UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

PhD

PROGRAM IN TRANSLATIONAL
AND MOLECULAR MEDICINE

DIMET

UNIVERSITY OF MILANO-BICOCCA

SCHOOL OF MEDICINE AND SCHOOL OF SCIENCE

Defining an innovative and safe non-viral gene delivery system: perspective analysis for gene therapy applications

Coordinator: Prof. Andrea Biondi
Tutor: Prof. Giuliana Ferrari
Co-Tutor: Prof. Alessandra Recchia

Dr. Giandomenico TURCHIANO
Matr. No. 734645

XXV CYCLE
ACADEMIC YEAR
2011-2012

Alla mia famiglia.

Table of contents

Introduction

Biology is the science that studies life. One of the first concept that the students are faced with is the main definition of cell and virus: cells are living because they can organize an environment, they are in homeostasis, they reproduce, grew and develop, assume energy and material from environment and transform them, they respond to external stimuli and show environment adaptations. Viruses are considered not properly living organism, although biology's books spent several pages on this topic, because they cannot accomplish all the features that a cell does. Viruses are defined as obliged intracellular parasites, and thanks to some cellular mechanism they comes to life and run through a life cycle and propagate themself. Sometimes the cell takes advantage from a virus life cycle but the real matter is that both, cells and viruses, found mechanisms to stay alive, propagate, perturb the hosting environment taking advantages from each other. This last concept is kind of philosophical but the biology is teaching us that there are several level of "life" to take in count: prion protein can propagate and be pathogenic; the firsts attempt of life are supposed to be only supported by RNA activity without any DNA or protein hint; DNA Transposable Elements (TE) can "jump" from one location to another in a genome and are considered molecular parasites. These examples were studied because they create an effect in the hosting environment. In this thesis I will deeply analyze a specific level of "life" that was discovered, "resurrected" and harnessed as a biology tool: the *Sleeping Beauty* Transposon.

Transposon history

In the 1942 Barbara McClintock wondered what create a certain type of effect on maize crop kernels: all of them were yellow but some blue, red and brown. The synthesis of kernel pigments is controlled by the genes of the maize plant but that genetic behaviour was not following the laws of the inheritance stated by Gregor Mendel. How maize kernels sometimes have numerous spots or dots rather than being evenly coloured as would be expected on the basis of mendelian genetics? She found some recurrent mutations on the anthocyanines genes, involved in the kernels pigmentation, that could disrupt the proteins frame sequence[1]. After these observations she could conclude that changes in genic expression result from chromosome alterations at the locus of the gene and these are initiated by units other than those composing the gene itself. After extensive crop self pollinated experiments she could surprisingly observe that these units may be transposed from one location to another within the chromosome complement. Furthermore, she could identify the action of one particular system that modified the expressions in a number of different loci; the so-called Dissociation-Activator (Ds-Ac) two-unit system. Both Ds and Ac are single chromosomal units and can undergo transposition. It also been learned that in order the transposition of Ds to occur, Ac must be present in the nucleus[2].

These studies represent the first characterizations of the transposon (Ds) and the transposase (Ac) activities. Today we know that our genome was modelled by the activities of several Transposable Elements that make up the 45% of whole genome content (Figure 1).

Transposon categories

Transposons are best viewed as molecular parasites that propagate themselves using resources of the host cell. However, unlike viruses that produce infectious particles, transposons are not infectious, and their life cycle are therefore confined into intracellular space. Due to this restriction, transposable elements have to coexist with host cells and therefore evolved different strategies to balance host interaction[3]. Despite their parasitic nature, there is increasing evidence that transposable elements have a powerful force in gene evolution. Indeed, about 50 human genes are derived from transposable elements [4], among them genes that are responsible for immunoglobulin gene recombination in all vertebrates.

Transposable elements fall into two major classes.

The first is retrotransposons that transpose through a RNA intermediate and include long interspersed elements (LINEs), short interspersed elements (SINEs), and long terminal repeat (LTR) retrotransposons. Retroelements transpose through a replicative (copy and-paste) process, in which the donor element itself does not move but, instead, its copy autonomously produced

integrates elsewhere in the genome. Similar to retroviruses, retrotransposons use reverse transcription as an essential component of their life cycle[5].

Members of the second class are called DNA transposons, because they move directly as DNA, by a conservative (cut-and-paste) mechanism of transposition, in which the element gets excised from the donor locus and is subsequently reinserted elsewhere (Figure 2) or through a replicative transposition, not yet well characterized, that likely replicate using a self-encoded DNA polymerase[6] and probably transpose through a replicative, copy-and-paste process.

Retrotransposon

Retrotransposons can be subdivided into two groups, distinguished by the presence or absence of Long Terminal Repeats (LTRs).

LTR retrotranspson

Human LTR elements account for ~8% of the genome. They have direct LTR and variable sizes from 100 bp to 5 kb. They replicate in a copy-and-paste mechanism through reverse transcription of their RNA and integration of the resulting cDNA into another locus. They are also defined endogenous retroviruses because share with them the mechanism of replication, with the difference that retrotransposons do not form infectious particles. The LTRs are direct sequence repeats that flank the internal coding region,

which, in all autonomous (functional) LTR retrotransposons, includes genes encoding both structural and enzymatic proteins. The *gag* gene encodes structural proteins that form the virus-like particle (VLP), where the reverse transcription takes place. The *pol* gene encodes several enzymatic functions, including a protease that processes the Pol polyprotein into the reverse transcriptase (RT) that copies the retrotransposon's RNA into cDNA, and the integrase that catalyzes the insertion of the cDNA into the genome[7, 8].

