Transcriptional activities of the *Sleeping Beauty* transposon and shielding its genetic cargo with insulators

Oliver Walisko, Andrea Schorn, Frank Rolfs, Anantharam Devaraj, Csaba Miskey, Zsuzsanna Izsvák, and Zoltán Ivics

Published in final edited form as:
Molecular Therapy. 2008 Feb; 16(2): 359-369 | doi: 10.1038/sj.mt.6300366
Nature Publishing Group ►

Final Draft

Transcriptional activities of the *Sleeping Beauty* transposon and shielding its genetic cargo with insulators

Oliver Walisko[1], *Andrea Schorn*[1], *Frank Rolfs*[1], *Anantharam Devaraj*[1], *Csaba Miskey*[1], *Zsuzsanna Izsvák*[1,2], *and Zoltán Ivics*[1]

[1] Max Delbrück Center for Molecular Medicine, Berlin, Germany
[2] Institute of Biochemistry, Biological Research Center of the Hungarian Academy of Sciences, Szeged, Hungary

ABSTRACT | The *Sleeping Beauty* (SB) transposable element shows efficient transposition in human cells, and provides long-term transgene expression in preclinical animal models. Random chromosomal insertion of SB vectors represents a safety issue in human gene therapeutic applications, due to potential genotoxic effects associated with transposon integration. We investigated the transcriptional activities of SB in order to assess its potential to alter host gene expression upon integration. The untranslated regions (UTRs) of the transposon direct convergent, inward-directed transcription. Transcription from the 5'-UTR of SB is upregulated by the host-encoded factor high-mobility group 2-like 1 (HMG2L1), and requires a 65–base pair (bp) region not present in commonly used SB vectors. The SB transposase antagonizes the effect of HMG2L1, suggesting that natural transposase expression is under a negative feedback regulation. SB transposon vectors lacking the 65-bp region associated with HMG2L1-dependent upregulation exhibit benign transcriptional activities, at a level up to 100-times lower than that of the murine leukemia virus (MLV) long terminal repeat (LTR). Incorporation of chicken β-globin HS4 insulator sequences in SB-based vectors reduces the transactivation of model promoters by transposon-borne enhancers, and thus may lower the risk of transcriptional activation of host genes situated close to a transposon insertion site.

Introduction

One promising approach to the correction of genetic disorders is based on stable genomic integration of a functional copy of a gene that has a disease-causing defective allele. [1] Several vector platforms exist for the delivery of therapeutic gene constructs into cells. Virus-based vectors are widely used in gene therapy applications, [2,3,4] because they allow efficient delivery of the therapeutic gene into a target cell population. However, the use of viruses can lead to both immunological complications, as exemplified by the sudden death of a patient treated with an adenoviral vector [5] and to genotoxic effects of vector integration. [6,7] The use of integrating retroviruses carries a significant potential for insertional mutagenesis and transcriptional activation of proto-oncogenes by strong enhancer/promoter elements in the long terminal repeats (LTRs) of the virus. [8,9] In fact, all classes of endogenous LTR retroelements display both a genomic distribution and an orientation bias in mammalian genomes in that those copies that are in the same orientation as the nearest gene are significantly underrepresented in regions within 5 kb of cellular genes. [10] Thus, LTR elements may have been excluded from gene regions because of their potential to affect gene transcription. The potential of retroviral vectors to elicit a genotoxic response is further enhanced by biased *de novo* integration into transcriptional units of actively transcribed genes [11,12,13] or near the start of transcriptional units (*i.e.*, promoters). [13,14] Indeed, the development of leukaemia in patients treated for X-linked severe combined immunodeficiency is attributed to vector insertions into or near the known proto-oncogene *LMO2*, [15] which has led to a temporary interruption of gene therapy trials for the treatment of X-linked severe combined immunodeficiency. [16]

As an alternative to viral vectors, transposons can be used as non-viral vehicles for gene transfer applications (see ref. 17 for a recent review and references therein). Transposons are naturally occurring, defined DNA segments, which move from one location to another in a given genome. Transposons can be developed into gene vector systems that, similar to integrating viruses, mediate stable insertion and long-term expression of transgene constructs. In contrast to viruses, transposon-based gene transfer vectors are easier to engineer (since they are simple, plasmid-based vectors with no packaging limitations) and manufacture (as they pose no safety concerns). Thus, transposon-based gene vectors represent a non-viral approach with an integrating feature that can be an alternative to virus-based methodologies. However, since transposons lack an extracellular stage in their life-cycle (*i.e.*, they do not have a machinery for cellular entry), cellular uptake and delivery of the transposon vector represents a major challenge.

Sleeping Beauty (SB), a reconstructed transposon from fish, [18] has been widely used in gene transfer studies, and represents a milestone in the application of transposition-mediated gene delivery in vertebrates. [19] SB vectors can provide long-term gene expression *in vivo*, and there has been a steady growth in interest in applying the SB system for the treatment of a number of conditions including blood disorders, diabetes, Huntington disease, cancer, diseases of the lung and cardiovascular diseases (reviewed in refs. 17,20,21,22).

In its natural configuration, the SB transposon consists of a single gene encoding the SB transposase that catalyzes the strand-cleavage and strand-transfer reactions involved in the transposition process (Fig.1). The transposase gene is flanked by imperfect terminal inverted repeats (IRs), which contain the binding sites for the transposase.[18] These binding sites are repeated twice per IR in a direct fashion [direct repeats (DRs)], resulting in an IR/DR organization of the IRs (Fig.1). The left IR is separated from the transposase coding sequence by a 160–base pair (bp) stretch of DNA (Fig.1) with no apparent function in the transposition reaction. [23] For gene delivery purposes, SB is typically used as a two-component vector system, in which a genetic cargo is flanked by the transposon IRs, and the transposase is

supplied *in trans* (Fig.1). Transposition is catalyzed by the transposase which, upon binding to the DRs and synapsis of the transposon ends, excises the element from its donor site, and reintegrates it into a target site in a process called cut-and-paste transposition. SB exclusively integrates into TA dinucleotides that are duplicated upon transposition to yield TA target site duplications.[18] Inspection of SB insertion distribution at the genome level has revealed that SB integration occurs fairly randomly [24,25] with a small bias toward genes and upstream regulatory regions. [25] However, microarray analyses revealed no correlation between the integration profile of SB and the transcriptional status of targeted genes, [25] suggesting that SB might have a favorable safety profile as compared to widely used viral approaches.

We addressed the safety issues concerned with SB-based gene vectors by investigating transcriptional activities of the untranslated regions (UTRs) of the transposon that include the IRs. Both UTRs exhibit moderate, inward-directed promoter activity, at a level >40-fold lower than the murine leukemia virus (MLV) retrovirus LTR. Mapping in the 5'-UTR revealed an enhancer-like region that significantly contributes to the overall transcriptional activity. We identified an HMG-box-containing cellular protein [high-mobility group 2-like 1(HMG2L1)] that acts as a transcription factor at the 5'-UTR of SB. We further show that cargo gene constructs in SB transposon vectors can significantly transactivate nearby promoters, and that this can efficiently be shielded by flanking the transgene construct with chicken *β*-globin HS4 insulators inside the transposon vector. Thus, insulator elements can successfully be incorporated in the next generation of transposon vectors for enhanced safety with respect to accidental transactivation of adjacent genes.

Results

The UTRs of SB exhibit moderate promoter activity

Some of the 5'-UTRs upstream of the initiation codon of the transposase gene contain promoter motifs, [26] suggesting that they might have functions associated with control of transposition activity. However, previous studies did not reveal an internal promoter in the Tc1 element; instead they showed that the elements are transcribed by read-through transcription from *Caenorhabditis elegans* genes. [27] In contrast, the 5'-UTR of the human *Hsmar1* element was shown to have fairly robust transcriptional activity. [28] Because transposon insertion close to genes may lead to the development of cancer due to transcriptional upregulation by the transposon's intrinsic enhancer/promoter elements, investigations into the transcriptional activities associated with SB transposon sequences bear significant relevance concerning the safety of SB transposon-based vectors in gene therapy applications.

To assess the potential of the 5'-UTR (including the left IR and ~160-bp DNA of unknown function, Fig.1) of SB to drive transcription, the transposase gene was replaced by a luciferase reporter gene at the ATG start codon of the coding region, and transcriptional activities were measured in transient transfection experiments in human HeLa cells. Transcription driven by the 5'-UTR of SB (Fig.2a, lane 3) is ~18-fold higher than transcription of a promoter-less sequence (Fig.2a, lane 1), ~4.6-fold higher than transcription driven by a TATA-box minimal promoter (Fig.2a, lane 2), and about 2.5-fold higher than transcription driven by the 5'-UTR of the closely related *Frog Prince* (FP) transposon (Fig.2a, lane 7). The 5'-UTR drives expression of the SB transposase at a level sufficient to detect SB transposition in a colony-forming transposition assay in HeLa cells (Fig.2b). To test for directionality in promoter activity, the orientation of the 5'-UTR of SB was reversed relative to the luciferase gene, resulting in significant reduction of luciferase expression down to the activity of the TATA-box minimal promoter (compare lanes 2, 3, and 4 in Fig.2a). A comparison with the transcriptional activity of the thymidine kinase promoter as well as that of the MLV LTR revealed a ~4-fold and a ~44-fold more efficient reporter gene expression than that driven by the 5'-UTR of SB, respectively (compare lanes 2–5 in Fig.2c). Taken together, these findings reveal the moderate ability of the SB 5'-UTR to drive the expression of the transposase gene in natural elements, but it is apparently unable to drive expression in the opposite direction, suggesting directionality in promoter activity as can also be found for retroviral LTRs (Fig.2c, lanes 4 and 5).

