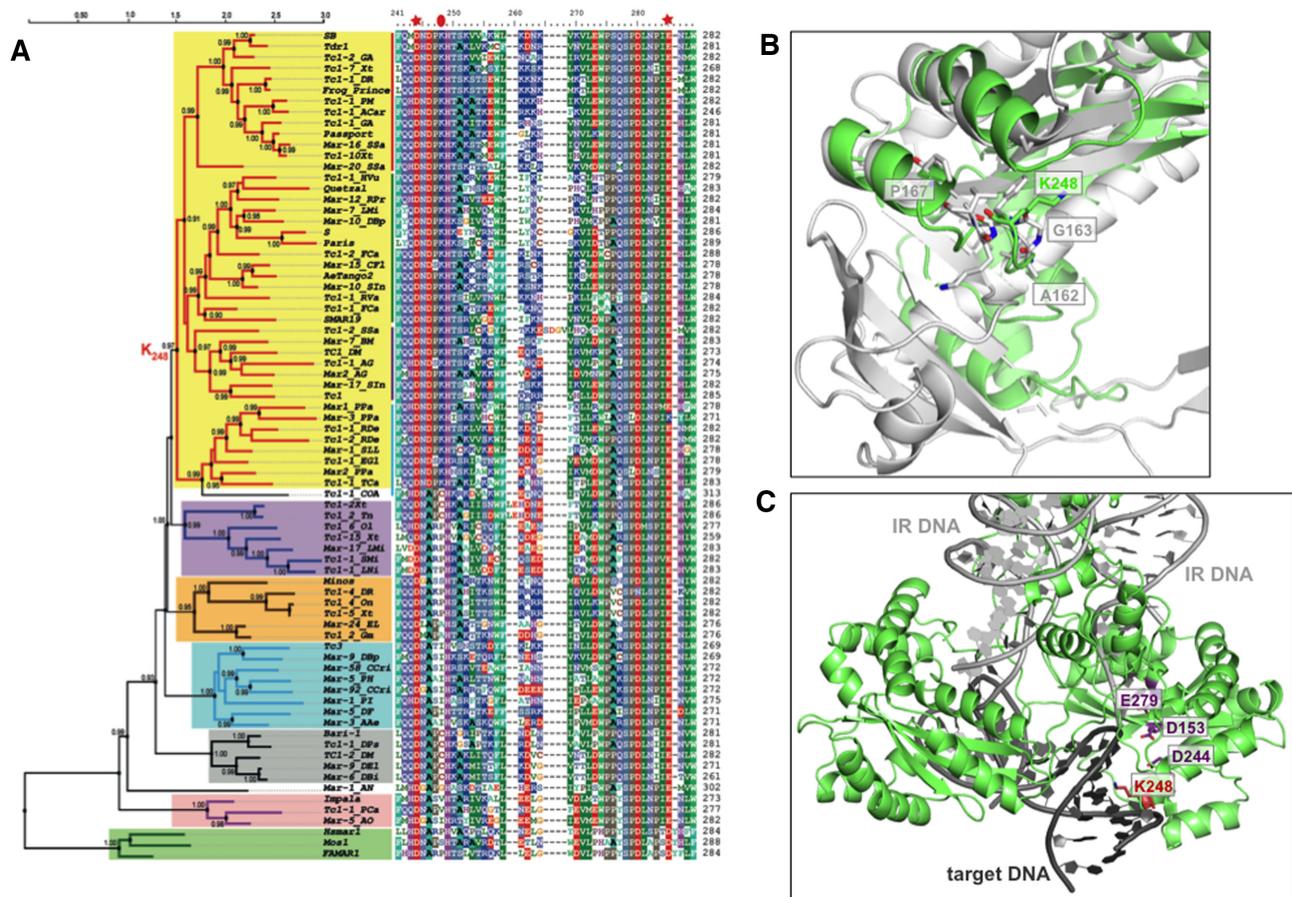


Figure 1. Phylogeny of Tc1/*mariner* transposase amino acid sequences. (A) The maximum-likelihood phylogenetic tree obtained from the multiple alignment of the full-length transposase sequences by using PhyML is accompanied by a portion of the multiple alignment that corresponds to a region of the SB transposase between amino acid positions 241 and 282. The numbers in the right column indicate the positions of the last shown amino acids in the corresponding transposase sequences. Red asterisks mark the last two amino acid residues of the Tc1/*mariner* catalytic triad. The red oval indicates the diagnostic position of the multiple alignment that corresponds to K248 in the SB transposase. All aLRT estimates of the bootstrap support values ≥ 0.9 are indicated by the numbers at the tree vertices. Different groups constituting the Tc1 clade are marked by rectangles of different colors. The *mariner* clade transposases are marked by the green rectangle. The new K248 group of Tc1 transposases is marked by the yellow rectangle and red branches. The Tc1-2Xt, Minos, Tc3, Bari and Impala groups of the Tc1 clade are marked by the purple, light-brown, blue, gray and pink rectangles, respectively. The K248 group is composed of four subgroups called SB, S, Tc1 and FPP that are marked by the red, green, violet and blue vertical lines, respectively. Transposon nomenclature, host species and GenBank accession numbers can be found in Supplementary Information. (B) SB transposase target capture complex superimposed onto a Tn10 transposase model. SB is in green, Tn10 is in gray. Protein chains are shown in cartoon representation. K248 in SB and Tn10 residues where mutations cause exc^+/int^- phenotype are shown in sticks representation with atomic coloring. (C) Location of K248 in the SB target capture complex. Transposon and target DNA are shown as gray and black cartoon, respectively. The catalytic D244 residue and K248 are highlighted in sticks representation. K248 is situated in close proximity to the target DNA.

Strong conservation of K248 in a particular group of Tc1 transposases clearly implies functional importance of this residue. Previously, a mutant screen identified 13 amino acids in the Tn10 transposase that display an exc^+/int^- phenotype similar to that of the SB K248A mutant. Thus, to compare the structural context of these protein residues, we have modeled the Tn10 transposase structure and overlaid this with the SB transposase target capture complex (TCC). Tn10 was modeled based on the related bacterial transposase Tn5 (63,71), and the SB TCC model was built on the recently determined SB100X catalytic domain structure (66,67). Despite low sequence identity ($<5\%$) between the two transposases, their core RNaseH-fold superposes well, with major differences restricted to the distinct insertion domains (a beta-stranded domain in Tn5 and an extended clamp-loop in SB). Strikingly, K248 of SB overlaps

with a group of Tn10 mutations (A162, G163 and P167) (Figure 1B) that have very severe defects in transposon integration (54). Notably, K248 is situated in close proximity to target DNA in the SB TCC model, implying a role in target binding and/or integration (Figure 1C). We conclude that phylogenetic conservation and structural position imply a fundamental role of K248 in transposon integration.

Saturation mutagenesis of K248 in the *Sleeping Beauty* transposase identifies mutants with an excision-proficient/integration-deficient phenotype

In order to assess the relative effects of single amino acid replacements at position 248 on transposition, the SB100X transposase was subjected to saturation mutagenesis by incorporating all possible amino acids by site-directed PCR

mutagenesis. All constructs encoding mutant versions of the SB100X transposase showed protein expression levels equivalent to that of SB100X by Western blot analysis (Supplementary Figure S2).

We evaluated the mutants for their excision and full transposition (excision + integration) activities relative to SB100X in human cells. To measure transposon excision, a FACS-based excision assay was performed. For this, we transiently co-transfected plasmids expressing the different transposases with a transposon donor construct, in which a GFP coding sequence is disrupted by an SB transposon. Precise excision of the SB transposon restores the GFP ORF, leading to fluorescence that can be detected by FACS analysis (67) (Figure 2A). The analysis revealed that all mutations had a negative impact on excision activity; roughly 50% (10 out of 19) of the mutants retained activity in a range of 15–80% relative to SB100X, whereas the remaining 50% (9 out of 19) displayed barely measurable activity or complete loss of excision (Figure 2B). This is not surprising given the strong conservation of lysine in this position across the K248 group of Tc1 family transposases, as described above. In order to assess the relative integration activity of each transposase mutant, we performed a full transposition assay that is based on the generation of antibiotic-resistant cell clones as a result of integration of antibiotic resistance gene-carrying SB transposons from donor plasmids into the chromosomes of transfected cells. Unexpectedly, the vast majority (15 out of 19) of mutants displayed a complete loss of function as defined by a colony forming potential similar to a catalytically inactive DDE mutant, which was used as negative control (Figure 2C). The data thus reveal a segregation of transposon excision and integration activities with two, arbitrarily defined clusters of mutants: those that retain $\geq 60\%$ excision activity of SB100X and show $\geq 10\%$ integration (K248C, K248I and K248V, $P < 0.01$), and those that retain 15–80% excision activity of SB100X, but show a complete loss of integration (K248A, K248N, K248L, K248M, K248S and K248T, $P < 0.001$ for excision and $P < 0.01$ for integration).