Non-LTR retrotransposon

Non-LTR retrotransposons (also called polyA retrotransposons, or target-primed retrotransposons), are ancient genetic elements that have persisted in eukaryotic genomes for hundreds of millions of years, and are perhaps best known for their enormous spreading in the human genome. LINEs (L1) count 18% of the whole human genome and SINEs (Alu, SVA elements) the 10% (Figure 1). The average human genome contains 80-100 L1 elements competent for retrotransposition whose encoded proteins mobilize their L1s in cis. The L1 Reverse Transcriptase (RT) is also required for trans-mobilization of SINE considered non-autonomous elements. Full lenght, autonomous non-LTR retrotransposons typically contain one or two open reading frames (ORFs). All of non-LTR retrotransposons contain an RT domain. Also present is an endonuclease domain, although encoded endonuclease protein is not an absolute requirement for non-LTR retrotransposition[5, 9]. The integration target site is

dictated by the encoded endonuclease domain, that in most cases has a relaxed target site sequence. The importance of the endonuclease domain lies in producing a primer for reverse transcription. After or during the minus strand DNA synthesis, a second strand nick of the target site occurs. This process, still not entirely clear and probably different between the various elements, creates a primer for plus strand synthesis. The newly synthesized minus strand change templates from the retrotransposon RNA to the target site DNA top strand. Once DNA synthesis is complete, the remaining nicks are ligated to fill up the gap.[9, 10].

DNA transposon

Class II transposable element or DNA transposons are mobile DNA that move utilizing a single or double-stranded DNA intermediate. Eukaryotic DNA transposons can be divided into three major subclasses: (I) those that excise as double-stranded DNA and reinsert elsewhere in the genome, “cut-and-paste” transposons; (II) those that utilize a mechanism probably related to rolling-circle replication, Helitrons[11]; and (III) Mavericks, whose mechanism of transposition is not yet well understood, but that likely replicate using a self-encoded DNA polymerase[6]. Both Helitrons and Mavericks probably transpose through a replicative, copy-and-paste process.

To date, ten superfamilies of cut-and-paste DNA transposons are recognized such as hAT, TC1/mariner, piggyBac, etc. Elements

belonging to the same superfamily can be linked through transposases sequence philogeny. Some of these cut-and-paste transposon hold inverted terminal repeats (ITR). These sequences, flanking the transposon core element, are specifically recognised by the cognate transposase and used as substrate for excision and re-ligation[12]. The catalytic activity could leave staggered ends and produce a footprint in the excision site or be precise without any gain, loss or duplication of genetic material. This process, unlike that for retrotransposons, doesn't produce copies of the transposed element, unless it occurs during DNA replication, in which case one of the two daughter cells will present an additional copy of the TE (Figure 2).

DNA transposons are present in many organisms but are far more frequent in prokaryotes than in eukaryotes, especially vertebrates. For example, Class I TEs make up roughly 42% of the human genome, while DNA transposons account for about 3% of its length (Figure 1). On the other end, bacterial TEs belong almost exclusively to Class II.

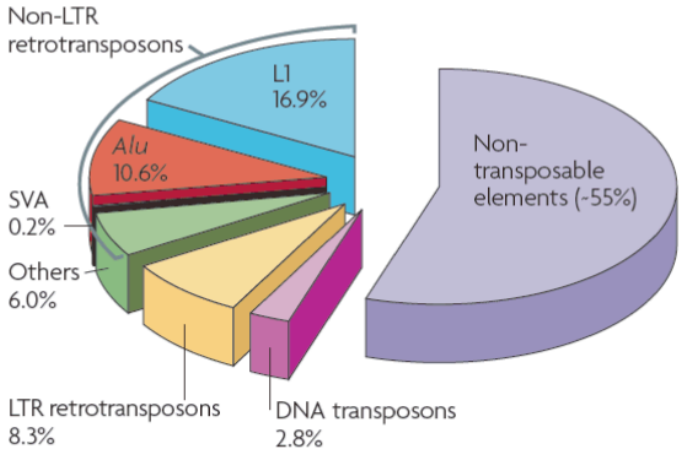


Figure 1. The transposable element content of the human genome[13]

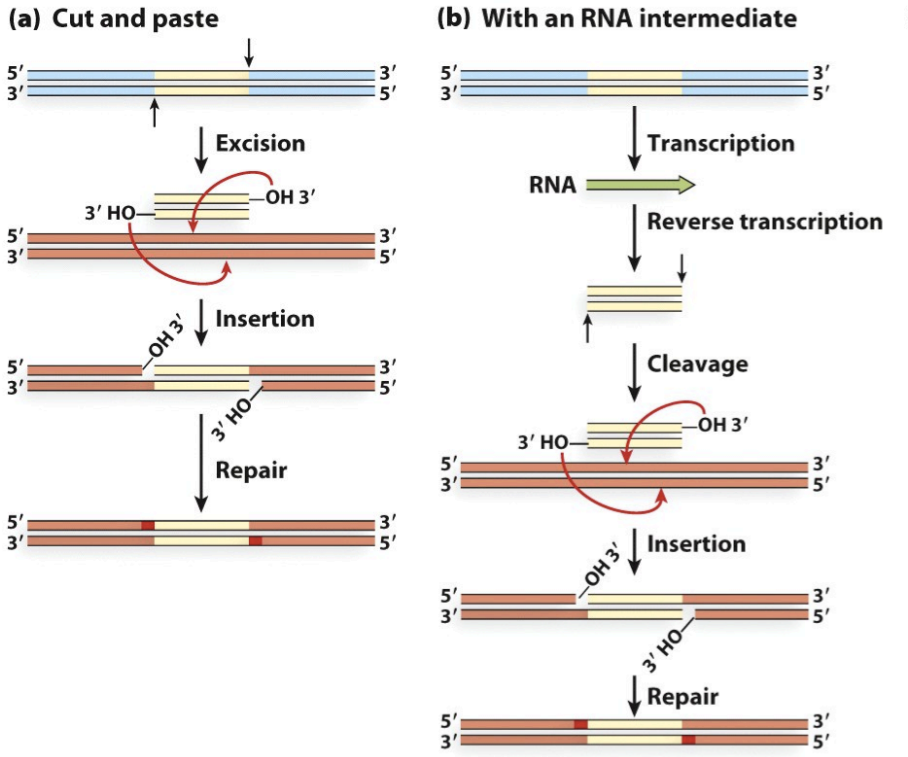


Figure 2. Transposon mechanism of action. A) DNA transposon cut-and-paste schematic mechanism of action in which the re-insertion step is concluded by repairing cellular factor. B) Retrotransposon copy-and-paste

mechanism of action through reverse transcription of their RNA and integration of the resulting cDNA into another locus.