In the natural arrangement of SB transposon components, the transposase coding region is followed directly by the right IR (Fig.1). Thus, the 3'-UTR practically consists of the right IR. Since the IRs of the SB transposon share a significant sequence similarity, we included the right IR of SB in the promoter analysis. As shown in Fig.2a, the right IR can drive expression towards the inside of the element, but at a lower efficiency than the 5'-UTR (compare lanes 3 and 5). In addition, similar to the 5'-UTR, the right IR appears to be unable to drive the expression towards the outside of the element (compare lanes 1 and 6 in Fig.2a).

In order to identify sequence elements in the 5'-UTR responsible for promoter activity, a series of deletion constructs was generated. As shown in Fig.2d, removal of the left IR sequence containing the transposase binding sites did not result in a significant reduction of reporter gene expression (Fig.2d, compare lanes 3, 4, and 5). By contrast, partial deletion of the ~160-bp intervening sequence resulted in a drastic reduction of promoter activity (Fig.2d, compare lanes 3, 6, and 7); thus, this region is expected to contain sequence elements that are required for initiation of transcription.

Taken together, these results suggest that the ~160-bp intervening sequence that separates the left IR from the ATG codon of the transposase coding region is primarily responsible for driving transposase expression under natural conditions. Importantly, this region is not included in the commonly used SB-derived vectors (because it is not required for transposition), and the remaining transcriptional activities associated with the transposon IRs are negligible when compared to retrovirus LTRs.

Physical interaction of the functional components of the SB transposon with HMG2L1

We reported previously that the SB transposase interacts with the DNA-binding protein HMGB1, [29] the DNA repair factor Ku [30] and with the transcriptional regulator Miz-1. [31]

In search for other cellular factors interacting with the SB transposase, a human HeLa complementary DNA (cDNA) library was screened using yeast two-hybrid technology

with the SB transposase as bait. The screen resulted in a 765-bp cDNA fragment encoding the central region of the HMG2L1 protein (Fig.3a, top panel). The recovered HMG2L1 fragment fused to a Gal4 transactivation domain (AD) together with the full-length SB transposase fused to a Gal4 DNA-binding domain (BD), efficiently activates reporter gene expression (indicative of protein–protein interaction), whereas the BD-SB transposase fusion alone and the AD-HMG2L1 fragment fusion alone do not (Fig.3a). This suggests that the human HMG2L1 protein interacts with the SB transposase through its central region. We engineered a full-length version of the HMG2L1 protein (see Materials and Methods) to test whether it retains its ability to interact with the SB transposase, using an *in vitro* pull-down experiment (Fig.3b). The transposase was fused to the maltose binding protein (MBP), and immobilized on agarose beads that were subsequently incubated with radiolabeled, full-length HMG2L1 protein. As shown in Fig.3b, a fraction of the HMG2L1 protein was retained on MBP-SB beads containing the transposase, but not on MBP control beads lacking the transposase. *In vivo* interaction of the SB transposase with hemagglutinin (HA)-tagged HMG2L1 (HMG2L1/HA) was investigated using coimmunoprecipitation with an anti-HA antibody, blotting and hybridization with an antibody against the SB transposase. SB transposase was precipitated in lysates coexpessing HMG2L1, but not in lysates expressing HA-tagged SETMAR (a transposase-derived human protein [28]) used as a control (Fig.3c). Taken together, the results establish the physical interaction of the SB transposase with HMG2L1 in cells, and suggest that this interaction may contribute to the regulation of SB transposition.

HMG2L1 is a member of the HMG-box family of transcription factors, which specifically bind their target DNA through their HMG-box domains, and regulate transcription of target genes (for review see ref. 32). Based on its predicted role in transcriptional regulation and its potential to interact with the SB transposase, we hypothesized that HMG2L1 may regulate transcription of the transposase gene. To investigate the potential physical interaction of HMG2L1 with transposon DNA, *in vivo* chromatin immunoprecipitation was used following cotransfection of cells with plasmid DNA containing the 5'-UTR of SB and a vector expressing HMG2L1-HA. After chemical cross-linking, HMG2L1-bound DNA was precipitated using anti-HA antibody coupled to agarose beads, and amplified using a diagnostic polymerase chain reaction (PCR). As shown in Fig.3d, PCR products were only recovered in the presence of HMG2L1/HA and the 5'-UTR of SB (lane 2), and were highly enriched in antibody-treated samples. These results suggest an interaction between HMG2L1 and SB transposon DNA *in vivo*.

HMG2L1 induces transcription of the transposon 5'-UTR

Next, the effect of HMG2L1 protein on luciferase reporter gene expression was assayed as described above. As shown in Fig.4a, expression of HMG2L1 upregulated transcription from the 5'-UTR of SB 10- to 15-fold, independent of the orientation of the 5'-UTR with regard to the luciferase reporter gene (lanes 3 and 8). Induction of transcription by HMG2L1 is specific to SB transposon DNA, since HMG2L1 failed to induce transcription of the promoter-less, and TATA-box minimal promoter-containing, control constructs (lanes 1 and 2 in Fig.4a). Specificity for the SB 5'-UTR is further evidenced by the finding that HMG2L1 also failed to induce transcription from the right IR in either orientation (lanes 9 and 10), as well as the 5'-UTR of the FP transposon (lane 11). To identify the region within the 5'-UTR responsible for activation by HMG2L1, the mapping constructs shown in Fig.2d (lanes 3 to 7) were tested for their ability to be transcriptionally activated by coexpression of HMG2L1 (Fig.4a, lanes 3–7). A 65-bp deletion immediately upstream of the luciferase coding region completely abrogated the induction of transcription by HMG2L1 (lane 6), whereas deletions of IR sequences had no apparent effect on transcriptional activation (lanes 4 and 5). A direct comparison of data presented in Fig.2d (lanes 3–7) and Fig.4a (lanes 3–7) reveals that efficiency of transcription is intimately coupled with inducibility by HMG2L1, and that removal of a 65-bp region directly upstream from the transposase ATG renders the 5'-UTR of SB unable either to drive efficient transcription (Fig.2d, lanes 3 and 6) or to respond to HMG2L1-dependent transcriptional activation (Fig.4a, lanes 3 and 6). A human cell is not a natural environment for the SB transposase that is originated from fish. In order to investigate the evolutionary conservation of HMG2L1's effect on the SB transposon, the ability of the *Xenopus laevis* ortholog of HMG2L1 (ref. 33) to upregulate transcription from the 5'-UTR of SB was tested. Similar to that observed with the human protein, xHMG2L1 induced transcription from the 5'-UTR of SB ~5-fold, and this effect required the presence of the 65-bp region (Supplementary Figure S1).

The results suggest that HMG2L1 is a key component of the transcriptional machinery that drives transposase expression from the 5'-UTR of the transposase gene, and that a 65-bp region upstream of the transposase ATG plays a critical role in mediating HMG2L1's effect on transcriptional activity.

Transcription at the 5'-UTR of the transposon is negatively regulated by the SB transposase

We identified HMG2L1 as a component for transcription of the transposase gene, and the above data suggest that its interaction with the SB transposase plays a role in this process. To determine the biological relevance of this interaction, the SB transposase was coexpressed with HMG2L1, and transcriptional activities associated with the 5'-UTR of SB were measured in transient luciferase reporter assays. Coexpression of the SB transposase with HMG2L1 not only abolished HMG2L1-mediated transcriptional activation (Fig.4b, lanes 2 and 3), but apparently had a repressing effect on transcription by the 5'-UTR (Fig.4b, lanes 1 and 3). Furthermore, when the transposase was expressed in the absence of exogenously introduced HMG2L1, a considerable reduction in promoter activity became evident (Fig.4b, lanes 1 and 4), probably due to interactions of the SB transposase with endogenous HMG2L1 protein. To test whether transcriptional repression by the transposase requires primary binding of the transposase to its binding sites, transcriptional activities of a reporter construct lacking the left IR (which contains the binding sites) but retaining the ~160-bp intervening sequence between the left IR and the ATG codon of the transposase coding region were tested. Lack of the transposase binding sites did not affect the ability of the transposase to antagonize HMG2L1-induced transcription (Fig.4b, lanes 5–7).

We conclude that transposase expression in the context of the naturally occurring transposable element is subject to negative feedback regulation, with the transposase acting as a transcriptional repressor. The data suggest that no tethering of the transposase to the IR is required for its repressor function.

Shielding of promoter transactivation by chicken β-globin insulators incorporated into SB vectors

Transposons are emerging alternatives to retroviral vectors for use in gene therapy applications, and the results presented in this study suggest that SB-based vectors may have a safety advantage as compared to retrovirus-based vectors due to the lack of strong element-intrinsic promoter activities (Fig.2a and **c**). However, transactivation of host gene expression upon vector integration may eventually arise from strong promoter/enhancer elements that are not components of the transposon vector itself, but instead are components of the cargo transgene cassette.

In order to simulate transposon insertion upstream of a gene, an SB transposon carrying an SV40-neo cassette (representing a model therapeutic gene) was cloned immediately upstream of the TATA-box minimal promoter-driven luciferase gene (representing a model host gene). This arrangement mimics the integration of an SB transposon carrying a therapeutic gene in close proximity to a host gene driven by its own weak promoter. As shown in Fig.5a, insertion of the SV40-neo cassette carried by the SB transposon leads to a 40- to 100-fold activation (depending on the orientation of insertion) of luciferase gene expression (lanes 1, 2, and 3), consistent with transactivation of the TATA-box promoter by the SV40 enhancer. Indeed, the transactivating ability of an "empty" transposon is reduced as compared to the cargo-containing transposon (lanes 4 and 5). The SV40-neo cassette can also upregulate the thymidine kinase promoter >3-fold (Supplementary Figure S2).