From the excision-proficient, integration-deficient (exc^+/int^-) mutants described above, K248S and K248T displayed the highest excision rates ($\sim 48\%$ and $\sim 76\%$, respectively, as compared to SB100X and measured by the GFP-based assay) (Figure 2 and Supplementary Figure S3). To further analyze the ability of these transposase mutants to mobilize a single, chromosomally located transposon, a transposon remobilization assay was used. This assay is suitable for independent, quantitative scoring of either excision or excision + integration activities of SB transposases (Figure 3A). The remobilization assay demonstrated that the K248S and K248T mutants are able to excise a chromosomally integrated SB transposon (K248S: $\sim 6\%$; K248T: $\sim 21\%$ relative excision activity compared to SB100X, $P < 0.001$), but fail to reintegrate the excised transposon back into the genome (Figure 3B). In these assays, calculating with a 70% transfection efficiency of the reporter cells, we estimate an excision rate of 0.0068% and 0.023% per transfected cell for the K248S and K248T mutants, respectively.

The D246A replacement rescues the integration-deficient phenotype of K248 mutants

Our phylogenetic analysis shown in Figure 1A revealed segregation of a distinct clade of transposases that do not contain the D246/K248 dyad. In these transposases the D246/K248 dyad is replaced by an A246[PICasy]248 dyad (where ‘PICasy’ represents six alternative amino acid residues at position 248 with the most frequent residues P, I and C in capital). These non-K248 transposons are represented by the *Minos* (insects and vertebrates), Tc3 (nematodes, insects, protists, crustaceans), Tc1–2Xt (vertebrates, insects), *Bari* (fruit flies) and *Impala* (fungi) groups (Figure 1A), and by transposases from the *mariner* clade and bacterial transposases that are ancestral to the *mariner* and Tc1 clades (data not shown).

Because the non-K248 transposases invariably contain A in position 246 (Figure 1A), we wondered if a D-to-A replacement in position 246 of the SB transposase could possibly rescue the integration defect in our exc^+/int^- mutants. We introduced the D246A mutation into the newly discovered K248S and K248T mutants and into the K248A mutant described previously (59). Strikingly, the double-mutant transposases displayed a full rescue of transposition activity as measured by the colony-forming assay (Figure 4). These results imply a cooperation between amino acids in position 246 and 248 in SB transposition.

The K248S and K248T *Sleeping Beauty* transposase mutants generate canonical excision footprints

DNA cleavage during transposon excision can occur at different positions relative to the transposon ends. Second strand cleavage occurs directly opposite the first strand cleavage site in V(D)J recombination (19) and for the bacterial Tn5 (72) and Tn10 elements (14), thereby generating blunt ended products. In case of the Tc1/*mariner* elements the non-transferred strand is cleaved a few nucleotides inside the transposon (22,53,73–75), thereby generating 3'-overhangs (Supplementary Figure S1). The prominent pathway for repairing transposon excision sites in somatic mammalian cells is nonhomologous end-joining (NHEJ), which generates transposon ‘footprints’ that are identical to the first or last 2–4 nucleotides of the transposon in Tc1/*mariner* transposition. Thus, canonical footprints generated by the SB transposase at the excision site are composed of 7 bp-sequences, either 5'-TACAGTA-3' or 5'-TACTGTA-3', comprising the terminal three base pairs of the transposon flanked by TA dinucleotides (Supplementary Figure S1) (22,53). Although NHEJ repair predominantly produces canonical footprints after SB excision in zebrafish embryos (76), mouse embryonic stem cells (53), Chinese hamster ovary cells (22) and somatic tissues of transgenic mice (77), repair in mouse spermatids tends to generate 1- and 2-bp footprints (78), and human HeLa cells produce an even more significant footprint sequence diversity (79,80). One likely explanation for dissimilar transposon footprints in different cell types/species is cell type-dependent variation in the activity of a critical host factor involved in DSB repair after transposon excision.

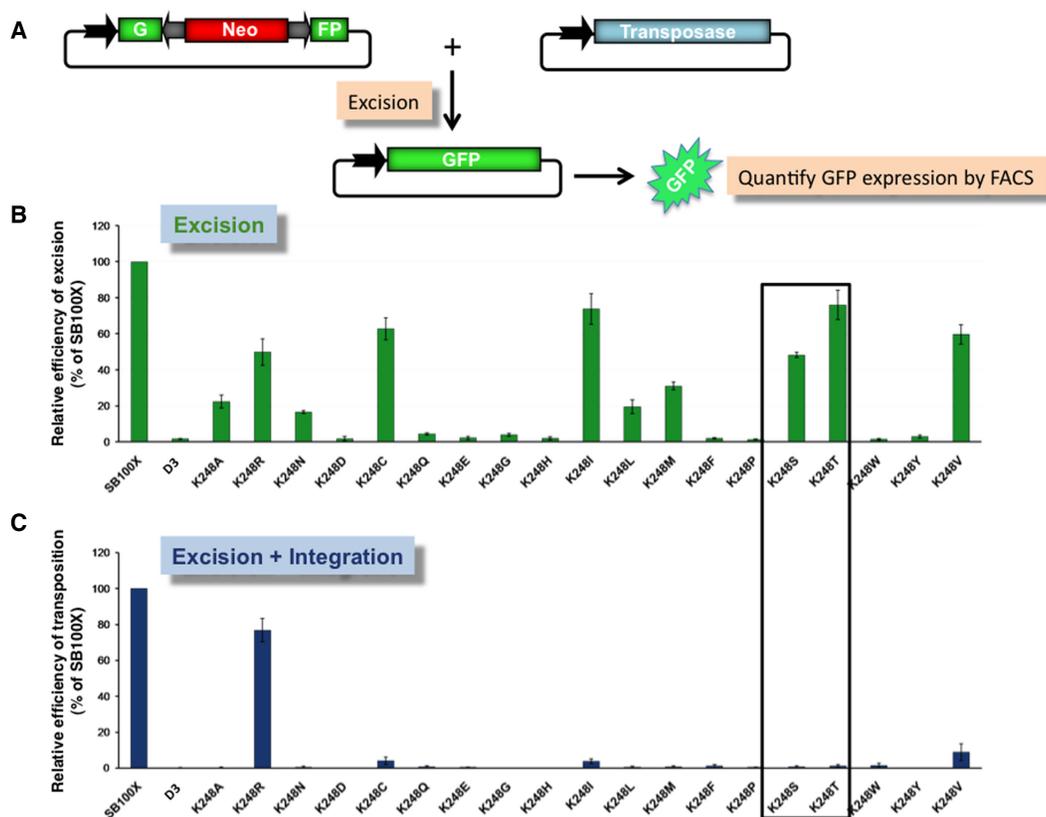


Figure 2. Excision and integration activities of *Sleeping Beauty* transposase mutants. (A) Transposon excision assay. A genetically tagged SB transposon disrupts the GFP coding sequence maintained on a plasmid. Cells transfected with this construct do not express GFP. In the presence of SB transposase excision occurs, and in a fraction of the products the GFP coding sequence is restored, thereby leading to green fluorescence that can be quantified. (B) Relative excision efficiencies. Plasmids expressing transposase mutants were transiently cotransfected with a transposon-donor plasmid (pCMV(CAT)-GFP//T2Neo) into HeLa cells. The frequency of excision is indicated by GFP fluorescence intensity, determined by FACS analysis and normalized to SB100X. An inactive SB transposase (D3) was included as negative control. Data are represented as mean \pm SD, $n = 3$ biological replicates. Differences in excision activities are significant as determined by Student's *t*-test for K248T $P = 0.018$, for all other mutants $P < 0.001$. (C) Relative transposition efficiencies. Plasmids expressing transposase mutants were transiently cotransfected with a transposon-donor plasmid (pT2B/puro) into HeLa cells. Cells were selected for puromycin resistance and stained with methylene blue to identify viable cell colonies. The frequency of integration is indicated by the number of puromycin resistant colonies and was normalized to SB100X; D3 was included as negative control. Data are represented as mean \pm SD, $n = 4$ biological replicates. Differences in transposition activities are significant as determined by Student's *t*-test for K248R $P = 0.012$, for all other mutants $P < 0.01$.

Both of the transposon excision assays described above are suitable for quantitative measurement of transposon excision, but they can only score events associated with canonical excision and thus restoration of the GFP or puro coding sequences. It is conceivable that the K248S and K248T mutants have lower scorable excision activities than SB100X in our assays, because a sizeable fraction of excision events are imprecise; these excision events would be incompatible with both marker expression and subsequent transposon integration. Thus, in order to gain a deeper qualitative insight into the excision reaction catalyzed by the SB transposase mutants K248S and K248T, we applied a PCR-based assay, which is suitable to isolate, amplify and sequence the transposon excision sites following transient transfection of transposon donor and transposase expression plasmids into cultured cells (22) (Figure 5A). We randomly picked and sequenced 25–30 products representing transposon excision sites generated by SB100X (positive control) and the K248S and K248T mutants. Consistent with previous observations for SB excision in human HeLa cells (79,80),

sequencing of the PCR products revealed some DNA sequence heterogeneity at the transposon excision sites generated by SB100X; alongside with canonical footprints, products displaying deletions at the footprint were also observed (Figure 5B and Supplementary Figure S4), consistent with error-prone NHEJ repair at the ends of the broken DNA following transposon excision (22,79). Excision footprints generated by the K248S and K248T mutants were indistinguishable from those generated by SB100X (Figure 5B and Supplementary Figure S4). In addition, we addressed if the K248S and K248T mutants maintain fidelity of footprint formation following excision from a genomic context. Transgenic HeLa and HepG2 cells containing a single copy of the SB transposon were transfected with transposase expression plasmids, and genomic DNA subjected to PCR with primers flanking the transposon. Sequencing of the PCR products revealed indistinguishable footprints generated by SB100X, K248S and K248T displaying the characteristic, canonical junctions of donor DNA following transposon excision (Supplementary Figure S5). We conclude