Transposon Element in genome evolution

The extremely high density of Transposon Elements (TE) in our genome poses the question: what is their evolutionary significance and impact during human evolution? TEs have the potential to influence the evolutionary trajectory of their host via alterations of gene function through insertion, via induction of chromosomal rearrangements, or providing a source of coding and non-coding genetic material that allows the emergence of genetic novelty (such as new genes and regulatory sequences). Studying the sequences phylogeny was possible to assume the TE mediated origin for some genomic elements. One of the main example of this phenomenon derive from the origin of the adaptive immune system; V(D)J recombination is the process by which a virtually infinite population of distinct antibodies can be generated in B and T lymphocytes. The two essential components of V(D)J recombination are (I) the RAG1 and RAG2 proteins, which interact to form the recombinase responsible for the joining and transfer activities; and (II) the recombination signal sequences (RSS) flanking the V(D)J segments, which define the specific sequences bound, cleaved, and joined by the RAG1/2 protein complex[14]. The analogy of the process of V(D)J recombination to a transposition reaction is striking. RAG1/2 can catalyze transposition of a DNA segment flanked by RSS in vitro [15]. Additionally, it has been observed that several eukaryotic

transposases utilize a cleavage chemistry similar to that seen in V(D)J recombination[16]. Another example of TE "domestication" come from the formation of eukaryotic telomerase. Comparative sequence analysis of the Telomerase Reverse Transcriptase proteins using the same set of RT motifs places telomerase in the group containing non-LTR retrotransposons. In other organisms the telomere maintenance is carried by non-LTR retrotransposons through a process called target-primed reverse transcription, in which the RT uses a 3'-OH at the DNA break as a primer. Similarly, telomerase uses the 3'-OH at the natural chromosome DNA end as a primer. Therefore, the close relationship of Telomerase to the non-LTR RT can be derived from functional criteria as well as by the RT motif analysis[17]. Moreover, the ATRN gene provides a good example of how TE induce an alternative mRNA processing that can enable functional diversification of one gene. A subset of ATRN transcripts are cleaved and polyadenylated within an L1 element that has retrotransposed into an intron. Other transcripts splice around the L1 element and incorporate an additional five exons. Transcripts polyadenylated within the L1 element encode a soluble form of Attractin (protein encoded by ATRN gene) that is released by activated T lymphocytes as part of the basic inflammatory response. The alternative transcripts encode a protein with transmembrane and cytoplasmic domains that is membrane-bound. This isoform is similar to murine Atrn, which functions as a receptor involved in pigmentation and energy metabolism[18]. This is a clear example of how a single retrotransposition event can increase transcript diversity with

direct consequences on cellular function. Recently there has been much debate about the extent and significance of somatic retrotransposition. L1 expression increases as neural stem cells commit to a neuronal lineage, and this was reported to be associated with frequent L1 retrotransposition[19].

There are many other example, recently reviewed, of how TEs improve cell mechanism, cause diseases or impact with specific features[12]. It's estimated that the most active element are the SINE Alu non-LTR retrotransposon, with one de novo germline insertion per 20 births[20] while de novo insertions of L1 elements occur at the rate of approximately 1 in 108 births [21].

"Sleeping Beauty" Transposon

Tc1/mariner transposon superfamily

The Tc1/mariner superfamily takes its name from the two well characterized members, *Tc1* from C. elegans and *mariner* from Drosophila mauritiana, widely spread in nature. Elements of this superfamily can be found in a great variety of organism, including vertebrates. They are generally 1300-2400bp long and their structure consists in a central element, coding for the transposase, flanked by Inverted Repeated sequences (IRs). These two elements differ in all subgroups giving them specific features.

A first group of elements, comprising Tc1 and mariner, has IRs of approximately 100bp and present a single transposase binding site per repeat. The IRs of a second group, to which Tc3 element

belongs, can be more than 400bp long and possess two binding sites, although the internal pair isn't required for transposition. A third group, named IR-DR, has a pair of binding sites containing short, 15-20 bp direct repeats (DRs) located at the ends of inverted repeats that are 200-250 bp long. Both the outer and the inner pairs of transposase-binding sites are required for transposition. Taken together, the Tc1/mariner superfamily contains elements of simple structure, in which the transposase gene is flanked by a pair of transposase binding sites, and elements of a more sophisticated structure carrying multiple binding sites selected to confer specificity of the transposition reaction.

Other than the IRs, a highly variable element among Tc1/mariner members is the transposase primary sequence: in fact, there is only about 15% amino acid identity between the various families. Despite this, many structural elements are conserved in all Tc1-like transposases. Three major domains can be identified in all of these enzymes: a DNA-binding domain at the N-terminus, a catalytic domain at the C-terminus and a Nuclear Localization Signal (NLS) in between the two, partially juxtaposed to the binding domain.

1) The DNA-binding domain of Tc1-like elements consists of two helix-turn-helix (HTH) motifs, the first of which reminds the structure of the paired domain of certain transcription factors such as Pax elements. The second HTH motif is, instead, similar to a homeo-domain. Other transposases, such as that of the pogo element found in D.

melanogaster, have binding domains made up of only one HTH motif. Another important element conserved in Tc1/mariner transposases is a GRPR-like sequence found between the two HTH subdomains. This motif corresponds to an AT-hook, a structural element that binds DNA in the minor groove of A-T base pairs[22].

2) The nuclear localization signal (NLS) partially overlaps the DNA-binding domain. Thanks to this signal the Tc1/mariner transposases use the host cellular machinery to be transported to the nuclear compartment. The NLS is also flanked by phosphorylation signals for casein Kinase II whose phosphorilation seems to be required in the transposition event. Importance of the NLS is also highlighted by the fact that even a single point mutation at this site could impair transposition.
3) Catalytic domain at the C-terminus of the transposase is responsible for the excision and integration of the TE. This domain presents a highly conserved DDE motif (or DDD in some elements, such as mariner). This structural element is fundamental for the catalytic activity of the transposase and is also found in many other transposases and integrases, such as that of HIV-1.

The transposition mechanism is complex, as it requires many steps and can differ among several elements of the Tc1/mariner superfamily. A common element is the excision of the TE by staggered cuts, which leaves behind a few (2-3) bases of the transposon, forming a footprint useful in identifying excision sites.