We next tested whether the safety of the model transposon vector can be further improved by flanking the SV40 cassette by chicken *β*-globin insulator (HS4) sequences. The HS4 sequence at the 5'-end of the chicken *β*-globin locus has the two defining properties of an insulator: it prevents an enhancer from acting on a promoter when placed between them ("enhancer blocking"), and acts as a barrier to chromosomal position effect when it surrounds a stably integrated reporter. Further dissection of the core revealed that HS4 is a compound element in which the enhancer blocking and barrier activities can be separated. [34] A CTCF binding site in a 250-bp core element of HS4 is necessary and sufficient for enhancer-blocking activity. [34,35] The SV40-neo cassette was flanked by the core HS4 elements in the SB vector in two possible orientations with regard to the transgene cassette. The insulated transposons displayed a 7- to 51-fold reduction in luciferase transcativation as compared to uninsulated vectors (compare lane 2 to lanes 6 and 7; and lane 3 to lanes 8 and 9 in Fig.5a. Enhancer blocking was efficient in both orientations of transposon integration with regard to the luciferase gene, and in both orientations of the insulators within the transposon (Fig.5a).

We also tested both transactivation as well as enhancer blocking in reporter constructs driving luciferase from the cyclin D1 promoter (Fig.5b). Transcriptional effects on the cyclin D1 gene are highly relevant, as its overexpression was detected in a variety of cancers. [36] Fig.5b shows that the SV40-neo transgene cassette carried by the SB transposon leads to a threefold to fourfold (depending on the orientation of insertion) activation of cyclin D1 promoter-driven luciferase gene expression (lanes 1, 2, and 3). Upregulation of the cyclin D1 promoter is clearly attributable to the SV40-neo cassette, because an "empty" transposon did not transactivate in either orientation (lanes 4 and 5). Apparently, in the context of the cyclin D1 promoter, enhancer blocking was efficient in only one orientation of transposon integration with regard to the luciferase gene, where the insulated transposon displayed a 1.6-fold reduction in luciferase transactivation as compared to the uninsulated vector (compare lanes 3 and 7 in Fig.5b).

To determine whether the additional HS4 insulator sequences would affect transposition efficiency, we compared the transposition efficiency of an uninsulated (T/neo) transposon with an insulated version using a standard transposition assay in HeLa cells. Both transposon constructs were co-transfected together with a helper-plasmid expressing the SB transposase or with a control plasmid expressing an unrelated protein, and antibiotic-resistant colonies were counted. As shown in Fig.5c, the insulated SB transposons can efficiently transpose, but incorporation of the HS4 sequences led to a ~2-fold decrease in transpositional efficiency, possibly due to steric hindrance of transposase action at the transposon IRs by CTCF binding. Furthermore, the HS4 insulator sequences apparently have no effect on transgene expression, as judged by transient neomycin phosphotransferase expression levels provided by insulated and uninsulated transposon constructs (Fig.5d).

Taken together, incorporation of HS4 insulator sequences in SB-based vectors reduces transactivation of promoters by transposon-borne enhancers, and thus may significantly increase the safety of these vectors due to the reduced risk of transcriptional activation of host genes situated close to a transposon insertion site.

Transcriptional activities of transposon vectors in primary human T cells

Activities of promoter/enhancer elements may be the subject of tissue/cell type-specific regulation. In order to substantiate our observations regarding the transcriptional activities of SB transposon-derived gene vectors, some of the luciferase reporter constructs described above were transfected into primary human T cells. The results largely confirmed the data obtained in HeLa cells, and can be summarized as follows. Transcription driven by the 5'-UTR of SB (Fig.6, lane 2) was about sevenfold higher than transcription driven by a TATA-box minimal promoter (lane 1), and this activity significantly dropped when the transposon region, critical for transctivation by HMG2L1, was deleted (lane 3). Outward-directed transcription from the 5'-UTR (lane 4), as well as promoter activity of the 3'-UTR of SB in both orientations (lanes 5 and 6) were negligible. Transposon-associated promoter activities remained significantly lower than those of the MLV LTRs (lanes 7 and 8). Finally, as seen in HeLa cells, an SV40-neo cassette-containing SB transposon can significantly upregulate the TATA-box promoter in T cells (lanes 9 and 10), which can efficiently be blocked by flanking the transgene cassette with HS4 insulators (lane 11).

Discussion

In this study we investigated the transcriptional activities residing within the SB transposable element and its derivatives in order to assess the potential of SB-based vectors to alter host gene expression upon integration in gene therapy applications.

We found that transposon DNA sequences flanking the transposase coding region (*i.e.*, the 5'- and 3'-UTRs) have moderate promoter activities directed towards the inside of the element (Fig.2a and b). Consistent with our findings, both sense and antisense transcripts of Tc1-like transposons were detected in catfish. [37] Convergent transcription of transposons raises the possibility for the formation of transposon-specific double-stranded RNA molecules that may serve as triggers for transposon regulation by RNA interference: an idea that remains to be tested by future investigations.

We identified the cellular HMG2L1 protein as a cellular interactor of the SB transposase (Fig.3). HMG2L1 is an HMG-box DNA-binding domain-containing protein [38] that shares structural similarity with the LEF-1, Sox-4 and SRY proteins. These proteins bind to DNA in a sequence-specific fashion through minor groove contacts, and alter DNA conformation upon binding. [32] Bending of the DNA helix might facilitate the interaction between proteins bound at distant sites in the enhancer region, thereby activating the transcription of target genes. Importantly, members of the SRY family of transcription factors have been found to regulate transcription of the human L1 retrotransposon. [39] In light of the interaction between the SB transposase and HMG2L1, we consider two different mechanisms as to how regulation of transposition by HMG2L1 could be achieved. First, HMG2L1 can act as a transcription factor of the SB transposase gene. Indeed, we provide evidence that HMG2L1 upregulates transcription from the 5'-UTR of SB, and have identified a 65-bp region of the 5'-UTR that is required for HMG2L1-dependent transcriptional regulation (Fig.4a). We discovered that the SB transposase antagonizes the effect of HMG2L1 at the 5'-UTR of the transposon (Fig.4b), suggesting that the SB transposase exerts a negative feedback regulation on its own expression (Fig.7). This model postulates a sensitive balance in the regulation of transposase expression that is calibrated by transposase concentrations in the cell (Fig.7), whereby low concentrations allow more transposase to be made, whereas high concentrations lead to shutting off transposase expression. In addition to the control of transposase expression, interaction between the transposase and HMG2L1 might possibly regulate transcription of yet unknown cellular target genes, thereby affecting transposition. We demonstrated such a mechanism for the SB transposase/Miz-1 interaction that downregulates cyclin D1 expression, resulting in an extended G_1 phase of the cell cycle and more efficient transposition. [31] HMG2L1 has been shown to negatively regulate the Wnt/*β*-catenin signalling pathway; [33] thus, it may be that the SB transposase/HMG2L1 interaction modulates transcription of Wnt/*β*-catenin target genes, which in turn affects transposition. Future investigations will have to clarify if transposon regulation through such a mechanism exists.

SB-based technologies for non-viral gene transfer gained significant ground over the past couple of years. [17] Thus, the results presented in this study bear practical relevance to the use and safety of SB transposon vectors in clinically relevant applications. First and foremost, the 65-bp region within the 5'-UTR of the SB transposon that mediates HMG2L1's activity on transcriptional regulation is not included in SB vectors, because the ~160-bp DNA situated between the left IR and the transposase coding region (Fig.1) is not required for transposition. [23] We further show that transcription from the UTRs towards the outside of the SB transposon is negligible, and occurs at rates comparable to that by the eukaryotic core promoter TATA-box in both HeLa and primary human T cells (lanes 2, 4, and 6 in Fig.2a and lanes 1, 4, and 6 in Fig.6, respectively). The UTRs of the SB element drive outward-directed transcription up to 100-times less efficiently than the strong enhancer/promoter activity of the MLV retrovirus LTR (Fig.2c). A clear distinction between an SB vector used for gene therapy and an SB vector engineered for the purpose of gain-of-function mutagenesis is of importance. SB transposition has been elegantly applied for *in vivo* somatic mutagenesis in the mouse for oncogene discovery. [40,41] The transposon vectors used in those experiments were specifically engineered to contain strong, viral enhancers and splice donor signals to purposefully overexpress genes near transposon insertion sites. This is clearly not the case in a typical SB vector used for gene therapeutic purposes (Fig.1) that would carry only a therapeutic expression cassette flanked by the transposon IRs. In this context, it is important to note that no dominant adverse effects associated with SB vector integration have been observed in experimental animals, [21] not even in a cancer-predisposed genetic background. [42]

We have shown that the cargo transgene sequence carried by SB vectors can exert a profound effect on the activity of a transcription unit linked *in cis* to the transposon vector (Fig.5 and 6), due to the transcriptional enhancer element of the transgene cassette. This presents a safety issue, because therapeutic expression cassettes may inadvertently upregulate a proto-oncogene or other signaling factor that happens to be close to the transposon insertion site. Two strategies are currently considered to lower the risk of insertional mutagenesis by integrating vector systems. [1] The first is the development of technologies that allow target-selected vector integration, so that transgene insertions could potentially be guided into safe sites in the human genome. [43] Gene transfer based on the site-specific integration system ΦC31 bacteriophage [44] is being developed for such purposes. In addition, fusion proteins consisting of HIV-1 integrase and E2C, a synthetic zinc finger DNA-binding protein, were found to bias integration of retroviral DNA near the E2C-binding site in tissue culture cells. [45] Furthermore, SB transposase fusions with E2C and Gal4 have been shown to bias insertion patterns into specific target site-containing plasmids, [46] whereas use of a fusion made up by a protein interaction domain of SB and the tetracycline repressor allowed the recovery of targeted chromosomal transposition events in cultured human cells. [47]