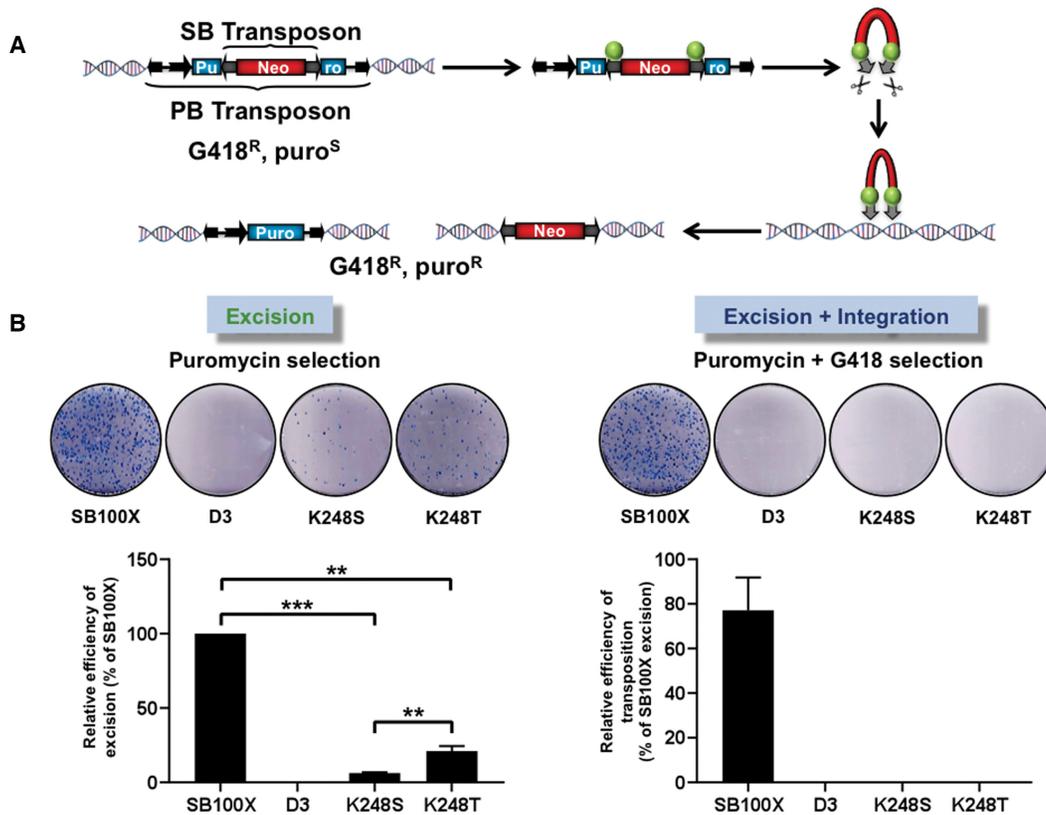


Figure 3. Mobilization of a chromosomally integrated *Sleeping Beauty* transposon. (A) The HeLa-derived reporter cell line contains a single copy of a *piggyBac* (PB) transposon carrying a puro antibiotic marker, whose coding sequence is disrupted by an SB transposon carrying a neo expression cassette. These cells are thus resistant to G418 and sensitive to puromycin. Expression of SB transposases in these cells results in mobilization of the SB transposon out of the puro marker, thereby reconstituting the puro ORF and resulting in puro resistance. In case the excised transposon does not reintegrate into the genome the resulting cell will be puro-resistant but G418-sensitive. In case the excised SB transposon reintegrates somewhere else in the genome the resulting cell will be resistant to both puro and G418. (B) Relative efficiencies of the excision and integration activities of the K248S and K248T mutants. Positive and negative controls were as described in Figure 2. Data are represented as mean \pm SD, $n = 3$ biological replicates. Asterisks indicate significant differences as determined by Student's *t*-test ** $P < 0.001$; *** $P < 0.0001$.

that excision and footprint formation by the mutants qualitatively follow the canonical pathway of SB transposition.

Transposons excised by the K248S and K248T *Sleeping Beauty* transposase mutants generate extrachromosomal circles

To investigate the fate of transposon molecules excised by K248S and K248T transposases, a circle formation assay was performed. Earlier work suggested that transposon excision by the K248A variant generates circular molecules (59). The ability of the K248S and K248T transposase mutants to give rise to circular molecules after excision was determined using a donor plasmid (pT/zeo) that contains a transposon carrying an origin of replication and a zeocin antibiotic resistance gene (zeo) (70). Following transient transfection of the transposon/transposase components into HeLa cells, non-integrated/circularized transposon molecules can be recovered from low molecular weight DNA preparation, and transformation into *E. coli* followed by zeocin selection (Figure 6A). Assuming canonical transposon excision, the sequences of the transposon end junc-

tions resulting from ligation (circularization) can be predicted (Figure 6B), and the circles can be detected by PCR.

Figure 6C demonstrates that, unlike SB100X, the mutant transposases K248S and K248T frequently generate circular transposon molecules upon excision. Analysis of the transposon circle junction sequences generated by the transposase mutants K248S and K248T revealed two classes of products. Some junctions correspond to simple religation of transposon ends following canonical, precise transposon excision, whereas other junctions contained five additional nucleotides (TACAG from the transposon end), which cannot easily be reconciled with a canonical excision/reintegration process (Figure 6D). Regardless of the pathway that generates these circles, we found that they are unable to undergo transposition by the SB100X transposase when co-transfected into HeLa cells (Figure 6E), suggesting that they represent dead-end products of SB transposition.

Generation of transgene-free induced pluripotent stem cells by cassette removal with K248T

To explore the applicability of the most effective transposase mutant K248T in transient transgenesis, we de-

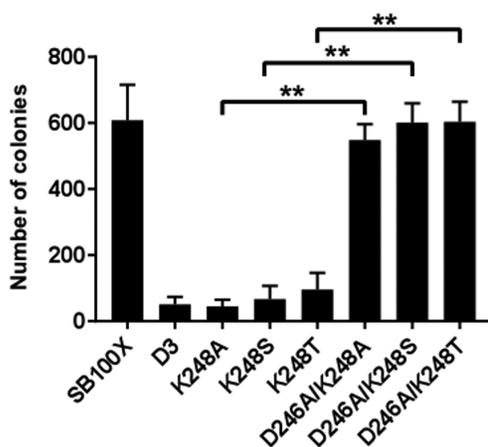


Figure 4. The D246A replacement rescues transposon integration activity of K248 mutants. Plasmids expressing the indicated single- and double mutants of the SB100X transposase along with SB100X (positive control) and catalytically inactive SB transposase (D3, negative control) were transiently cotransfected with a puromycin-tagged transposon donor plasmid into HeLa cells. The frequency of integration is indicated by the number of puromycin-resistant colonies. Data are represented as mean \pm SD, $n = 3$ biological replicates. Asterisks indicate significant differences as determined by Student's t -test $^{***}P < 0.001$.

employed a previously characterized mouse iPSC line containing a single SB transposon reprogramming vector insertion on chrX (26). The reprogramming vector contains an OSKML (Oct4, Sox2, Klf4, c-Myc and Lin28 reprogramming factors) expression cassette driven by the CAG promoter and includes a pu Δ tk positive/negative selection cassette (Figure 7A). The pu Δ tk cassette allows positive selection due to resistance to puromycin and negative selection based on sensitivity to 1-(2-deoxy-2-fluoro-1- β -D-arabino-furanosyl)-5-iodouracil (FIAU). This iPSC line has been generated in the transgenic OG2 background, which contains a GFP transgene driven by the Oct4 promoter (81). Thus, iPSCs cells derived from OG2 MEFs express GFP due to the activation of the Oct4 promoter in pluripotent stem cells.

Mouse iPSCs were transiently transfected with the SB transposase K248T and subjected to FIAU selection to sort for OSKML cassette loss. A diagnostic splinkerette PCR on genomic DNA prepared from a FIAU-resistant iPSC clone revealed loss of the transposon in the mouse genome (Figure 7B). These 'cassette-removed' iPSCs expressed GFP from the Oct4 promoter and expressed the endogenous pluripotency marker nanog (Figure 7C), thereby indicating that the K248T transposase allows removal of reprogramming genes in iPSCs and that these cells maintain pluripotency even in the absence of external pluripotency factors.