Another common feature of many Tc1/mariner elements is their integration at TA sites. Insertion site preference is also influenced by sequence composition and physical properties of the DNA flanking the TA site[23-25]

Sleeping Beauty reconstruction

DNA transposons have been engineered as tools for insertional mutagenesis and germline transgenesis in invertebrate model organisms. The elements that have been extensively used are: the P element in Drosophila and the Tc1 element in Caenorhabditis Elegans[23, 26]. Unfortunately, no such active, endogenous system was available in vertebrates until recently. This is because the vast majority of DNA transposons in vertebrate genomes were inactivated millions of years ago, thereby prohibiting the isolation of an element convertible into a gene transfer vector. To circumvent this problem, the idea of adapting invertebrate transposons for use in vertebrates has been around for quite some time. Another approach to remedy the problem in vertebrates was the molecular reconstruction of a Tc1/mariner type, active vertebrate element from ‘‘dead’’ transposon fossils found in fish genomes, which was named Sleeping Beauty (SB)[27]. The resurrection of SB generated a considerable interest in developing new, safe, simple, and efficient technologies for somatic gene transfer in humans, for gene therapeutic purposes.
In the 1997 Zoltan Ivics and Zsuzsanna Izsvak derive a majority-rule consensus sequence and identify conserved protein and DNA

sequence motifs that likely have functional importance by the alignment of 12 partial salmonid-type Tc1-like element (TcE) sequences found in 8 fish species. Five regions were highly conserved in all TcE transposases: a bipartite nuclear localization signal in the N-terminal half of the transposase, a glycine-rich motif close to the center of the transposase without any known function at present, and three segments in the C-terminal half comprising the DDE domain. Multiple sequence alignment also revealed a fairly random distribution of mutations in transposase coding sequences although fewer amino acid substitutions were detected at the conserved motifs as compared to protein regions between the conserved domains. This implies that some selection had maintained the functional domains before inactivation of transposons took place. Starting from these analyses was possible to reactivate the transposase restoring the open reading frame by removing the premature translational stop codons and frameshifts. They undertook two reconstruction strategies: one fuses the putative functional transposase domains together and, the second synthesizes variants of the protein sequence for all of the mutant amino acids that differs from the consensus transposase sequence. Consequently, a series of constructs, was made by PCR mutagenesis and tested in a PCR assay that detect the transposon excision from his origin plasmid. The result was a transposase gene encoding 340 aminoacids, designed SB10.

Similar studies were performed for the Inverted Repeated elements. The “Sleeping Beauty” transposon IR have 210-250 bp at their termini and directly repeted DNA sequence motifs (DRs)

at the ends of each IR. The match, between the sequences analysed, was less than 80% at the central region of the IRs, but was perfect at the DRs, suggesting that this non-random distribution of dissimilarity could be the result of positive selections that have maintained functionally crucial sequence motifs in the IRs. Therefore, mutations within the IRs but outside the DRs would probably not impair the ability of the element to transpose. Finally Ivics et al. choose, as a model substrate, a single salmonid-type TcE from Tanichthys albonubes (designed as T), whose sequence was 3.8% divergent from the salmonid consensus and has intact DR motifs.

The T Inverted Repeat element and the SB10 transposase together constitute the first Sleeping Beauty transposon/transposase system[27].

Sleeping Beauty development

The “resurrection” of an active vertebrate DNA transposon based system, aroused the interest of the scientific community for the enormous potential as a tool in mammalian cellular biology. This interest was converted in an extensive engineering effort in order to ameliorate the transposition activity modifying both components of the SB system, the transposon IR and the transposase enzyme.

Transposon IR development

Alignment studies on IRs gave indication that even though the DR sequences in the SB transposon are nearly the same they are not identical. In view of this observation Zongbin Cui produces in the 2002 several transposon constructs in order to define structure and functions of IRs such as modification of the outer and inner DR sites, left and right IR, and, spacer sequence that interposes the DRs[28]. During these studies he reverts a presumptive mutation in the internal DR of the right IR and this could increase the transposition activity of about the 80%. Moreover he could observe how the dinucleotide TA placed at the end of IR transposon modifies the excision and the integration efficiency: absence of this dinucleotide almost abolish the activity of the SB system while extra dinucleotides (TATA) at the end of both the IRs double the transposition. Then, he could revert three other point mutations found by the alignment of the sequence between the DRs and the respective fragment in defective transposon placed in fish genomes[29]. Thanks to these three mutations the transposition efficiency increased of about 40%.

Finally, he combined all of the site-specific mutations that increased transposition activity with the double-TA flanks to the transposon to produce a new transposon construct IR, called T2, with a fourfold enhanced activity over the standard T construct.

In the same year, Izsvak et al, published an improvement of the T transposon upon similar manipulation of the wilde type sequence[30]. They show how the transposon flanked by two left

IR showed an almost 3-fold increase in transpositional efficiency compared with the wild-type element with one left and one right IR. Surprisingly, transposons flanked by two right IRs of SB showed a reduced activity. This result indicates that some elements contained in to the left IR that are required for an higher transposition efficiency. An accurate analysis on the left IR reveals the presence of a 11-bp sequence that resemble the 3'-half of a DR binding site. This "Half-Site" is also recognised in other two Tc1-like transposons, the S and Bari1 elements in Drosophila. Because of its evolutionary conservation among transposon in phylogenetically distant species, they hypothesize that this Half-Site has some function in the transposition of these elements.

Still, another work shows how was possible to manipulate the transposon structure to enhance the activity and to effort the transposition efficiency drop for transposon of bigger size than 10 kb. Hatem Zayed in the 2004 published the "Sandwich" (SA) version of the transposon[31] in which he emulates the structure of a Tc1/mariner Paris transposon[32]. This TE was found longer than 10kb with a complex structure consisting of two virtually identical copies of Paris IR flanking a non-repetitive sequence. These copies of Paris differ in a single base pair in one of the terminal inverted repeats, resulted in a TA dinucleotide duplication and possess an uninterrupted ORF, which encodes a putative protein with strong similarities to transposase of the Tc1/mariner type[32]. In light of this natural construct, Zayed reproduces the SB IR with two complete transposon elements, placed in a head to head orientation, flanking a DNA sequence,

thereby forming a sandwich-like arrangement (Figure 3). To guarantee the activity of this construct, he mutated the terminal 5'–CA bases of the right IR to 5'–GC. This modification doesn't allow the transposase enzyme to use the DR sequences, facing the transposon internal DNA, as a catalytic substrate for transposition. Therefore, only the external (respect to the whole transposon sequence) DRs are subjected to cleavage and insertion.