The second strategy is the transcriptional confinement of an expression unit upon genomic integration, which serves two purposes: allow position-independent expression of the transgene (for efficiency), and prevent transactivation of a cellular gene (for safety). The HS4 chromatin insulator of the chicken *β*-globin locus has both of these activities, [34] and was shown to improve the expression performance of murine retrovirus, [48,49] lentivirus [50,51] as well as adeno-associated virus [52]

vectors by protecting them from chromosomal position effects. In addition, a suppression of clonal dominance was found with HS4-insulated lentiviral vectors, [53] suggesting reduced upregulation of host genes upon vector integration due to enhancer blocking by the insulator. We addressed the potential utility of HS4 insulators in enhancer-blocking activity. When flanking an SV40-neo transgene unit within an SB transposon vector, HS4 elements effectively blocked transactivation of a nearby TATA-box minimal promoter in both HeLa and primary human T cells (Fig.5a and 6). Enhancer blocking in the context of the cyclin D1 promoter was less efficient, and showed a bias with respect to the orientation and distance of the SV40 enhancer and the target promoter (Fig.5b). Since HS4 elements have been shown to provide more efficient enhancer-blocking activity when present in multiple tandem copies, [35,54] it may be necessary to incorporate multicopy CTCF binding sites in future transposon vector designs. To sum up, SB transposon-based vectors have a favorable safety profile, because they are fairly inert in their transcriptional activities, and because insulator elements can successfully be incorporated in the next generation of transposon vectors. Thus, SB vectors are expected to have only a limited ability to upregulate a cellular gene located in the vicinity of a transposon insertion site.

Materials and Methods

Plasmid constructs. Mutations were introduced into the 5'-UTR of SB to match the consensus sequence of the salmonid subfamily of Tc1-like elements by QuikChange Multi Site-Directed Mutagenesis (Stratagene, La Jolla, CA). Five point mutations were introduced using the primers L-IRmultimut1, L-IRmultimut2, L-IRmultimut3, and L-IRmultimut4 (sequences of primers used in this study can be found in Supplementary Materials and Methods). To generate a 5'-UTR/ luciferase fusion, the modified 5'-UTR sequence was amplified by PCR using primers IRLuc1, IRLuc2reverse, and IRLuc3reverse. The PCR fragment was cloned into the *Sal* I/*Nar* I sites of pCMVtkLuc [55] to result in pSBIR-L-Luc. Additional luciferase reporter constructs are listed below together with the primers used to amplify the sequence of interest: SBIR-L(reversed)-Luc [primers: SBIR(rev.comp.)/*Sal* I; SBIR(rev.comp)/NarI; SBIR(rev.comp.)]; SBIR-R-Luc [primers: SBIR-R/*Sal* I, SBIR-R/*Nar* I, SBIR-R)]; SBIR-R(reversed)-Luc [primers: SBIR-R2/*Sal* I, SBIR-R2/*Nar* I, SBIR-R2]; FPIR-L-Luc [FP-IR-left (*Sal* I), FPIR-Luc1, FPIR-Luc2/*Nar* I]; SBIR-L(130-391) [primers: SB-IR1; IRLuc3reverse], SBIR-L(228-391)-Luc [SBIR(left; nt 228)/*Sal* I; IRLuc3reverse]; SBIR-L(1-325)-Luc [primers: IRLuc1, IR-Luc-8, IR-Luc9], SBIR-L(1-260)-Luc [primers: IRLuc1, IR-Luc4/*Nar* I, IR-Luc5). The Mo-MuLV LTR was isolated from pLXSN (M28248) by *Spe* I/*Sac* II digest, and cloned into the *Sal* I site of pLuc.

To generate an HS4-insulated T/neo transposon, the HS4 insulator core element was excised from pNI-CD (a kind gift of Gary Felsenfeld) by *Asc* I digest and inserted into the *Eco* RI and *Bam* HI sites of T/Neo to flank the SV40-neo expression cassette at either side in pT/Neo-HS4. Uninsulated (T/Neo) and insulated (T/Neo-HS4) transposon versions were cloned into the *Sal* I site of pTATA-Luc^{+55} in two possible orientations to determine the impact of the transposon on adjacent promoter elements. pCMV-HA/SETMAR was obtained by inserting two copies of the hemagglutinin peptide YPYDVPDYA upstream of the coding region of the human SETMAR protein by PCR, and cloning into the *Bam* HI/*Eco* RI sites of pCMV-SETMAR. [28]

Cloning the human HMG2L1 cDNA. A partial HMG2L1 sequence was generated from three overlapping cDNA clones; one recovered from the two-hybrid screen and two commercially available EST clones (IMAGp998J2210102Q2 and IMAGp958N021240Q2, RZPD, Berlin, Germany). The overlapping cDNAs were fused together using *Avr* II and *Bst* XI restriction sites resulting in a partial HMG2L1 cDNA sequence [HMG2L1(MDLL-IMPGL)] lacking the N-terminal region. The N-terminal region of HMG2L1 (encompassing amino acids 1–110) was assembled by a single-step oligo-assembly method as described. [56] Briefly, 16 oligodeoxyribonucleotides (N1–N16) were used, which collectively encode both strands of the N-terminal region of the *HMG2L1* gene. The oligos were assembled in a PCR using Pwo DNA polymerase (Roche Basel, Switzerland). The generated PCR product was gel-purified, and fused to the partial HMG2L1 cDNA at an *Ava* II restriction site resulting in a full-length HMG2L1 cDNA. An expression construct was generated by PCR amplification of the *HMG2L1* gene [primers: HMG2L1-Start(MAYDDS-IMPGL)/*Hind* III, HMG2L1-Stop/*Xho* I] and cloning into the *Hin* dII/*Xho* I sites of pcDNA3.1(+)/zeo (Invitrogen Karlsruhe, Germany). For coimmunoprecipitation and chromatin immunoprecipitation, a HA-tagged version of HMG2L1 was generated by PCR [primers: HMG2L1-Start(MAYDDS)/*Hin* dIII, HMG2L1-HA-Stop/*Xho* I]. *X. laevis HMG2L1* cDNA was amplified from pCS2-xHMG2L1 [33] by PCR with primers XHMG2L1-Start/*HIN* DIII and XHMG2L1-HA-Stop/*Xho* 1, and cloned into the *Xho* I/*Hin* dIII sites of pcDNA3.1/Zeo(+) (Invitrogen Karlsruhe, Germany).

Cell culture, transfections and transposition assay. HeLa cells were cultured at 37 C and 5% CO_2 in Dulbecco's modified Eagle's medium (Gibco /Invitrogen, Karlsruhe, Germany) supplemented with 10% fetal bovine serum (PAA, Pasching, Austria). One day prior to transfection, cells were seeded onto 6-well plates, and incubated for 16–20 hours until 50–80% confluence was reached. Cells were transfected with purified plasmid DNA (Qiagen, Hilden, Germany) using Fugene6 tranfection reagent (Roche Basel, Switzerland). Transposition assays in HeLa cells were done as described. [18] 10^5 cells were transfected with 200 ng of the transposon donor plasmid pT/neo or pT/neo-HS4 and 90 ng of the transposase-expressing vector pCMV-SB10. [18] For the isolation of T cells, fresh human blood treated with heparin (Ratiopharm, Ulm, Germany) as coagulant was diluted 1:1.3 in Rosewell Park Memorial Institute (RPMI) medium supplemented with 10% fetal bovine serum (RPMI$^+$). Ficoll-Paque Plus (Amersham, Buckinghamshire, UK) was overlayed with blood/RPMI$^+$, centrifuged at 650 g at room temperature, and peripheral blood mononuclear cells were removed from the interphase. Peripheral blood mononuclear cells including T cells were washed twice with RMPI$^+$, once in PBS, and 5 x 10^6 cells were transfected with 3 µg of pCMV-*β* (Invitrogen, Karlsruhe, Germany) plus 2 µg of luciferase reporter constructs (in equal molar ratios) using the Human T Cell Nucleofector Kit (Amaxa, Cologne, Germany).

Reporter gene assays. HeLa cells were transfected as described above with 50 ng luciferase reporter construct, 500 ng pCMV-HMG2L1 or pCMV (as control) and 50 ng

pCMV-*β* as an internal control for transfection efficiency. Two days post-transfection, cells were harvested from 6-ml plates using 400 µl lysis buffer [25 mmol/l Tris-phosphate (pH 7.8), 2 mmol/l dithiothreitol, 2 mmol/l trans-1,2-diaminocyclohexane-N,N,N',N'-tetraacetic acid, 10% glycerol and 1% Triton X-100]. Light intensity was measured in a Lumat LB 9507 luminometer (Berthhold Technologies, Bad Wildbad, Germany) by mixing 100 µl of luciferase reagent [20 mmol/l Tricine (pH 7.8), 1.07 mmol/l $(MgCO_3)_4Mg(OH)_2$ x $5H_2O$, 2.67 mmol/l $MgSO_4$, 0.1 mmol/l EDTA (pH 7.8), 33.3 mmol/l dithiothreitol, 270 µmol/l Coenzyme A, 470 µmol/l luciferin, 530 µmol/l adenosine triphosphate] with 30 µl cell extract. The resulting readings were expressed as relative light units, and normalized to *β*-galactosidase activity. Briefly, 30 µl of protein extract was mixed with 970 µl reaction buffer [100 mmol/l HEPES (pH 7.3), 150 mmol/l NaCl, 4.5 mmol/l L-Aspartate (hemi-Mg salt), 1% bovine serum albumin, 0.05% Tween 20 and 1.6 mmol/l chlorphenol red-*β*-D-galactopyranosid]. After stopping the reaction (0.5 ml 3 mmol/l $ZnCl_2$), the reaction time was recorded, and OD_{578} of the samples was measured. For each sample, the amount of total protein was determined by Bradford Assay (BioRad, Munich, Germany), and OD_{595} readings were recorded. Units of *β*-galactosidase were calculated using the formula: units *β*-galactosidase = 1000 x OD_{578}/(reaction time x OD_{595}). As a measure for promoter activity, light intensities (relative light unit) were normalized to units of *β*-galactosidase.