DISCUSSION

Here we conducted a mutagenesis screen in the SB transposase, and identified two novel mutants, K248S and K248T, that exhibit an exc⁺/int⁻ phenotype. In the transposition reactions catalyzed by these transposase variants the excision and integration functions are clearly disconnected,

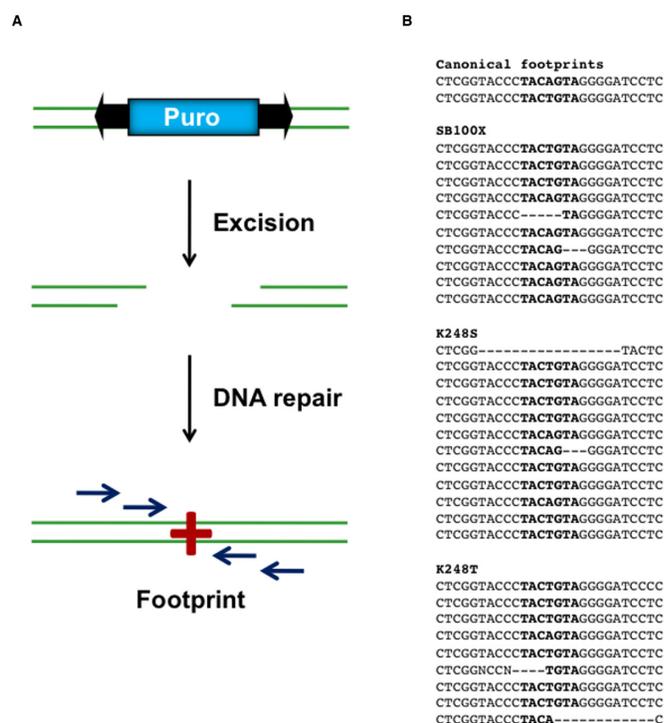


Figure 5. Analysis of excision footprints by plasmid-based excision assay. (A) Schematic of the plasmid-based excision assay. HeLa cells were transfected with a puromycin-marked (Puro) donor plasmid along with a transposase-encoding expression plasmid. The gap in the donor-plasmid left by the transposon is re-ligated and leaves a footprint at the site of the initial transposon. Two days post-transfection, plasmid DNA was isolated from cells, and analyzed in nested PCR reactions to determine the footprints of the excision reactions. (B) Excision footprints generated by SB100X, K248S and K248T. Footprints are shown in bold.

displaying high efficiency transposon excision with no detectable level of re-integration in human cells.

Regions within the catalytic core domain of retroviral integrases and the Tn10 transposase have been shown to be involved in mediating interactions with the target DNA (82–84), and are thus critically important for genomic integration. K248 is situated directly C-terminal to the D244 catalytic residue (the second D in the DDE triad) in the SB transposase (Figure 1C). This region is analogous to the α 2 helix in HIV integrase, which was shown to be required for retroviral integration (82,83). In addition, the proximity of K248 to target DNA in the SB TCC model (Figure 1C) suggests that the defect underlying the int⁻ phenotype is related to a reduced ability of the transposase to either interact with the target DNA or perform strand transfer. Notably, the previously reported exc⁺/int⁻ PB mutants were also replacements of conserved, positively charged amino acids (lysines and arginines) in the catalytic core domain of the transposase (58), and exc⁺/int⁻ Tn10 transposase mutants were shown to be defective in target DNA capture (84).

Amino acid sequence alignments revealed a distinct group of Tc1 family transposases that do not contain the characteristic D246/K248 amino acid dyad (Figure 1A). We have shown that a D246A replacement can fully rescue in-

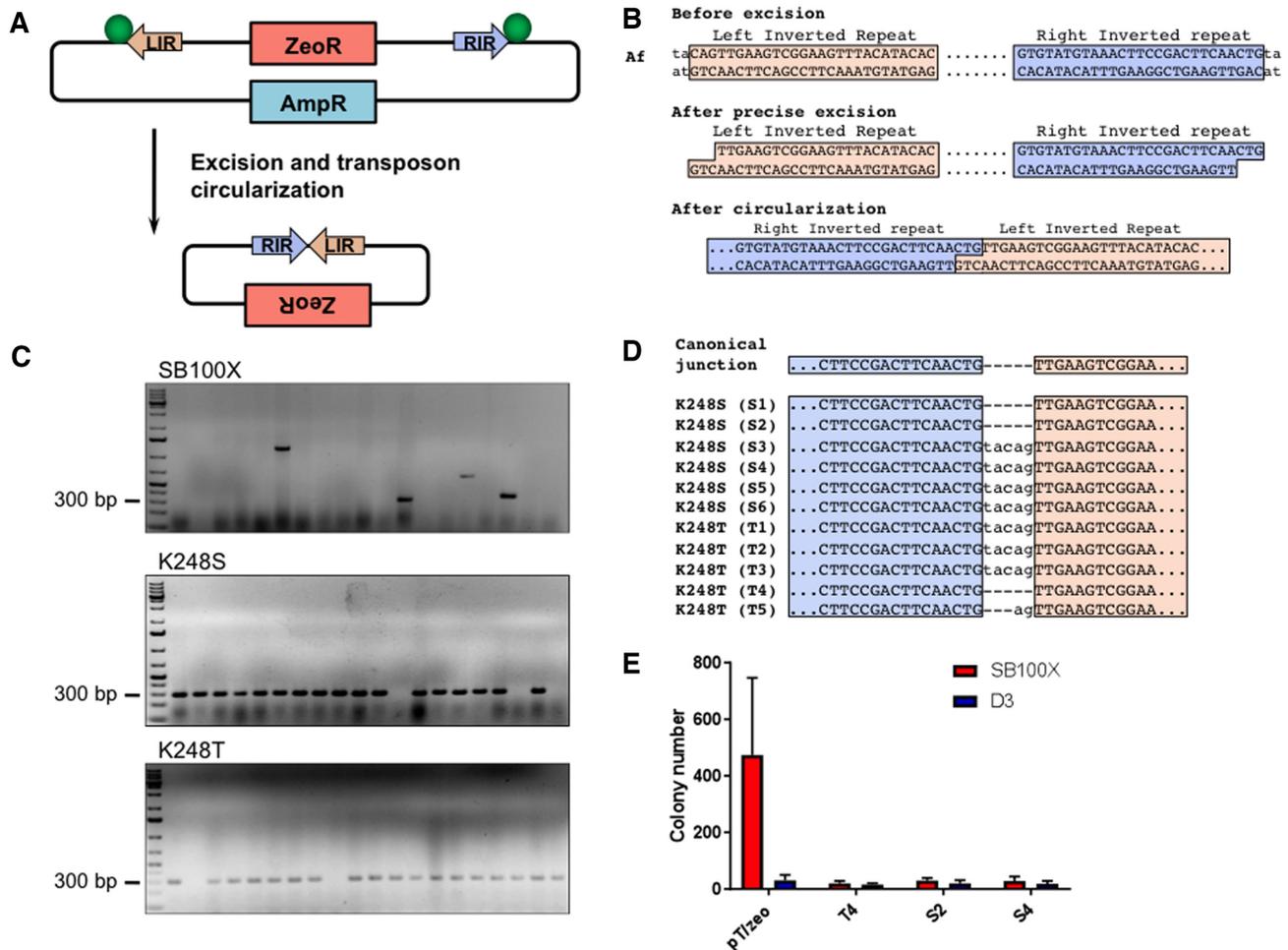


Figure 6. The K248S and K248T *Sleeping Beauty* transposase mutants generate transposon circles. (A) SB transposon circle donor plasmid (pT/zeo) and SB transposase-generated transposon circle. The transposase (green circle) binds to the left and right TIRs (LIR and RIR), and catalyzes the excision of the transposable element from the donor plasmid. DNA breaks at the donor site are repaired by the host DNA repair machinery. (B) Expected junction sequence after precise transposon excision and circularization of the donor plasmid. The central T:T mismatch is expected to be resolved by mismatch repair to yield a CTG or CAG sequence. (C) PCR-based circle formation assay. HeLa cells were transfected with a transposase expression plasmid (SB100X, K248S, K248T) along with the transposon donor plasmid pT/zeo. Circularized excised transposon molecules were recovered from low molecular weight DNA preparation followed by transformation into *E. coli* and zeocin selection. Twenty individual *E. coli* colonies from each transposase group were subjected to PCR to detect a circularized transposon, revealed by a ~300-bp band. The positions of the 300-bp size marker are indicated on the left. (D) Analysis of transposon circle junction sequences. Sequences of independent transposon circle junctions generated by K248S or K248T. On top, the expected junction sequence formed after canonical transposon excision and religation is shown. (E) Transposon circles do not support SB transposition. Some of the products generated by K248S and K248T and pT/zeo (positive control) were re-introduced into HeLa cells together with SB100X or with D3 (negative control) and zeocin-resistant cell colonies were counted. Data are represented as mean \pm SEM, $n = 3$ biological replicates.

tegration activity of K248S and K248T (Figure 4). Because the D246A mutation in itself is detrimental to SB transposition (59), these results establish an epistatic relationship between amino acid residues in positions 246 and 248 during transposition. All transposase sequences that contain K in position 248 have D in position 246. K248 is positioned near D246 and both residues are located on the target DNA binding surface of SB in our TCC model. Thus, it is tempting to speculate that in D246/K248-containing transposases, the positively charged side chain of K248 shields the negative charge of D246 to prevent charge repulsion with the target DNA. Removal of K248 would expose the negative charge of D246, largely compromising target DNA binding. Simultaneous mutation of K248 and D246 would negate the charge conflict and could rescue transposition.

This is consistent with the conservation of A246 in transposases that lack K248 and rescue of transposition activity in SB double mutants.