The structure of the transposon is therefore as follows: intact left IR- body of SB element-disabled right IR- DNA sequence (marker gene or sequence of interest)- disabled right IR- body of SB transposase- intact left IR. They compare ca. 7.7kb sized T and SA transposon in a transposition assay and found a three fold enhanced activity for the Sandwich construct. This is probably due to the 8 transposase binding-sites that can stabilize the synaptic complex formation even in the presence of large cargoes. A comparison between this new sandwich construct and the T2 version of sleeping beauty is still required and will be described in this work.

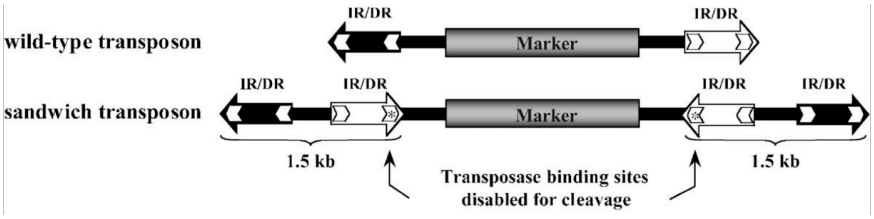


Figure 3. Sandwich transposon. The sandwich transposon has two complete transposon elements at both ends. Highlighted with asterisks the mutated internal IRs unable to function as catalytic substrate of transposition [31].

SB Transposase development

A first improvement of the transposase enzyme was performed in 2003 by introducing amino acid substitutions by site-directed PCR mutagenesis[33]. Originally, after a comparative phylogenetic analysis between SB and active mariner transposases, 14 possibly beneficial mutations were identified and tested. Four of them were selected due to their positive effect on transposition and were incorporated into the original "resurrected" transposase SB10 to produce a new version of the enzyme, designed SB11, that possessed threefold increase in transposition efficiency. They also use the SB11 in combination with the T2 transposon but the transposition rate doesn't ameliorate and remains comparable with the one achieved with the T transposon.

Another similar approach was used in Zayed et al. work [31] in which they analysed transposase protein structure of different TE and tried to replicate into the SB system. They describe three approaches. First, modifications of a linker region between the catalytic and DNA-binding domains were tested. Since this helical region appears to interfere with DNA-binding, native amino acids were substituted with proline in order to disrupt the structure. This approach however, produced no positive results. A second strategy was based on non-conserved acidic amino acid residues substitution with basic amino acids, since these modifications were shown to produce hyperactive forms in other transposases. This strategy was more successful, for the D260K substitution resulted in hyperactivity of the transposase. The last approach

focused on using naturally occurring hyperactive mutations like R115H. This mutation was combined with D260K, therefore producing a more refined version of SB transposase with a 3.7-fold increase in transposition efficiency over SB10. Transposition of this hyperactive system is further enhanced by the transient overexpression of HMGB1, a cellular cofactor of SB transposition[31]. Other mutations focused in the N-terminal domain realized a further enhancement in activity by 9-fold respect to SB. It was also tested in mice transfecting the plasmids by hydrodynamics injection[34]. In the 2005, Baus et al., join all these mutations previously published to obtain an enzyme with a 17-fold higher transposition activity[35]. They focused their attention on the catalytic portion of the gene looking for some common amino acid shared by other transposase, counting also the SB mutants described by Geurts at al.[33] and Yant et al.[34]. Combining this new enzyme with the pT2 contruct the transposition is raised by 30-fold respect to the SB10/T transposon system[35]. A ~100-fold more efficient transposase (respect to SB10), named SB100X, was developed in 2009 by Mates et al.[36] thanks to a high-throughput PCR-based DNA-shuffling strategy. First, by analyzing conserved residues in other Tc1 related transposases, 25 hyperactive mutations were found and were used, in combinations with other previously discovered, to create a library of 41 transposase clones, each containing a single hyperactivating mutation. Then, these clones were fragmented using restriction endonucleases to produce a set of combinatorial units that were further broken in a random fashion

by DNase I treatment. Fragments of 30-70bp were randomly assembled by self-priming PCR in the presence of bridging oligonucleotides. Primers designed to hybridize to the end of the transposase gene were then used to amplify the recombined gene pool and the resulting library was cloned in an eukaryotic expression vector. The ~2000 variants obtained were screened for transpositional activity and the repeatedly recovered ones were subsequently manually combined, leading to the development of SB100X.

Sleeping Beauty molecular activity

As a member of the TC1/mariner superfamily, the SB transposone transposase system has a peculiar molecular activity. The transposase gene can be separeted from the transposon sequence and replace by other DNA sequences because the enzyme can mobilize transposon in trans. The transposase binding sites of SB are repeated twice per IR in a Direct Orientation (four-times in the sandwich transposon). Specific binding to the DRs is mediated by an N-terminal paired-like DNA-binding domain of the transposase that overlaps with a nuclear localization signal. The SB transposition is a cut-and-paste process that could be subdivided into four steps: binding of the transposase to its sites within the transposon IRs, formation of a synaptic complex in which the two ends of the elements are paired and held together by transposase subunits, excision from the donor site and finally integration in the target site[37]. The first step in the transposition reaction is

the binding of transposase molecules to the IRs' binding sites. Since the internal binding sites show a greater affinity to the transposase, these sites are bound first. The binding step is highly specific as it has been shown that SB transposase can discriminate its substrate from a similar sequence differing in only three bases[22]. Two domains are involved in the DR recognition, PAI and RED. The first recognises the 3' portion of the 28 bp composed DR and the HDR sequence, while the RED domain binds the 5' portion increasing binding specificity[38]. In cooperation with the main PAI-RED domain, the GRRR motif function as an AT-hook, contributing to specific substrate recognition.

The following formation of the synaptic complex is mainly in charge to the PAI domain. This portion has a capacity to bind the DR sequence and to stabilize the tetrameric complex created by four SB transposase held close from the four Direct Repeats. The HDR was shown to be important but not essential for the transposition helping to stabilize the synaptic complex. Another important element in this step is the High Mobility Group Protein B1 e B2 (HMGB). This is a nuclear protein associated with eukaryotic chromatin and has the ability to bend DNA. A whole work, describes the direct interaction of the SB transposase with HMGB1[39] in vivo thus forming a ternary complex with the transposon DNA. It suggests that the transposase may actively recruit HMGB1 to transposon DNA via protein-protein interactions therefore bringing transposases closer to each other and facilitating the formation of the synaptic complex as represented in Figure 4.