Yeast two-hybrid assay. In total, 2 x 10^7 independent transformants of a pretransformed HeLa cDNA library (Matchmaker, Clontech, Mountain View, CA) were screened with the SB transposase as a bait protein according to the manufacturer's instructions, using medium stringency selection for protein–protein interactions. Library plasmids were rescued from positive yeast colonies by transformation of *Escherichia coli* and sequenced. The recovered truncated HMG2L1 cDNA was co-transfected with the bait construct to verify the interaction. Yeast clones containing both plasmids were selected and subsequently plated onto medium selecting for protein–protein interaction and incubated at 30 °C. After 12 days, plates were scored for growth by visual inspection.

MBP pull-down. MBP-SB transposase fusion protein and MBP protein expression was induced in *E. coli* strain BL21 (0.4 mmol/l isopropyl-*β*-D-thiogalactopyranoside, 30 °C for 5 hours). Soluble MBP protein was isolated from cells using BugBuster reagent (Novagen Merck, Darmstadt, Germany) according to the manufacturers instructions, and bound directly to amylose resin (New England BioLabs, Ipswich, MA). Protein concentrations were estimated by gel electrophoresis and coomassie-staining, and were adjusted by dilution with unbound beads. Protein-loaded beads were equilibrated for 4 hours at 4 °C on a rotary shaker in binding buffer [20 mmol/l HEPES (pH 7.5), 0.4 mol/l KCl, 25% (vol/vol) glycerol, 1 mmol/l EDTA, 2 mmol/l $MgCl_2$ and 5 mmol/l dithiothreitol]. *In vitro* translation of full-length HMG2L1 was performed with rabbit reticulocyte lysate (Promega, Madison, WI) and [^{35}S]-methionine (Amersham, Buckinghamshire, UK). 10 µl samples of radiolabeled protein were added to the MBP-SB-, and MBP-bound resin, and incubated for 4 hours as above. Beads were collected by centrifugation, and washed 10-times in 1 ml binding buffer with 60 mmol/l KCl and without bovine serum albumin. The proteins were resolved on a 12.5% sodium dodecyl sulfate polyacrylamide gel. Bands were visualized by autoradiography.

Coimmunoprecipitation. Whole-cell extracts were prepared using extraction buffer (50 mmol/l Tris-HCl pH 7.4, 150 mmol/l NaCl, 1 mmol/l EDTA, 1% NP-40 and 0.25% Na-deoxycholate) supplemented with protease inhibitor cocktail (Roche, Basel, Switzerland). For immunoprecipitations, equal amounts of lysate (containing 5 mg of total cellular protein from HeLa cells) were precleared with protein G-agarose beads (Sigma-Aldrich, Munich, Germany). Precleared extracts were incubated with 1 µg rat monoclonal anti-HA antibody (Roche) for 2 hours at 4 °C. Precipitates were washed extensively in extraction buffer, bound complexes were eluted with 2 x sodium dodecyl sulfate polyacrylamide gel electrophoresis sample buffer and resolved by 7.5–10% sodium dodecyl sulfate polyacrylamide gel electrophoresis. Immunoblotting was performed according to standard procedures, and proteins detected with mouse monoclonal anti-SB transposase antibody (R&D Systems, Minneapolis, MN) and chemiluminescence using ECL Advance Western Blotting Detection Kit (Amersham Bioscience).

Chromatin immunoprecipitation. HeLa cells were transiently transfected with CMV-HMG2L1-HA and pSBIRL-Luc as described above. Two days post-transfection, cross-linking, chromatin isolation and enzymatic shearing of chromatin were performed using the ChIP-IT Enzymatic Shearing Kit (Active Motif, Rixensart, Belgium) according to the manufacturer's instructions. Next, 150-µl aliquots of sheared chromatin were diluted with 1.35 ml lysis buffer [50 mmol/l HEPES (pH7.5), 140 mmol/l NaCl, 1% Triton X-100, 0.1% sodium deoxycholate] supplemented with protease inhibitors (Complete Mini, Roche, Basel, Switzerland). Chromatin preparations were precleared twice by incubation with 50 µl Protein G-agarose (Sigma-Aldrich, Munich, Germany) for 1 hour. Then the liquid fractions were split into 500-µl aliquots for immunoprecipitation. Three micrograms of a polyconal HA antibody (Abcam) were added to the samples and incubated overnight at 4 °C. Precipitation was performed by addition of 40 µl of preblocked protein G-agarose [1.5 µg herring sperm DNA (Sigma-Aldrich, Munich, Germany) per µl beads, 30 minutes, 4 °C] for 2 hours at 4 °C. Beads were rigorously washed four times by repeated steps of pelleting and resuspending with 1 ml of lysis buffer, lysis buffer plus 500 mmol/l NaCl, wash buffer [10 mmol/l Tris-HCl (pH 8.0), 250 mmol/l LiCl, 1 mmol/l EDTA, 0.5 % NP-40, 0.5 % sodium deoxycholate] and Tris/EDTA. Protein–DNA complexes were eluted from the beads by incubation with 250 µl of elution buffer [50 mmol/l Tris-HCl (pH 8.0), 10 mmol/l EDTA, 1% sodium dodecyl sulfate] for 20 minutes at room temperature. Formaldehyde crosslinks were reversed for 4 hours at 65 °C in 200 mmol/l NaCl, 20 µg RNAse A (Sigma-Aldrich, Munich, Germany), and ethanol precipitated. Samples were digested with 20 µg Proteinase K (Invitrogen, Karlsruhe, Germany) for 1 hour at 55 °C, and purified using the QIAquick PCR purification Kit (Qiagen, Hilden, Germany). Samples were subjected to diagnostic PCR analysis using primers LucFw and LUC1, and the products were separated and visualized by gel electrophoresis.

Acknowledgments

The HS4 insulator element was a kind gift from G. Felsenfeld NIH, Bethesda, MD. The pCS2-xHMG2L1 construct was obtained from Seiichi Urushiyama, Medical Research Institute, Tokyo, Japan. We thank Andrea Katzer MDC, Berlin, Germany for her help in cell transfections. Work in the authors' laboratory is supported by EU FP6 grants INTHER LSHB-CT-2005-018961 and TRANSFOG LSHC-CT-2004-503438, and a grant from the Deutsche Forschungsgemeinschaft SPP1230 "Mechanisms of gene vector entry and persistence".

Corresponding Author

Zoltan Ivics, Max Delbrück Center for Molecular Medicine, Robert Rössle Strasse 10, D-13092 Berlin, Germany. E-mail: zivics@mdc-berlin.de