Excision by the K248S and K248T SB transposases leads to production of sealed transposon circles (Figure 6C), which cannot proceed to integration (Figure 6E), stalling the transposition reaction after excision. Intriguingly, sequence analysis of transposon circles suggests an alternative mode of excision by the K248S and K248T SB transposases by a mechanism, in which initially exposed 3'-OHs attack the complementary strands to form hairpins on the transposon ends with concomitant release from the donor DNA. This alternative pathway would result in canonical footprints at donor sites, whereas hairpin resolution at a shifted position can explain the extended junction sequences (Sup-

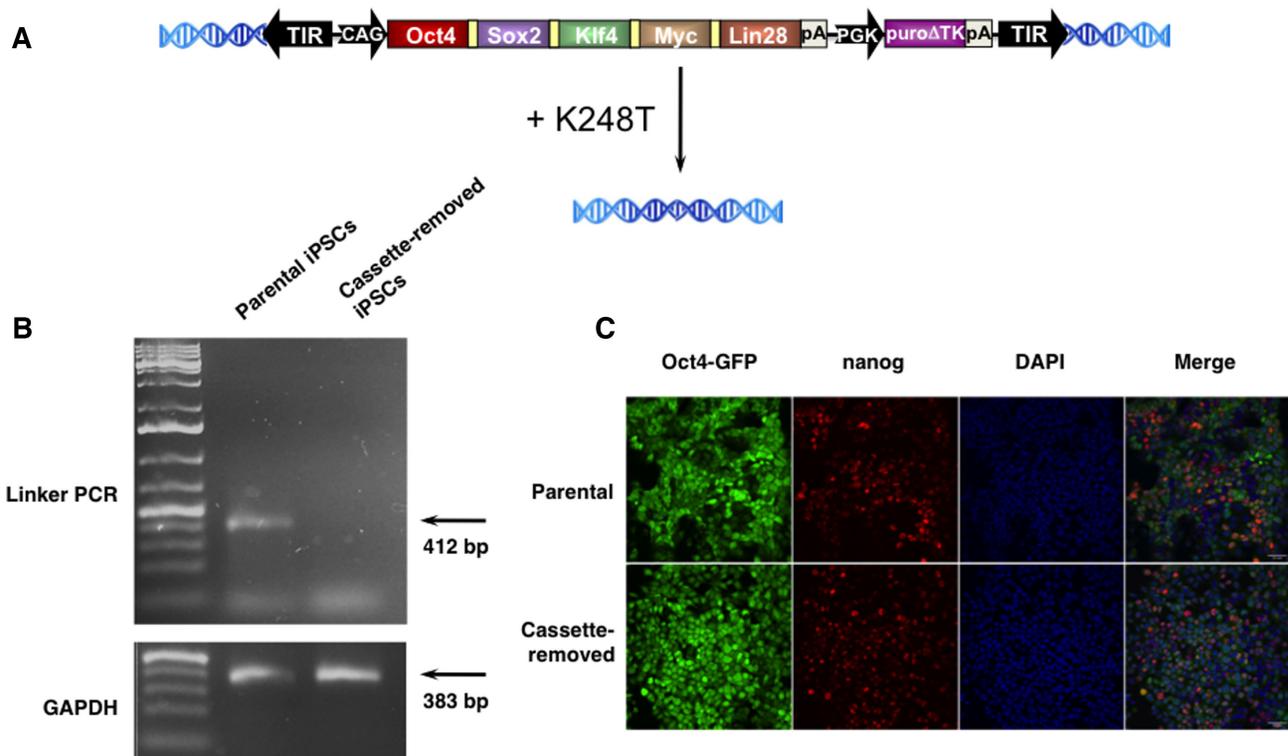


Figure 7. Reprogramming cassette removal with K248T in mouse induced pluripotent stem cells. (A) Transient transgenesis with an SB vector containing reprogramming genes and their removal with transient expression of the K248T SB transposase. Schematic representation of the SB transposon-based reprogramming vector pT2-CAG.OSKML-pu Δ tk (26). Individual genes in five-factor cassettes were linked by 2A self-cleaving peptide sequence and expressed from the CAG promoter. Black arrows represent SB transposon TIRs; pA, bovine growth hormone polyadenylation signal; pu Δ tk, PGK promoter-driven pu Δ tk expression cassette. The OSKML factors plus selection genes were removed from a single-copy mouse iPSC clone with the help of the SB100X transposase mutant K248T. (B) Linker-mediated PCR genotyping of a mouse iPSC clone before and after cassette removal. The lack of a PCR product in lane 2 of the agarose gel indicates that the reprogramming cassette was successfully removed with K248T (top gel). The quality of the genomic DNAs was verified by control PCR using GAPDH primers (bottom gel). (C) Phenotyping of a reprogramming factor-free mouse iPSC clone. Reprogramming factor-free mouse iPSCs express the endogenous nanog pluripotency marker (immunostaining) and GFP from the endogenous Oct4 promoter. Scale bar: 50 μ m.

plementary Figure S6). Importantly, it has recently been shown that all the chemical steps of *mariner* transposition are executed by a single transposase dimer, in which one monomer performs two sequential strand cleavage and one strand transfer reactions at the same transposon end (85), thereby providing strong evidence that a single DDE/D active site can hydrolyze DNA strands of opposite polarity. The *mariner* transposase cleaves the non-transferred strand first (21), and we infer that the first cleavage event during wild-type SB transposition also occurs at the non-transferred strand of the SB transposon. It is conceivable that the K248S and K248T mutations convert the active site into one that cleaves only one strand, and a hairpin on the transposon end generated by K248S and K248T implies a nick at the transferred strand (Supplementary Figure S6), just like in Tn5, Tn10 and *piggyBac* transposition (12–14). Alternatively, the footprints of transposon excision (Figure 5B) and circularization (Figure 6D) may be explained with altered cleavage position by the K248S and K248T mutants, but the exact sites are difficult to predict in the absence of a robust *in vitro* excision reaction.

Circularization of excised DNA without subsequent genomic integration is reminiscent of V(D)J recombination in the immune system of jawed vertebrates [reviewed in (86)].

In this process, that occurs during lymphocyte development, pre-existing V (variable), D (diversity), and J (joining) gene segments are rearranged to generate a large repertoire of T-cell surface receptor (TCR) and immunoglobulin molecules necessary for the recognition of diverse pathogens. The recombination event involves *cis*-acting sequences known as recombination signal sequences (RSSs) (functional equivalents of transposon TIRs) that flank each receptor gene segment and requires two proteins encoded by the recombination-activating genes, RAG1 and RAG2 (functional equivalents of a transposase). The V(D)J recombination reaction is subdivided into two stages, a cleavage phase and a joining phase [reviewed in (86)]. The complex formed by the RAG1 and RAG2 proteins introduces DSBs in the DNA between the RSS and the neighboring coding DNA *via* a nick-hairpin mechanism (Supplementary Figure S7). This mechanism shares significant similarities with the excision step of the cut-and-paste transposition process (87), with the fundamental difference that the DNA sequence elements excised by RAG1/RAG2 do not undergo genomic integration. Repair factors of the NHEJ pathway join the two signal ends resulting in circular, extrachromosomal circles (signal joints), which are lost from the cell (Supplementary Figure S7). Both of the RAG proteins have

been proposed to originate from a transposon (88,89), and are thought to have undergone a ‘domestication’ process, during which the ancestral transposon/transposase components have acquired regulatory components suppressing genomic integration of excised DNA (90). Remarkably, the DNA transposition reactions catalyzed by our new K248S and K248T SB transposases resemble a V(D)J recombination-like process, with excision resulting in sealed transposon circles and no re-integration (Supplementary Figure S7). Thus, it is tempting to speculate that one of the early evolutionary adaptations that gave rise to the domesticated RAG1 recombinase might have been amino acid replacement mutations, similar to the ones that we report here for the SB transposase, in the ancestral RAG transposase some 500 million years ago, resulting in an $\text{exc}^+/\text{int}^-$ phenotype, which has undergone fixation and has been under selection ever since. Indeed, recent evidence highlights the importance of an M848R replacement in an ancestor of RAG1 that, together with an acidic region in RAG2, dramatically suppresses RAG-mediated transposition (91). Importantly, an inverse (R848M) mutation converts RAG1 into an active transposase by stimulating the transposition reaction at a post-cleavage step (91).

RAG1 is not the only example for a transposase-derived protein, whose domestication likely involved the loss or severe repression of integration after transposon excision. *Oxytricha trifallax*, a unicellular, ciliated protozoan, has two types of nuclei in the same cytoplasm. Diploid micronuclei are transcriptionally silent during vegetative growth but transmit the germline genome through sexual conjugation. Macronuclei, on the other hand, govern somatic gene expression but undergo degradation, so that the new macronuclei develop from germline micronuclei. Macronuclear development in *Oxytricha* involves deletion of thousands of copies of germline transposons (92), designated telomere-bearing elements (TBEs) (93). These DNA sequences are thus restricted to the micronucleus and absent in the macronucleus. TBEs encode a transposase characterized by a DDE catalytic motif (94). It has been shown that the TBE transposase generates transposon circles upon excision, suggesting that circle formation and suppression of genomic re-integration might be linked (95), and that the TBE transposase has been domesticated to serve as an ‘excisase’ during macronuclear development in *Oxytricha*.