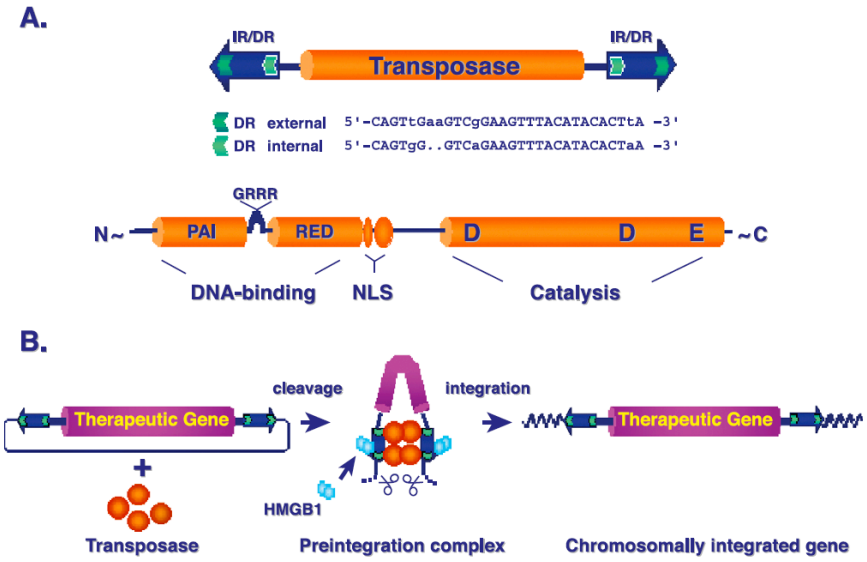


Figure 4. Sleeping Beauty system. A) Schematic representation of transposon. The IR/DR domains placed at the ends of the transposon serve as binding portion (in green) for the transposase. The sequences of internal and external are reported to highlight the differences between them. In the lower part, the scheme of SB transposase protein with its DNA binding domains PAI-RED and GRRR AT-hook motif. Going thorough C-terminal termini is possible to observe the Nuclear Localization Signal NLS and the catalytic DDE domain. B) Four transposase units bind the IR/DR elements and recruit the HMGB1 protein for an efficient formation of the preintegration complex [37].

Once the synaptic complex is completely and correctly assembled, the excision of the transposon, catalyzed by the C-terminal DDE catalytic domain of the transposase, can take place. In this reaction, a single-strand nick is introduced and it is not clear whether the second strand cut happens through transesterification or by hydrolisis. The excision reaction produces cuts staggered by three nucleotides 5'-CTG-3', therefore leaving 3' overhangs. Therefore, when the excision site is repaired, a footprint of the transposon is formed, rendering these sites easier to identify[38].

After the excision, the mobilized element can be inserted into a new genomic location thanks to the action of the transposase tetramer. The insertion TA site is also cut in a staggered way and is therefore duplicated upon insertion as depicted in Figure 5. Insertion step is always in charge of the DDE catalytic domain.

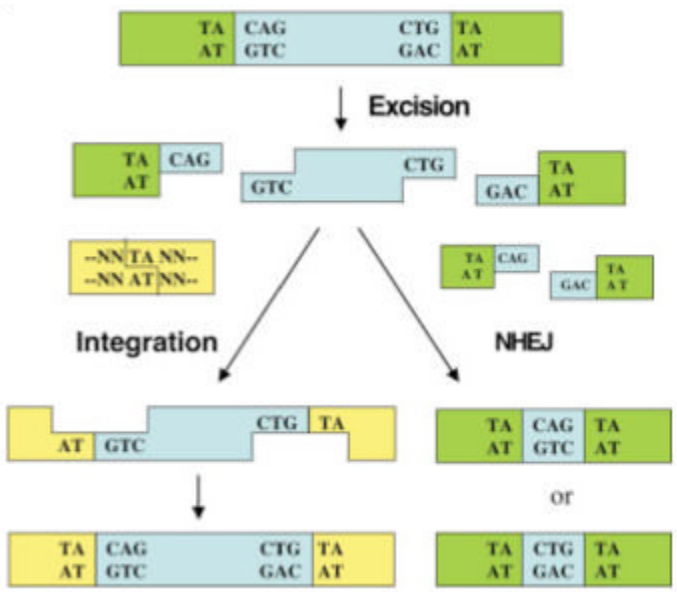


Figure 5. The *Sleeping Beauty* (SB) transposon and its transposition. "Cut-and-paste" mechanism of SB transposition. The two major steps involved in transposition, excision and integration of a transposon are shown. The two broken ends at the donor sites, joined together by non-homologous end-joining (NHEJ) enzymes encoded by the host, leave a footprint at the donor site [23, 40].

Several members of the DDE domain recombinase family integrate fairly randomly and no consensus sequence can be recognised outside the dinucleotide TA. Therefore, common feature affecting the DDE target selection could come from structural proprieties of the DNA sites. Vidgal has proposed conformational proprieties which affect the target selection by the

SB transposase rather than a specific base-pair specific interactions[24]. The hypothesis is that combination of a spatial optimum conformation and of a specific hydrogen bonding capacity recruits the transposase with a substantial degree of specificity[24]. In addition, as long-term genomic resident coevolving with their host, transposase are expected to have developed a number of interactions with host proteins. SB transposase activity is drove by a repair protein Ku70, the HMGB1 and the transcription factor Miz-1[39, 41, 42]. Other factor that can influence the integration efficiency is the cargo size of the transposon. It has been observed that, starting from a minimum transposition cassette size of ~2.5Kb, each increase of 1Kb to the length of the transposon results in a ~30% drop in transposition efficiency. The distance separating the transposon's ends in a plasmid construct is also an influencing element for it was proved that shortening the plasmid's backbone determines an increase in transposition efficiency[43]. This is compatible with the hypothesis that transposases bound to one IR are engaged in a bidirectional search along the DNA molecule for the other IR. It also explains the lower transposition rates observed for transposons inserted in linear constructs[37, 44].