References

1. Porteus, MH, Connelly, JP andPruett, SM (2006). A look to future directions in gene therapy research for monogenic diseases. *PLoS Genet* 2: e133.
2. Dobbelstein, M (2003). Viruses in therapy—royal road or dead end? *Virus Res* 92: 219-221.
3. Thomas, CE, Ehrhardt, A and Kay, MA (2003). Progress and problems with the use of viral vectors for gene therapy. *Nat Rev Genet* 4: 346-358.
4. Sinn, PL, Sauter, SL and McCray, PB Jr. (2005). Gene therapy progress and prospects: development of improved lentiviral and retroviral vectors—design, biosafety, and production. *Gene Ther* 12: 1089-1098.
5. Raper, SE, Chirmule, N, Lee, FS, Wivel, NA, Bagg, A, Gao, GP et al. (2003). Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer. *Mol Genet Metab* 80: 148-158.
6. Baum, C, von Kalle, C, Staal, FJ, Li, Z, Fehse, B, Schmidt, M et al. (2004). Chance or necessity? Insertional mutagenesis in gene therapy and its consequences. *Mol Ther* 9: 5-13.
7. VandenDriessche, T, Collen, D and Chuah, MK (2003). Biosafety of onco-retroviral vectors. *Curr Gene Ther* 3: 501-515.
8. Ott, MG, Schmidt, M, Schwarzwaelder, K, Stein, S, Siler, U, Koehl, U et al. (2006). Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1. *Nat Med* 12: 401-409.
9. Recchia, A, Bonini, C, Magnani, Z, Urbinati, F, Sartori, D, Muraro, S et al. (2006). Retroviral vector integration deregulates gene expression but has no consequence on the biology and function of transplanted T cells. *Proc Natl Acad Sci USA* 103: 1457-1462.
10. Medstrand, P, van de Lagemaat, LN and Mager, DL (2002). Retroelement distributions in the human genome: variations associated with age and proximity to genes. *Genome Res* 12: 1483-1495.
11. Scherdin, U, Rhodes, K and Breindl, M (1990). Transcriptionally active genome regions are preferred targets for retrovirus integration. *J Virol* 64: 907-912.
12. Schroder, AR, Shinn, P, Chen, H, Berry, C, Ecker, Jr and Bushman, F (2002). HIV-1 integration in the human genome favors active genes and local hotspots. *Cell* 110: 521-529.
13. Mitchell, RS, Beitzel, BF, Schroder, AR, Shinn, P, Chen, H, Berry, CC et al. (2004). Retroviral DNA integration: ASLV, HIV and MLV show distinct target site preferences. *PLoS Biol* 2: E234.
14. Wu, X, Li, Y, Crise, B and Burgess, SM (2003). Transcription start regions in the human genome are favored targets for MLV integration. *Science* 300: 1749-1751.
15. Hacein-Bey-Abina, S, Von Kalle, C, Schmidt, M, McCormack, MP, Wulffraat, N, Leboulch, P et al. (2003). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* 302: 415-419.
16. Check, E (2005). Gene therapy put on hold as third child develops cancer. *Nature* 433: 561.
17. Ivics, Z and Izsvak, Z(2006). Transposons for gene therapy! *Curr Gene Ther* 6: 593-607.
18. Ivics, Z, Hackett, PB, Plasterk, RH 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.
19. Izsvák, Z, Ivics, Z and Plasterk, RH (2000). Sleeping Beauty a wide host-range transposon vector for genetic transformation in vertebrates. *J Mol Biol* 302: 93-102.
20. Izsvák, Z and Ivics, Z (2004). Sleeping Beauty transposition: biology and applications for molecular therapy. *Mol Ther* 9: 147-156.
21. Essner, JJ, McIvor, RS and Hackett, PB (2005). Awakening gene therapy with Sleeping Beauty transposons. Curr Opin Pharmacol 5: 513-519.
22. Hackett, PB, Ekker, SC, Largaespada, DA and McIvor, RS (2005). Sleeping beauty transposon-mediated gene therapy for prolonged expression. *Adv Genet* 54: 189-232.
23. Zayed, H, Izsvak, Z, Walisko, O and Ivics, Z (2004). Development of hyperactive Sleeping Beauty transposon vectors by mutational analysis. *Mol Ther* 9: 292-304.
24. Vigdal, TJ, Kaufman, CD, Izsvák, Z, Voytas, DF and 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.
25. Yant, SR, Wu, X, Huang, Y, Garrison, B, Burgess, SM and Kay, MA (2005). High-resolution genome-wide mapping of transposon integration in mammals. *Mol Cell Biol* 25: 2085-2094.
26. Leaver, MJ (2001). A family of Tc1-like transposons from the genomes of fishes and frogs: evidence for horizontal transmission. *Gene* 271: 203-214.
27. Sijen, T and Plasterk, RH (2003). Transposon silencing in the Caenorhabditis elegans germ line by natural RNAi. *Nature* 426: 310-314.
28. Miskey, C, Papp, B, Mates, L, Sinzelle, L, Keller, H, Izsvak, Z et al. (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.
29. Zayed, H, Izsvak, Z, Khare, D, Heinemann, U and Ivics, Z (2003). The DNA-bending protein HMGB1 is a cellular cofactor of Sleeping Beauty transposition. *Nucleic Acids Res* 31: 2313-2322.
30. Izsvak, Z, Stuwe, EE, Fiedler, D, Katzer, A, Jeggo, PA and Ivics, Z (2004). Healing the wounds inflicted by sleeping beauty transposition by double-strand break repair in mammalian somatic cells. *Mol Cell* 13: 279-290.
31. Walisko, O, Izsvak, Z, Szabo, K, Kaufman, CD, Herold, S and Ivics, Z (2006). Sleeping Beauty transposase modulates

cell-cycle progression through interaction with Miz-1. *Proc Natl Acad Sci USA* 103: 4062-4067.

32. Bewley, CA, Gronenborn, AM and Clore, GM (1998). Minor groove-binding architectural proteins: structure, function, and DNA recognition. *Annu Rev Biophys Biomol Struct* 27: 105-131.

33. Yamada, M, Ohkawara, B, Ichimura, N, Hyodo-Miura, J, Urushiyama, S, Shirakabe, K et al. (2003). Negative regulation of Wnt signalling by HMG2L1, a novel NLK-binding protein. *Genes Cells* 8: 677-684.

34. Recillas-Targa, F, Pikaart, MJ, Burgess-Beusse, B, Bell, AC, Litt, MD, West, AG et al. (2002). Position-effect protection and enhancer blocking by the chicken beta-globin insulator are separable activities. *Proc Natl Acad Sci USA* 99: 6883-6888.

35. Bell, AC, West, AG and Felsenfeld, G (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 98: 387-396.

36. Fu, M, Wang, C, Li, Z, Sakamaki, T and Pestell, RG (2004). Minireview: Cyclin D1: normal and abnormal functions. *Endocrinology* 145: 5439-5447.

37. Nandi, S, Peatman, E, Xu, P, Wang, S, Li, P and Liu, Z (2006). Repeat structure of the catfish genome: a genomic and transcriptomic assessment of Tc1-like transposon elements in channel catfish (Ictalurus punctatus). *Genetica* 131: 81-90.

38. Seroussi, E, Kedra, D, Kost-Alimova, M, Sandberg-Nordqvist, AC, Fransson, I, Jacobs, JF et al. (1999). TOM1 genes map to human chromosome 22q13.1 and mouse chromosome 8C1 and encode proteins similar to the endosomal proteins HGS and STAM. *Genomics* 57: 380-388.

39. Tchenio, T, Casella, JF and Heidmann, T (2000). Members of the SRY family regulate the human LINE retrotransposons. *Nucleic Acids Res* 28: 411-415.

40. Collier, LS, Carlson, CM, Ravimohan, S, Dupuy, AJ and Largaespada, DA (2005). Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. *Nature* 436: 272-276.

41. Dupuy, AJ, Akagi, K, Largaespada, DA, Copeland, NG and Jenkins, NA (2005). Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. *Nature* 436: 221-226.

42. Carlson, CM, Frandsen, JL, Kirchhof, N, McIvor, RS and Largaespada, DA (2005). Somatic integration of an oncogene-harboring Sleeping Beauty transposon models liver tumor development in the mouse. *Proc Natl Acad Sci USA* 102: 17059-17064.

43. Portlock, JL and Calos, MP (2003). Site-specific genomic strategies for gene therapy. *Curr Opin Mol Ther* 5: 376-382.

44. Thyagarajan, B, Olivares, EC, Hollis, RP, Ginsburg, DS and Calos, MP (2001). Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. *Mol Cell Biol* 21: 3926-3934.

45. Tan, W, Dong, Z, Wilkinson, TA, Barbas, CF 3rd and Chow, SA (2006). Human immunodeficiency virus type 1 incorporated with fusion proteins consisting of integrase and the designed polydactyl zinc finger protein E2C can bias integration of viral DNA into a predetermined chromosomal region in human cells. *J Virol* 80: 1939-1948.

46. Yant, SR, Huang, Y, Akache, B and Kay, MA (2007). Site-directed transposon integration in human cells. *Nucleic Acids Res* 35: e50.

47. Ivics, Z, Katzer, A, Stuwe, EE, Fiedler, D, Knespel, S and Izsvak, Z (2007). Targeted sleeping beauty transposition in human cells. *Mol Ther* 15: 1137-1144.

48. Emery, DW, Yannaki, E, Tubb, J and Stamatoyannopoulos, G (2000). A chromatin insulator protects retrovirus vectors from chromosomal position effects. *Proc Natl Acad Sci USA* 97: 9150-9155.

49. Emery, DW, Yannaki, E, Tubb, J, Nishino, T, Li, Q and Stamatoyannopoulos, G (2002). Development of virus vectors for gene therapy of beta chain hemoglobinopathies: flanking with a chromatin insulator reduces gamma-globin gene silencing in vivo. *Blood* 100: 2012-2019.

50. Malik, P, Arumugam, PI, Yee, JK and Puthenveetil, G (2005). Successful correction of the human Cooley's anemia beta-thalassemia major phenotype using a lentiviral vector flanked by the chicken hypersensitive site 4 chromatin insulator. *Ann NY Acad Sci* 1054: 238-249.

51. Ramezani, A, Hawley, TS and Hawley, RG (2003). Performance- and safety-enhanced lentiviral vectors containing the human interferon-*β* scaffold attachment region and the chicken *β*-globin insulator. *Blood* 101: 4717-4724.

52. Inoue, T, Yamaza, H, Sakai, Y, Mizuno, S, Ohno, M, Hamasaki, N et al. (1999). Position-independent human *β*-globin gene expression mediated by a recombinant adeno-associated virus vector carrying the chicken *β*-globin insulator. *J Hum Genet* 44: 152-162.

53. Evans-Galea, MV, Wielgosz, MM, Hanawa, H, Srivastava, DK and Nienhuis, AW (2007). Suppression of clonal dominance in cultured human lymphoid cells by addition of the cHS4 insulator to a lentiviral vector. *Mol Ther* 15: 801-809.

54. Recillas-Targa, F, Bell, AC and Felsenfeld, G (1999). Positional enhancer-blocking activity of the chicken *β*-globin insulator in transiently transfected cells. *Proc Natl Acad Sci USA* 96: 14354-14359.

55. Altschmied, J and Duschl, J (1997). Set of optimized luciferase reporter gene plasmids compatible with widely used CAT vectors. *Biotechniques* 23: 436-438.

56. Stemmer, WP, Crameri, A, Ha, KD, Brennan, TM and Heyneker, HL (1995). Single-step assembly of a gene and entire plasmid from large numbers of oligodeoxyribonucleotides. *Gene* 164: 49-53.