A third example for a DNA recombination reaction where the production of episomal circles have been observed is retroviral integration [reviewed in (96)]. Some of the linear double-stranded DNA resulting from reverse transcription during the retroviral- and lentiviral life cycle is converted into circular episomes by cellular DSB repair factors. During this process, NHEJ generates 2-LTR circles, whereas homologous recombination (HR) generates 1-LTR circles, both containing the entire viral genome and either two or one LTR motifs, respectively. The formation of retroviral circles negatively correlates with the ability of the virus to integrate into the host genome. For example, mutations in the retroviral integrase DDE residues (such as D64V in the case of lentiviral vectors) that block proviral integration result in enhanced levels of retroviral circles. This observation served as a basis for the development of integration-deficient lentiviral vectors (IDLVs) that are typically used

for the generation of circular vector episomes in transduced cells for various genetic engineering applications [reviewed in (96,97)].

The wild-type, fully competent SB transposase can also generate transposon circles, albeit at a rather low frequency (Figure 6C), implying that transposon integration and circle formation are competitive processes. Indeed, circularization of DNA following excision by the K248S and K248T SB transposase mutants, during V(D)J recombination and TBE transposon excision in the *Oxytricha* macronucleus and in IDLVs collectively point to a general rule, in which in the absence of genomic integration DNA repair factors gain access to the ends of extrachromosomal, linear fragments of DNA, thereby leading to the formation of circles. In the case of SB transposition it is clear that these circles cannot be substrates for a new round of cut-and-paste transposition (Figure 6E), thus representing dead-end side products of the transposition reaction.

We demonstrated the utility of the SB transposase mutants in ‘transient transgenesis’ for the generation of reprogramming factor-free iPSCs (Figure 7). The key attributes of these mutants that match the requirements of applications as genetic tools include efficient excision (Figure 2) and canonical footprint formation (Figure 4). Although the amino acid substitutions in our mutants have a clearly negative impact on the efficiency of excision (Figures 2B and 3B), we note that these mutants have been generated in the SB100X hyperactive transposase background. Thus, the net efficiency of excision by our mutants still enables efficient generation of ‘transgene-free’ cells, especially in experimental situations where clonal cell propagation is possible (e.g. in iPSC culture, as demonstrated here). The benefit of applying ‘transient transgenesis’ by these SB variants might be best realized in genetic backgrounds that limit efficient reprogramming by non-integrating technologies, including fibroblasts from patients with ataxia telangiectasia (98,99). Finally, because the K248 residue of the SB transposase is proposed to be involved in interaction with the target DNA, some of the integration-proficient mutants that we generated in our screen (Figure 2) might display alterations in genome-wide target site selection properties, thereby providing a useful resource for future structure-function investigations of SB transposition.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Lacramioara Botezatu and Dan Shen for technical support.

FUNDING

EU FP7 InduStem (230675 to A.D. and Z.I.); institutional funds from the Paul Ehrlich Institute and the European Molecular Biology Laboratory (EMBL); EMBL International PhD Programme (fellowship to I.Q). Funding for open access charge: Institutional funds.

Conflict of interest statement. None declared.

REFERENCES

- Kulkosky, J., Jones, K.S., Katz, R.A., Mack, J.P. and Skalka, A.M. (1992) Residues critical for retroviral integrative recombination in a region that is highly conserved among retroviral/retrotransposon integrases and bacterial insertion sequence transposases. *Mol. Cell Biol.*, **12**, 2331–2338.
- Doak, T.G., Doerder, F.P., Jahn, C.L. and Herrick, G. (1994) A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common “D35E” motif. *Proc. Natl. Acad. Sci. U.S.A.*, **91**, 942–946.
- Kim, D.R., Dai, Y., Mundy, C.L., Yang, W. and Oettinger, M.A. (1999) Mutations of acidic residues in RAG1 define the active site of the V(D)J recombinase. *Genes Dev.*, **13**, 3070–3080.
- Landree, M.A., Wibbenmeyer, J.A. and Roth, D.B. (1999) Mutational analysis of RAG1 and RAG2 identifies three catalytic amino acids in RAG1 critical for both cleavage steps of V(D)J recombination. *Genes Dev.*, **13**, 3059–3069.
- Bujacz, G., Alexandratos, J., Wlodawer, A., Merkel, G., Andrade, M., Katz, R.A. and Skalka, A.M. (1997) Binding of different divalent cations to the active site of avian sarcoma virus integrase and their effects on enzymatic activity. *J. Biol. Chem.*, **272**, 18161–18168.
- Goldgur, Y., Dyda, F., Hickman, A.B., Jenkins, T.M., Craigie, R. and Davies, D.R. (1998) Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium. *PNAS*, **95**, 9150–9154.
- Mizuuchi, K. (1992) Polynucleotidyl transfer reactions in transpositional DNA recombination. *J. Biol. Chem.*, **267**, 21273–21276.
- Craig, N.L. (1995) Unity in transposition reactions. *Science*, **270**, 253–254.
- Turlan, C. and Chandler, M. (2000) Playing second fiddle: second-strand processing and liberation of transposable elements from donor DNA. *Trends Microbiol.*, **8**, 268–274.
- Hickman, A.B., Chandler, M. and Dyda, F. (2010) Integrating prokaryotes and eukaryotes: DNA transposases in light of structure. *Crit. Rev. Biochem. Mol. Biol.*, **45**, 50–69.
- Hickman, A.B. and Dyda, F. (2015) Mechanisms of DNA transposition. *Microbiol. Spectrum*, **3**, MDNA3–0034–2014.
- Mitra, R., Fain-Thornton, J. and Craig, N.L. (2008) piggyBac can bypass DNA synthesis during cut and paste transposition. *EMBO J.*, **27**, 1097–1109.
- Bhasin, A., Goryshin, I.Y. and Reznikoff, W.S. (1999) Hairpin formation in Tn5 transposition. *J. Biol. Chem.*, **274**, 37021–37029.
- Kennedy, A.K., Guhathakurta, A., Kleckner, N. and Haniford, D.B. (1998) Tn10 transposition via a DNA hairpin intermediate. *Cell*, **95**, 125–134.
- Bischerour, J. and Chalmers, R. (2009) Base flipping in tn10 transposition: an active flip and capture mechanism. *PLoS One*, **4**, e6201.
- Bischerour, J., Lu, C., Roth, D.B. and Chalmers, R. (2009) Base flipping in V(D)J recombination: insights into the mechanism of hairpin formation, the 12/23 rule, and the coordination of double-strand breaks. *Mol. Cell Biol.*, **29**, 5889–5899.
- Zhou, L., Mitra, R., Atkinson, P.W., Hickman, A.B., Dyda, F. and Craig, N.L. (2004) Transposition of hAT elements links transposable elements and V(D)J recombination. *Nature*, **432**, 995–1001.
- Hencken, C.G., Li, X. and Craig, N.L. (2012) Functional characterization of an active Rag-like transposase. *Nat. Struct. Mol. Biol.*, **19**, 834–836.
- Gellert, M. (2002) V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu. Rev. Biochem.*, **71**, 101–132.
- van Gent, D.C., Mizuuchi, K. and Gellert, M. (1996) Similarities between initiation of V(D)J recombination and retroviral integration. *Science*, **271**, 1592–1594.
- Dawson, A. and Finnegan, D.J. (2003) Excision of the Drosophila mariner transposon Mos1. Comparison with bacterial transposition and V(D)J recombination. *Mol. Cell*, **11**, 225–235.
- Izsvak, Z., Stuwe, E.E., Fiedler, D., Katzer, A., Jeggo, P.A. and Izsvak, Z. (2004) Healing the wounds inflicted by sleeping beauty transposition by double-strand break repair in mammalian somatic cells. *Mol. Cell*, **13**, 279–290.
- Richardson, J.M., Dawson, A., O’Hagan, N., Taylor, P., Finnegan, D.J. and Walkinshaw, M.D. (2006) Mechanism of Mos1 transposition: insights from structural analysis. *EMBO J.*, **25**, 1324–1334.
- Ivics, Z., Hackett, P.B., Plasterk, R.H. and Izsvak, Z. (1997) Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell*, **91**, 501–510.
- Ivics, Z., Li, M.A., Mates, L., Boeke, J.D., Nagy, A., Bradley, A. and Izsvak, Z. (2009) Transposon-mediated genome manipulation in vertebrates. *Nat. Methods*, **6**, 415–422.
- Grabundzija, I., Wang, J., Sebe, A., Erdei, Z., Kajdi, R., Devaraj, A., Steinemann, D., Szuhai, K., Stein, U., Cantz, T. et al. (2013) Sleeping Beauty transposon-based system for cellular reprogramming and targeted gene insertion in induced pluripotent stem cells. *Nucleic Acids Res.*, **41**, 1829–1847.
- Kues, W.A., Herrmann, D., Barg-Kues, B., Haridoss, S., Nowak-Imialek, M., Buchholz, T., Streeck, M., Grebe, A., Grabundzija, I., Merkert, S. et al. (2013) Derivation and characterization of sleeping beauty transposon-mediated porcine induced pluripotent stem cells. *Stem Cells Dev.*, **22**, 124–135.
- Talluri, T.R., Kumar, D., Glage, S., Garrels, W., Ivics, Z., Debowski, K., Behr, R. and Kues, W.A. (2014) Non-viral reprogramming of fibroblasts into induced pluripotent stem cells by Sleeping Beauty and piggyBac transposons. *Biochem. Biophys. Res. Commun.*, **450**, 581–587.
- Talluri, T.R., Kumar, D., Glage, S., Garrels, W., Ivics, Z., Debowski, K., Behr, R., Niemann, H. and Kues, W.A. (2015) Derivation and characterization of bovine induced pluripotent stem cells by transposon-mediated reprogramming. *Cell Reprogramm.*, **17**, 131–140.
- Sebe, A. and Ivics, Z. (2016) Reprogramming of human fibroblasts to induced pluripotent stem cells with sleeping beauty transposon-based stable gene delivery. *Methods Mol. Biol.*, **1400**, 419–427.
- Molina-Estevéz, F.J., Lozano, M.L., Navarro, S., Torres, Y., Grabundzija, I., Ivics, Z., Samper, E., Bueren, J.A. and Guenechea, G. (2013) Brief report: impaired cell reprogramming in nonhomologous end joining deficient cells. *Stem Cells*, **31**, 1726–1730.
- Copeland, N.G. and Jenkins, N.A. (2010) Harnessing transposons for cancer gene discovery. *Nat. Rev. Cancer*, **10**, 696–706.
- DeNicola, G.M., Karreth, F.A., Adams, D.J. and Wong, C.C. (2015) The utility of transposon mutagenesis for cancer studies in the era of genome editing. *Genome Biol.*, **16**, 229.
- Kawakami, K., Largaespa, D.A. and Ivics, Z. (2017) Transposons as tools for functional genomics in vertebrate models. *Trends Genet.*, **33**, 784–801.
- Ivics, Z., Garrels, W., Mates, L., Yau, T.Y., Bashir, S., Zidek, V., Landa, V., Geurts, A., Pravenec, M., Rulicke, T. et al. (2014) Germline transgenesis in pigs by cytoplasmic microinjection of Sleeping Beauty transposons. *Nat. Protoc.*, **9**, 810–827.
- Ivics, Z., Hiripi, L., Hoffmann, O.I., Mátés, L., Yau, T.Y., Bashir, S., Zidek, V., Landa, V., Geurts, A., Pravenec, M. et al. (2014) Germline transgenesis in rabbits by pronuclear microinjection of Sleeping Beauty transposons. *Nat. Protoc.*, **9**, 794–809.
- Ivics, Z., Mates, L., Yau, T.Y., Landa, V., Zidek, V., Bashir, S., Hoffmann, O.I., Hiripi, L., Garrels, W., Kues, W.A. et al. (2014) Germline transgenesis in rodents by pronuclear microinjection of Sleeping Beauty transposons. *Nat. Protoc.*, **9**, 773–793.
- Alessio, A.P., Fili, A.E., Garrels, W., Forcato, D.O., Olmos Nicotra, M.F., Liaudat, A.C., Bevacqua, R.J., Savy, V., Hiriart, M.I., Talluri, T.R. et al. (2016) Establishment of cell-based transposon-mediated transgenesis in cattle. *Theriogenology*, **85**, 1297–1311.
- Garrels, W., Mates, L., Holler, S., Dalda, A., Taylor, U., Petersen, B., Niemann, H., Izsvak, Z., Ivics, Z. and Kues, W.A. (2011) Germline transgenic pigs by Sleeping Beauty transposition in porcine zygotes and targeted integration in the pig genome. *PLoS One*, **6**, e23573.
- Garrels, W., Talluri, T.R., Apfelbaum, R., Carratala, Y.P., Bosch, P., Potzsch, K., Grueso, E., Ivics, Z. and Kues, W.A. (2016) One-step Multiplex Transgenesis via Sleeping Beauty Transposition in Cattle. *Sci. Rep.*, **6**, 21953.
- Katter, K., Geurts, A.M., Hoffmann, O., Mates, L., Landa, V., Hiripi, L., Moreno, C., Lazar, J., Bashir, S., Zidek, V. et al. (2013) Transposon-mediated transgenesis, transgenic rescue, and tissue-specific gene expression in rodents and rabbits. *FASEB J.*, **27**, 930–941.
- Narayanavari, S.A., Chilkunda, S.S., Ivics, Z. and Izsvak, Z. (2017) Sleeping Beauty transposition: from biology to applications. *Crit. Rev. Biochem. Mol. Biol.*, **52**, 18–44.