Yet, SB transposition seems to be affected by "Overproduction Inhibition", a phenomenon in which expression of the transposase over a certain threshold (whose level is variable and influenced by various cellular factors) impairs transposition. This inhibition seems to be due to an inefficient folding of the tetramer structure of the transposase with transposon DNA[37, 38, 44].

Integration profile analysis

A huge branch of the scientist's efforts focused on mapping the integration pattern of the most used integrating vectors such as Human Immunodeficiency Virus and Moloney Leukemia Virus[45-47]. Several concerns raised from the observation of what these integration events cause in treated cells. Retroviruses have an attraction for active genes and actually perturb their activity in different way: the viral promotor placed in the LTR could activate transcription of neighbor genes, cryptic splice sites could disrupt the exon of the hit gene causing a truncated protein.

The same questions were addressed to this new category of integrating non-viral vectors; the DNA transposon/transposase. The Sleeping Beauty Transposase integrates into a TA dinucleotide that is duplicated in the last step of transposition. There are about 200 million TA sites in the human genome, which is appealing for obtaining a wide distribution of integrated transposon. However not all TA sites are equally targeted. SB transposon appear to prefer "flexible" TA sites[24]. Sequence flanking the TA dinucleotide play a crucial role in site selection. It has been shown that SB prefers insertion into an AT palindromic repeat of 8bp: ATA**TA**TAT (bold TA represent the target site). Of particular importance is the fact that the first and last bases of this short sequence appear to be particularly conserved for they are found in 66% and 70% of the target sites, respectively. Affecting target site selection are also the physical properties of the

sequences surrounding the TA dinucleotide. First of all, a highly bendable structure may allow the transposase and other factors involved in transposition to deform the DNA molecule in order to achieve the correct spatial conformation necessary for the transposition reaction. The propensity of DNA to assume an A-DNA form (A-philicity) is another important physical factor, for the wide and shallow minor grove of this DNA structure allows proteins to easily access the DNA helix[24].

Overall, SB integration pattern was shown to be close to random. In 3T3 mouse cells and Huh-7 human cell lines were retrieved firsts statistics on transposon distribution[48]. In this work, Yant et al., compare SB insertion with a random library of 10000 reads generated in-silico and they found a close to random pattern with some enrichments in microsatellite repeated region and a under-representation in LINE-L1 repeats.

Another work compare integration patterns in human T-cells of SB transposon with other enabled transposase: Tol2 and piggyBac (PB)[25]. It was demonstrate that Tol2 and PB integration preferences resemble those seen in retroviral vectors: both preferentially integrate near Transcriptional Start Sites (TSSs), CpG islands and DNaseI hypersensitive sites. Again, it was confirmed the random distribution of SB transposon[25]. However potential genotoxic risks associated with SB insertions cannot fully discharged. The possibility of remobilization of a transposon from residual transposase activity is a theoretical concern. The SB transposase gene is provided in cis under a CMV promoter and its expression from an episomal plasmid was

shown to last few days after transfection. Nevertheless, the possibility of a few cells continuing to express transposase cannot be ruled out. When a SB transposon is re-mobilized, it will leave a foot-print of about 5 bp in most of the cases[40](Figure 5). In a exon this would result in a frameshift leading to an abnormal protein. But, exons represent no more than 2% of the human genome, the rate of excision of transposon in a cell is less than 10^{-4}, and moreover SB do not prefer transcriptional units, so the chance of an adverse event is estimated to be less than 10^{-6}[49]. To reduce the potential dangerous insertions risk, the SB system was modified to introduce a bias in the landing pattern. The transposase coding sequence was fused with elements that are able to bind a specific sequence in the human genome[50, 51]. These recent works show that is possible to retarget the integration pattern in safe loci but also that efficiencies of transposition and retargeting were not sufficient to already express their potential in Gene Therapy (GT).

Sleeping Beauty applications

A DNA transposon active in vertebrate cells gives the opportunity to enhance the biological studies and to raise other questions to the unknown molecular mechanism. With the knowledge provided by the Human Genome Project the scientific community wondered about identification of gene function. Several methods were employed to address this task, including mutational analysis. There are different strategies to create mutations for example

insertional mutagenesis by discrete pieces of foreign DNA. The random integration pattern provided by the SB transposon, combined with its ability to efficiently integrate versatile transgene cassettes into chromosomes established this system as an extremely useful tool for insertional mutagenesis in stem cells[52], somatic cells[53] and in animal models[54]. Different kind of constructs can be delivered within the transposon according to the goal:

-5' gene-trap splice acceptor and/or a polyadenylation sequences to disrupt the coding sequence upon insertion into introns[55];

-strong promoter/enhancer activity sequences to induce a gain-of-function mutation and drive transcription outwards from the vector[56]. Type of tool used in large-scale cancer gene discovery screens in experimental animals[57];

-short-hairpin RNA expression cassette to obtain stable knockdown cell lines by RNA interference. This tool was evaluated as a potential approach to the therapy of acquired immunodeficiency syndrome by knocking down CCR5 and CXCR4 cell surface antigens required for the viral HIV entry[58].

Moreover, DNA transposons are an attractive tool for the induced Pluripotent Stem Cells (iPSCs) field because of their capacity to stably integrate in safer loci compared to the retroviral vectors and the re-excision possibility produced by transient re-expression of the transposase after the completion of reprogramming[59, 60].