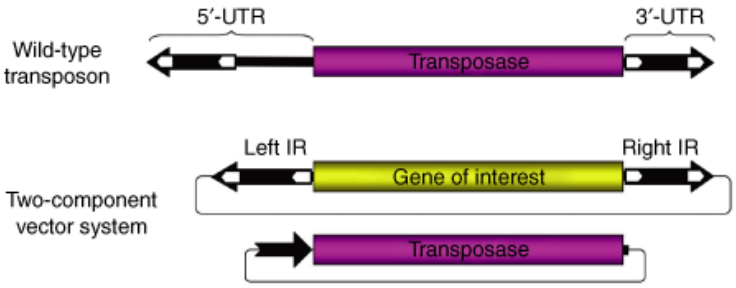

Fig.1: The *Sleeping Beauty transposon system*. Structure and functional components of the transposon are shown. In the wild-type, natural transposon, the central transposase gene is flanked by untranslated regions (UTRs) that include the terminal inverted repeats (IR, black arrows) that contain binding sites for the transposase (white arrows). The left IR is separated from the transposase translational start site by 160 base pairs of intervening sequence (black line), whereas the right IR directly follows the translational stop codon of the transposase coding region. For gene transfer, a two-component vector system is typically used. When used as a vector, the transposase coding region is replaced by a gene of interest within the transposable element that is maintained in a plasmid. This non-autonomous transposon can be mobilized if the transposase is supplied *in trans*, for example by expression from a separate plasmid vector containing a suitable promoter (black arrow).

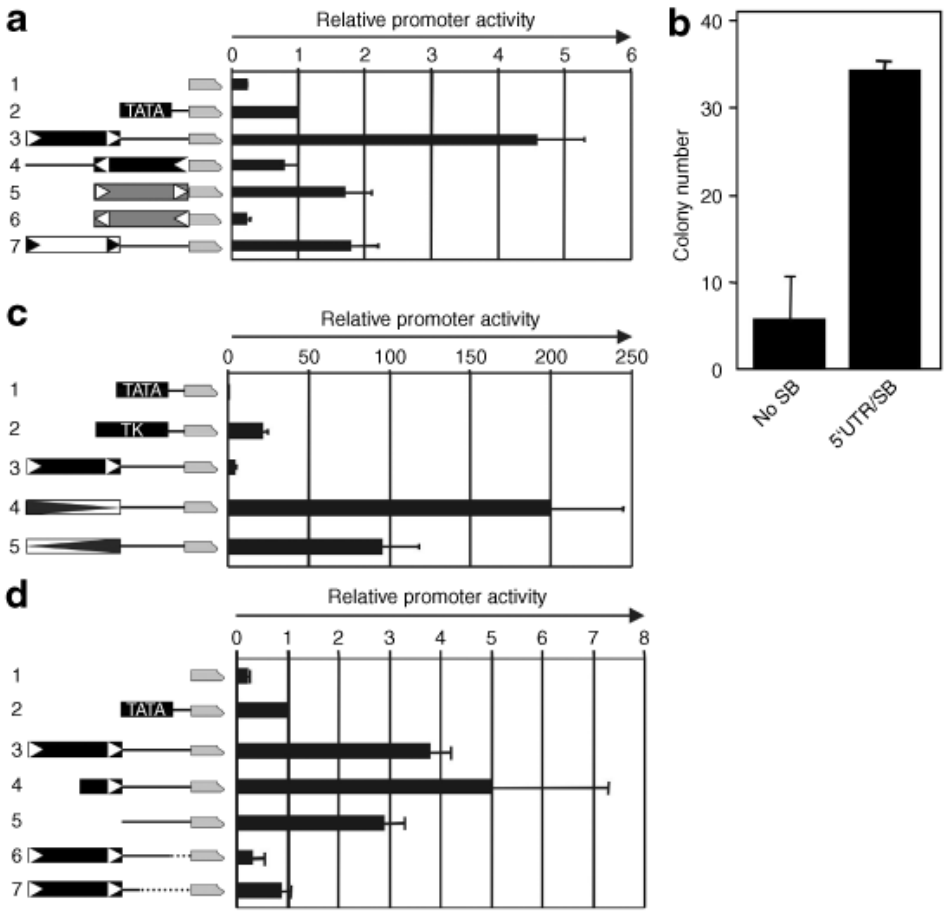

Fig.2: The untranslated regions (UTRs) of the *Sleeping Beauty* (SB) transposon exhibit moderate, directional promoter activities. (**a**) Transcriptional activities residing within the SB transposon. Promoter activities were determined by transient luciferase assays in HeLa cells. Activity of a minimal promoter (TATA-box) control was arbitrarily set to value 1. Transposon sequences flanking the transposase gene were placed in front of a luciferase reporter gene in two possible orientations (in the case of the 5'-UTR, the luciferase gene precisely replaces the transposase coding region). Black box: left IR/DR of SB; gray box: right IR/DR of SB; white box: left IR/DR of *Frog Prince*; small triangles in the boxes: transposase binding sites; black lines connecting the IR/DRs and the luciferase gene represent transposon sequences directly upstream of the transposase coding regions. (**b**) The 5'-UTR of SB can drive transposase expression at a level sufficient for the detection of chromosomal transposition events in cultured cells. A neo-tagged SB transposon plasmid (pT/Neo) was co-transfected together with an SB expression construct in which the transposase is expressed from the 5'-UTR of the transposon or with an empty cloning vector. The difference in numbers of G418-resistant cell colonies is evidence for transposition. (**c**) Promoter activity of the 5'-UTR of SB as compared to those of the thymidine kinase (TK) promoter as well as the murine leukemia virus retroviral long terminal repeats (white box with long black triangle). (**d**) Domain mapping of the region responsible for promoting transcription from 5'-UTR of SB. Assays were done as described in Fig.2a. All transfections were done at least in triplicates, and the error bars represent SEM.

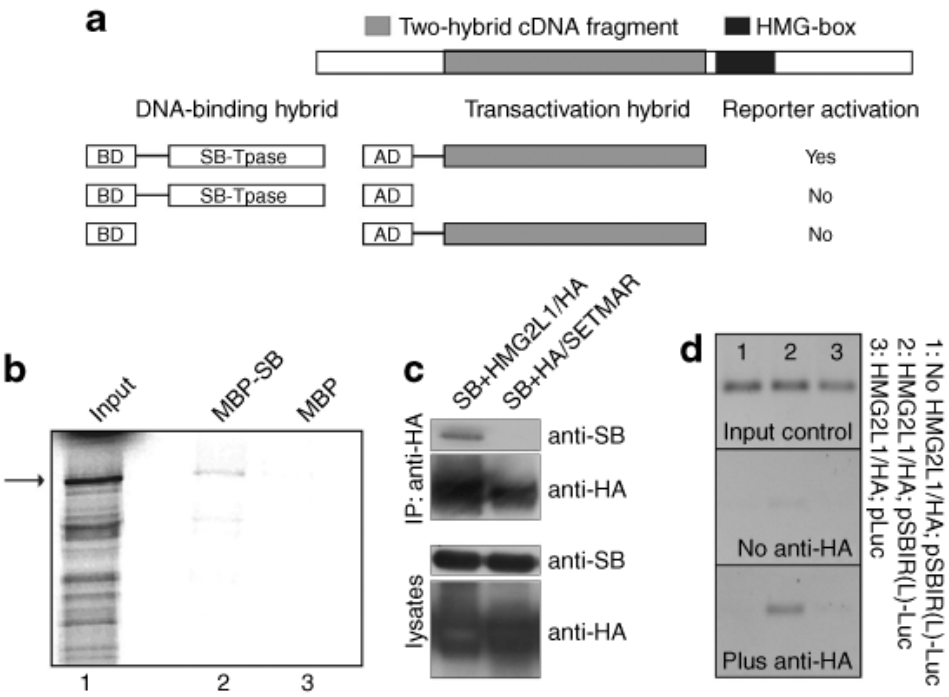

Fig.3: The human HMG2L1 protein physically interacts with functional components of *Sleeping Beauty* (SB). (**a**) SB transposase/HMG2L1 interaction in a yeast two-hybrid assay. Protein–protein interaction is assessed by activation of *ADE* 2 and *HIS* 3 nutritional reporter genes. The top construct displays the full-length coding region of HMG2L1. (**b**) *In vitro* interaction of SB transposase and HMG2L1. ^{35}S-labeled, full-length HMG2L1 was incubated with immobilized maltose binding protein (MBP) or with MBP-SB transposase fusion protein. Bound material recovered after extensive washing was separated by sodium dodecyl sulfate polyacrylamide gel electrophoresis and visualized by autoradiography. The arrow marks HMG2L1. (**c**) *In vivo* interaction between the SB transposase and HMG2L1. Immunoblots of total extracts of HeLa cells coexpressing SB and HMG2L1/HA, or HA/SETMAR as control, were hybridized with anti-SB, and anti-HA antibodies, following immunoprecipitation (IP) with an anti-HA antibody. The Western blots on the cell lysates show proper expression of the test proteins. (**d**) HMG2L1 interacts with the 5'-UTR of the SB transposon. Protein–DNA interaction was determined by *in vivo* chromatin immunoprecipitation. AD, transactivation domain; BD, binding domain; cDNA, complementary DNA; HA, hemagglutinin.