43. Hackett,P.B., Largaespada,D.A. and Cooper,L.J.N. (2010) A transposon and transposase system for human application. *Mol. Ther.*, **18**, 674–683.
44. Mátés,L., Chuah,M.K.L., Belay,E., Jerchow,B., Manoj,N., Acosta-Sanchez,A., Grzela,D.P., Schmitt,A., Becker,K., Matrai,J. *et al.* (2009) Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nat. Genet.*, **41**, 753–761.
45. Boehme,P., Doerner,J., Solanki,M., Jing,L., Zhang,W. and Ehrhardt,A. (2015) The sleeping beauty transposon vector system for treatment of rare genetic diseases: an unrealized hope? *Curr. Gene Ther.*, **15**, 255–265.
46. Hudecek,M. and Ivics,Z. (2018) Non-viral therapeutic cell engineering with the Sleeping Beauty transposon system. *Curr. Opin. Genet. Dev.*, **52**, 100–108.
47. Hudecek,M., Izsvak,Z., Johnen,S., Renner,M., Thumann,G. and Ivics,Z. (2017) Going non-viral: the Sleeping Beauty transposon system breaks on through to the clinical side. *Crit. Rev. Biochem. Mol. Biol.*, **52**, 355–380.
48. Di Matteo,M., Belay,E., Chuah,M.K. and Vandendriessche,T. (2012) Recent developments in transposon-mediated gene therapy. *Expert Opin. Biol. Ther.*, **12**, 841–858.
49. Schlaeger,T.M., Daheron,L., Brickler,T.R., Entwisle,S., Chan,K., Cianci,A., DeVine,A., Ettenger,A., Fitzgerald,K., Godfrey,M. *et al.* (2015) A comparison of non-integrating reprogramming methods. *Nat. Biotechnol.*, **33**, 58–63.
50. Kuehle,J., Turan,S., Cantz,T., Hoffmann,D., Suerth,J.D., Maetzig,T., Zychlinski,D., Klein,C., Steinemann,D., Baum,C. *et al.* (2014) Modified lentiviral LTRs allow Flp recombinase-mediated cassette exchange and in vivo tracing of “factor-free” induced pluripotent stem cells. *Mol. Ther.*, **22**, 919–928.
51. Woltjen,K., Michael,I.P., Mohseni,P., Desai,R., Mileikovsky,M., Hamalainen,R., Cowling,R., Wang,W., Liu,P., Gertsenstein,M. *et al.* (2009) piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature*, **458**, 766–770.
52. Yusa,K., Rad,R., Takeda,J. and Bradley,A. (2009) Generation of transgene-free induced pluripotent mouse stem cells by the piggyBac transposon. *Nat. Methods*, **6**, 363–369.
53. Luo,G., Ivics,Z., Izsvak,Z. and Bradley,A. (1998) Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells. *Proc. Natl. Acad. Sci. U.S.A.*, **95**, 10769–10773.
54. Haniford,D.B., Chelouche,A.R. and Kleckner,N. (1989) A specific class of IS10 transposase mutants are blocked for target site interactions and promote formation of an excised transposon fragment. *Cell*, **59**, 385–394.
55. Ruan,K. and Emmons,S.W. (1984) Extrachromosomal copies of transposon Tc1 in the nematode *Caenorhabditis elegans*. *PNAS*, **81**, 4018–4022.
56. van Luenen,H.G., Colloms,S.D. and Plasterk,R.H. (1993) Mobilization of quiet, endogenous Tc3 transposons of *Caenorhabditis elegans* by forced expression of Tc3 transposase. *EMBO J.*, **12**, 2513–2520.
57. Arca,B., Zabalou,S., Loukeris,T.G. and Savakis,C. (1997) Mobilization of a Minos transposon in *Drosophila melanogaster* chromosomes and chromatid repair by heteroduplex formation. *Genetics*, **145**, 267–279.
58. Li,X., Burnight,E.R., Cooney,A.L., Malani,N., Brady,T., Sander,J.D., Staber,J., Wheelan,S.J., Joung,J.K., McCray,P.B. Jr. *et al.* (2013) piggyBac transposase tools for genome engineering. *PNAS*, **110**, E2279–E2287.
59. Wang,Y., Pryputniewicz-Dobrinska,D., Nagy,E.E., Kaufman,C.D., Singh,M., Yant,S., Wang,J., Dalda,A., Kay,M.A., Ivics,Z. *et al.* (2017) Regulated complex assembly safeguards the fidelity of Sleeping Beauty transposition. *Nucleic Acids Res.*, **45**, 311–326.
60. Solovyyev,V.V. (2007) In: Balding,D., Cannings,C and Bishop,M (eds). *Handbook of Statistical Genetics*. Wiley-Interscience, p. 1616.
61. Katoh,K. and Standley,D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.*, **30**, 772–780.
62. Guindon,S., Dufayard,J.F., Lefort,V., Anisimova,M., Hordijk,W. and Gascuel,O. (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.*, **59**, 307–321.
63. Steiniger-White,M., Rayment,I. and Reznikoff,W.S. (2004) Structure/function insights into Tn5 transposition. *Curr. Opin. Struct. Biol.*, **14**, 50–57.
64. Karaca,E. and Bonvin,A.M. (2011) A multidomain flexible docking approach to deal with large conformational changes in the modeling of biomolecular complexes. *Structure*, **19**, 555–565.
65. Kelley,L.A., Mezulis,S., Yates,C.M., Wass,M.N. and Sternberg,M.J. (2015) The PyMol web portal for protein modeling, prediction and analysis. *Nat. Protoc.*, **10**, 845–858.
66. Abrusan,G., Yant,S.R., Szilagyi,A., Marsh,J.A., Mates,L., Izsvak,Z., Barabas,O. and Ivics,Z. (2016) Structural determinants of sleeping beauty transposase activity. *Mol. Ther.*, **24**, 1369–1377.
67. Voigt,F., Wiedemann,L., Zuliani,C., Querques,I., Sebe,A., Mates,L., Izsvak,Z., Ivics,Z. and Barabas,O. (2016) Sleeping Beauty transposase structure allows rational design of hyperactive variants for genetic engineering. *Nat. Commun.*, **7**, 11126.
68. van Zundert,G.C.P., Rodrigues,J., Trellet,M., Schmitz,C., Kastiris,P.L., Karaca,E., Melquiond,A.S.J., van Dijk,M., de Vries,S.J. and Bonvin,A. (2016) The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. *J. Mol. Biol.*, **428**, 720–725.
69. Grabundzija,I., Irgang,M., Mates,L., Belay,E., Matrai,J., Gogol-Doring,A., Kawakami,K., Chen,W., Ruiz,P., Chuah,M.K. *et al.* (2010) Comparative analysis of transposable element vector systems in human cells. *Mol. Ther.*, **18**, 1200–1209.
70. Izsvak,Z., Ivics,Z. and Plasterk,R.H. (2000) Sleeping Beauty, a wide host-range transposon vector for genetic transformation in vertebrates. *J. Mol. Biol.*, **302**, 93–102.
71. Davies,D.R., Goryshin,I.Y., Reznikoff,W.S. and Rayment,I. (2000) Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. *Science*, **289**, 77–85.
72. Goryshin,I.Y. and Reznikoff,W.S. (1998) Tn5 in vitro transposition. *J. Biol. Chem.*, **273**, 7367–7374.
73. van Luenen,H.G., Colloms,S.D. and Plasterk,R.H. (1994) The mechanism of transposition of Tc3 in *C. elegans*. *Cell*, **79**, 293–301.
74. Lampe,D.J., Churchill,M.E. and Robertson,H.M. (1996) A purified mariner transposase is sufficient to mediate transposition in vitro. *EMBO J.*, **15**, 5470–5479.
75. Miskey,C., Papp,B., Mates,L., Sinzelle,L., Keller,H., Izsvak,Z. and Ivics,Z. (2007) The ancient mariner sails again: transposition of the human Hsmar1 element by a reconstructed transposase and activities of the SETMAR protein on transposon ends. *Mol. Cell Biol.*, **27**, 4589–4600.
76. Dupuy,A.J., Clark,K., Carlson,C.M., Fritz,S., Davidson,A.E., Markley,K.M., Finley,K., Fletcher,C.F., Ekker,S.C., Hackett,P.B. *et al.* (2002) Mammalian germ-line transgenesis by transposition. *PNAS*, **99**, 4495–4499.
77. Horie,K., Kuroiwa,A., Ikawa,M., Okabe,M., Kondoh,G., Matsuda,Y. and Takeda,J. (2001) Efficient chromosomal transposition of a Tc1/mariner-like transposon Sleeping Beauty in mice. *PNAS*, **98**, 9191–9196.
78. Fischer,S.E., Wienholds,E. and Plasterk,R.H. (2001) Regulated transposition of a fish transposon in the mouse germ line. *PNAS*, **98**, 6759–6764.
79. Yant,S.R. and Kay,M.A. (2003) Nonhomologous-end-joining factors regulate DNA repair fidelity during Sleeping Beauty element transposition in mammalian cells. *Mol. Cell Biol.*, **23**, 8505–8518.
80. Liu,G., Aronovich,E.L., Cui,Z., Whitley,C.B. and Hackett,P.B. (2004) Excision of Sleeping Beauty transposons: parameters and applications to gene therapy. *J. Gene Med.*, **6**, 574–583.
81. Yoshimizu,T., Sugiyama,N., De Felice,M., Yeom,Y.I., Ohbo,K., Masuko,K., Obinata,M., Abe,K., Scholer,H.R. and Matsui,Y. (1999) Germline-specific expression of the Oct-4/green fluorescent protein (GFP) transgene in mice. *Dev. Growth Differ.*, **41**, 675–684.
82. Appa,R.S., Shin,C.G., Lee,P. and Chow,S.A. (2001) Role of the nonspecific DNA-binding region and alpha helices within the core domain of retroviral integrase in selecting target DNA sites for integration. *J. Biol. Chem.*, **276**, 45848–45855.
83. Harper,A.L., Sudol,M. and Katzman,M. (2003) An amino acid in the central catalytic domain of three retroviral integrases that affects target site selection in nonviral DNA. *J. Virol.*, **77**, 3838–3845.
84. Junop,M.S. and Haniford,D.B. (1997) Factors responsible for target site selection in Tn10 transposition: a role for the DDE motif in target DNA capture. *EMBO J.*, **16**, 2646–2655.

85. Claeys Bouuaert, C. and Chalmers, R. (2017) A single active site in the mariner transposase cleaves DNA strands of opposite polarity. *Nucleic Acids Res.*, **45**, 11467–11478.
86. Jones, J.M. and Gellert, M. (2004) The taming of a transposon: V(D)J recombination and the immune system. *Immunol. Rev.*, **200**, 233–248.
87. Agrawal, A., Eastman, Q.M. and Schatz, D.G. (1998) Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system. *Nature*, **394**, 744–751.
88. Kapitonov, V.V. and Koonin, E.V. (2015) Evolution of the RAG1-RAG2 locus: both proteins came from the same transposon. *Biol. Direct*, **10**, 20.
89. Huang, S., Tao, X., Yuan, S., Zhang, Y., Li, P., Beilinson, H.A., Zhang, Y., Yu, W., Pontarotti, P., Escrava, H. *et al.* (2016) Discovery of an active RAG transposon illuminates the origins of V(D)J recombination. *Cell*, **166**, 102–114.
90. Carmona, L.M. and Schatz, D.G. (2017) New insights into the evolutionary origins of the recombination-activating gene proteins and V(D)J recombination. *FEBS J.*, **284**, 1590–1605.
91. Zhang, Y., Cheng, T.C., Huang, G., Lu, Q., Surleac, M.D., Mandell, J.D., Pontarotti, P., Petrescu, A.J., Xu, A., Xiong, Y. *et al.* (2019) Transposon molecular domestication and the evolution of the RAG recombinase. *Nature*, **569**, 79–84.
92. Witherspoon, D.J., Doak, T.G., Williams, K.R., Seegmiller, A., Seger, J. and Herrick, G. (1997) Selection on the protein-coding genes of the TBE1 family of transposable elements in the ciliates *Oxytricha fallax* and *O. trifallax*. *Mol. Biol. Evol.*, **14**, 696–706.
93. Herrick, G., Cartinhour, S., Dawson, D., Ang, D., Sheets, R., Lee, A. and Williams, K. (1985) Mobile elements bounded by C4A4 telomeric repeats in *Oxytricha fallax*. *Cell*, **43**, 759–768.
94. Doak, T.G., Doerder, F.P., Jahn, C.L. and Herrick, G. (1994) A proposed superfamily of transposase genes: transposon-like elements in ciliated protozoa and a common “D35E” motif. *PNAS*, **91**, 942–946.
95. Williams, K., Doak, T.G. and Herrick, G. (1993) Developmental precise excision of *Oxytricha trifallax* telomere-bearing elements and formation of circles closed by a copy of the flanking target duplication. *EMBO J.*, **12**, 4593–4601.
96. Wanisch, K. and Yanez-Munoz, R.J. (2009) Integration-deficient lentiviral vectors: a slow coming of age. *Mol. Ther.*, **17**, 1316–1332.
97. Staunstrup, N.H. and Mikkelsen, J.G. (2011) Integrase-defective lentiviral vectors—a stage for nonviral integration machineries. *Curr. Gene Ther.*, **11**, 350–362.
98. Nayler, S., Gatei, M., Kozlov, S., Gatti, R., Mar, J.C., Wells, C.A., Lavin, M. and Wolvetang, E. (2012) Induced pluripotent stem cells from ataxia-telangiectasia recapitulate the cellular phenotype. *Stem Cells Transl. Med.*, **1**, 523–535.
99. Fukawatase, Y., Toyoda, M., Okamura, K., Nakamura, K., Nakabayashi, K., Takada, S., Yamazaki-Inoue, M., Masuda, A., Nasu, M., Hata, K. *et al.* (2014) Ataxia telangiectasia derived iPS cells show preserved x-ray sensitivity and decreased chromosomal instability. *Sci. Rep.*, **4**, 5421.