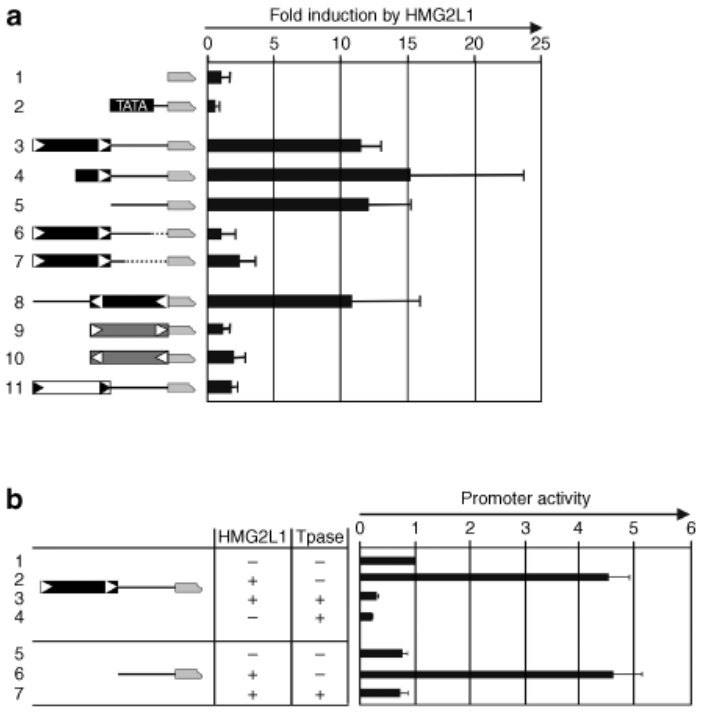

Fig.4: HMG2L1 mediates transcription from the 5'-untranslated region (5'-UTR) of the *Sleeping Beauty* (SB) transposon; the SB transposase is a negative regulator of HMG2L1-mediated promoter activity. (**a**) Transcriptional regulation of 5'-UTR sequences by HMG2L1 in transient transfections in HeLa cells. Data show fold induction of transcription in the presence of HMG2L1 as compared to values measured in the absence of exogenously introduced HMG2L1. (**b**) Promoter activity of the SB 5'-UTR sequence in the presence of HMG2L1 and SB transposase. Values obtained in the presence of empty expression vectors only were arbitrarily set to value 1. All transfections were done at least in triplicates, and the error bars represent SEM.

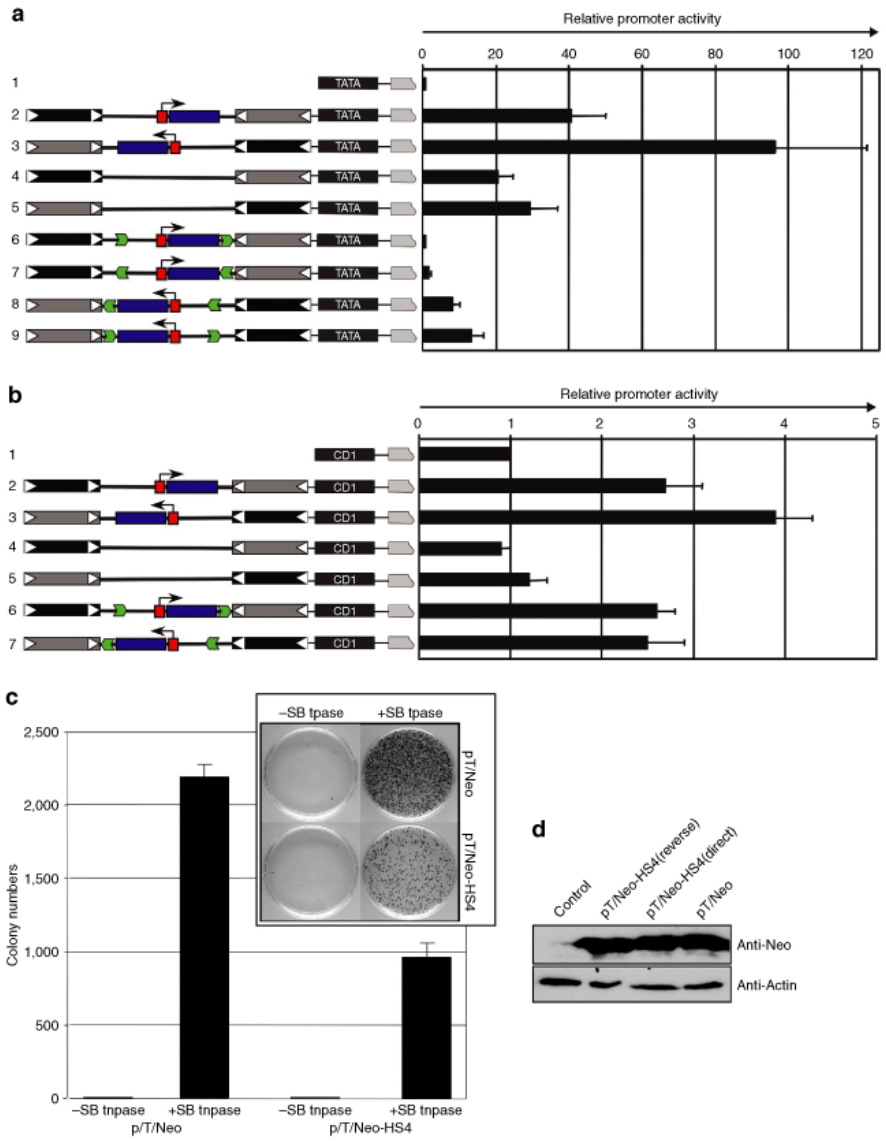

Fig.5: Upregulation of test promoters by vector-borne expression units, and its shielding with insulators. (**a**) A *Sleeping Beauty* (SB) transposon carrying a SV40-neo transgene cassette (T/Neo) can activate transcription of a nearby minimal promoter (TATA-box). The capacity of the cargo gene to activate adjacent promoter elements can be efficiently reduced by flanking the transgene with HS4 insulator core elements. Transcriptional activity was determined by transient luciferase assays in HeLa cells, and activity of the TATA-box was arbitrarily set to value 1. Black box: left IR of SB; gray box: right IR of SB; small triangles in the boxes: transposase binding sites; red box: SV40 enhancer/promoter element; blue box: neo marker; green arrows: HS4 insulator elements; arrows indicate the direction of transcription initiated at the SV40 promoter. (**b**) T/Neo can transactivate the cyclin D1 (CD1) promoter, and the HS4 insulator reduces transactivation in one orientation of transposon insertion with respect to the luciferase marker gene. Transcriptional activity was determined as described in Fig.5a, and activity of the CD1 promoter was arbitrarily set to value 1. All transfections in Fig.5a and b were done at least in triplicates, and the error bars represent SEM. (**c**) Transpositional activity of an HS4 insulator-carrying transposon (T/Neo-HS4) was compared to a T/neo transposon lacking HS4 sequence elements by a standard transposition assay. In control experiments, SB transposase was replaced by *β*-galactosidase, which is unable to support transposition. (**d**) HS4 insulator sequences do not affect expression levels of the neomycin phosphotransferase as determined by Western analysis. IR, inverted repeat.

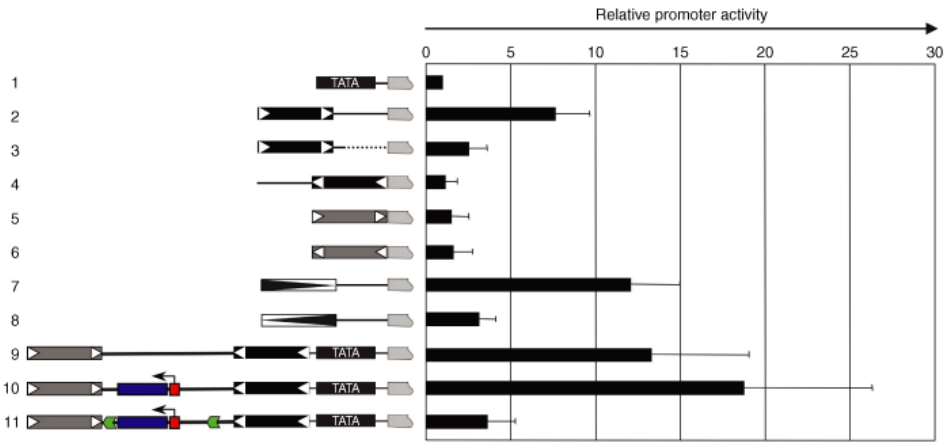

Fig.6: Promoter activities of the *Sleeping Beauty* (SB) transposon untranslated regions, upregulation of test promoters by transgene cassettes, and insulator effects in primary human T cells. Transcriptional activities of selected constructs shown in Fig.2, 3 and 5 were determined by transient luciferase assays in human T cells. Activity of a minimal promoter (TATA-box) control was arbitrarily set to value 1. Black box: left IR of SB; gray box: right IR of SB; small triangles in the boxes: transposase binding sites; white box with long black triangle: murine leukemia virus long terminal repeat; red box: SV40 enhancer/promoter element; blue box: neo marker; green arrows: HS4 insulator elements; arrows indicate the direction of transcription initiated at the SV40 promoter. Mean values and SEM of luciferase measurements are from four independent transfections using T cells isolated from four different blood donors.

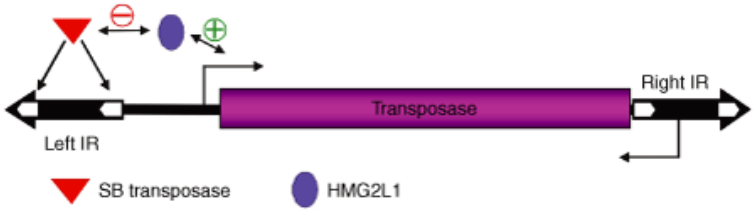

Fig.7: A model for transcriptional regulation of the *Sleeping Beauty* transposase gene. In the wild-type, natural transposon, the central transposase gene is flanked by untranslated regions (UTRs) that include the left and right inverted repeats (IRs, black arrows) that contain binding sites for the transposase (white arrows). Arrows indicate the direction of transcription that is initiated within the UTRs. Transcriptional start sites have not been mapped, so the positions of the arrows are arbitrarily chosen. HMG2L1 upregulates, whereas SB transposase downregulates transcription from the 5'-UTR.