However, the DNA transposon tool has seen a growth in interest for treatment of several diseases. It allowed the expression of

human α-1-antitrypsin cDNA for more than six months in mice and restored therapeutic levels of FIX in a mouse model for Hemophilia B[44, 61, 62]. In 2005, the SB system was used to mediate in vivo stable co-expression of two anti-angiogenesis factors, the statin-AE and sFlt1, in a Glioblastoma multiforme-derived cell line. One-three weeks after implantation of this cell line in nude mouse strains, tumoral mass arose. When a transposase-encoding plasmid was co-transfected with the SB vector in the tumor mass, complete regression of the tumor was observed in a significant percentage of the treated mice, indicating the potential application of this approach in an anti-angiogenic GT trial[63]. In a 2006 study, Liu et al. used a SB vector encoding for human indoleamine-2,3-dioxygenase (an enzyme with T-cell suppressive and anti-oxidant activity) to prevent chronic lung allograft rejection in rats, a process characterized by progressive accumulation of collagenous tissue and persistence of fibroblasts in the airways[64]. This approach was indeed successful in preventing allograft rejection as shown by the normal pulmonary functionality and histological appearance along with reduced collagen content. Moreover, this approach can possibly be used in the treatment of other disorders accompanied by fibrosis, such as mucoviscidosis. Finally, the first clinical trial performed with the SB system aimed to generate patient-derived T cell stably expressing a Chimeric Antigen Receptor (CAR) to redirect their specificity against CD19 in patients suffering B-cells lymphomas[65, 66]. The CAR is composed by a murine single-chain variable fragment, which mediates specific cytoxicity

toward CD19+ lymphoid tumor cells. In this study, T cells were purified by peripheral or umbilical cord blood and, after transfection with a CAR encoding SB vector, the cells were expanded on CD19+ artificial Antigen Presenting Cells in order to enrich the population with CAR+ T lympocytes[66]. A clinical study is currently evaluating safety and feasibility of autologous transplantation of modified T lymphocytes after chemotherapy for the treatment of B lymphoid malignancies[67]. The use of Sleeping Beauty system was applied in other disease model such as tyrosinemia[68], Huntington disease[69], sickle cell disease[70], mucopolysaccharidosis[71] and type I diabetes [72]. SB applications are expanding and several issues yet demand characterizations for further uses.

Gene Therapy

Gene therapy aims to correct the disease process by restoring, modifying or enhancing cellular functions through the introduction of a functional gene into a target cell. Under this definition almost any disease may, at least in theory, be a candidate target of gene therapy. The concept of gene therapy was born at the end of the 80s, together with new progress in molecular biology and the development of efficient techniques for gene transfer. In its initial formulation, gene therapy was supposed to be used in order to transfer normal copies of a mutated gene in its appropriate tissue with the aim of curing monogenic hereditary diseases with recessive phenotype (such as

cystic fibrosis, hemophilia, muscular dystrophy); however today the repertory of therapeutic genes includes also gene sequences or DNA fragments that try to influence and regulate the cellular functions or the host immune system. Consequently, the therapeutic potentiality of gene therapy also extends to new and highly diffused acquired diseases, such as tumors, cardiovascular and infectious diseases and neurodegenerative disorders.

Since the first gene therapy trial was approved in 1990, over 1800 clinical trials have been completed, are ongoing or have been approved worldwide. Almost two-third of all the trials performed have been carried out in the United States. Within Europe, the United Kingdom accounts for 11% of the world total with more than 200 trials, Germany 4,4%, France 2,9%, Switzerland 2,7% and Italy 1,1% (Figure 6). The vast majority of gene therapy clinical trials have addressed cancer (64,4%), followed by inherited monogenic disease (8,7%), such as cystic fibrosis, α1–antitrypisin deficiency severe combined immunodeficiency (SCID), and cardiovascular disease (8.4%). Retroviral vectors and adenoviral vectors have, so far, been the most commonly used vectors in gene transfer trials, while non-viral gene transfer has been assessed in roughly one quarter of all trials (Figure 6). Most trials are designed to assess only the safety of a particular gene therapy approach (Phase I) and only few gene therapies are being assessed in Phase II or Phase III efficacy trials.

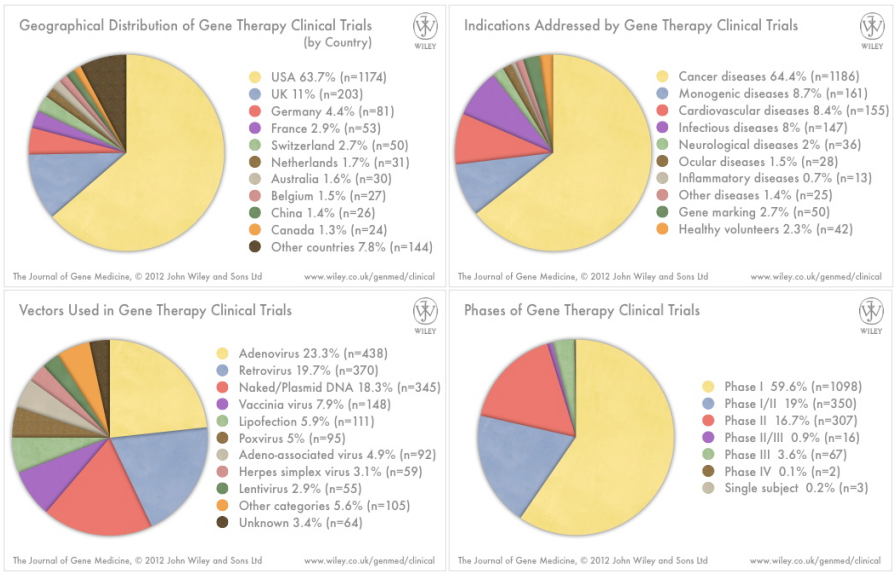


Figure 6. A survey of gene transfer clinical trials. Journal of gene Medicine clinical trials website: www.interscience.wiley.com

Ideally, gene therapy should be efficient, cell specific and safe; the many steps important for a successful approach include: a near understanding of the pathogenesis of disease; an appropriate target tissue of gene delivery; an effective therapeutic gene and an animal model that closely simulates disease for preclinical testing. At present, gene therapy is being contemplated only on somatic cells, although many efforts are being directed to the isolation and generation of pluripotent cells (iPSCs[59]) and how to efficiently differentiate them in the desired cell type avoiding hyper-proliferation of undifferentiated cells. Although many somatic tissues can receive therapeutic DNA, the choice of cell usually depends on the nature of the disease and on consideration such as the efficiency of gene delivery, protein modifications, immunological status, accessibility and economics. The rate-limiting step for successful

gene therapy, however, is the ability to transfer efficiently the appropriate therapeutic gene to the target tissue and to obtain sustained expression[73]. Although the nucleic acids containing the genes can be generated in the laboratory with relative ease, the delivery of these materials into a specific set of cells in the body is far from simple. Various methods have been developed for this purpose. At present there are two main categories of delivery vehicles: viral and non viral vectors [74]. Non viral vectors comprise direct injection of naked DNA or DNA associated with polylysine, cationic lipids or other physical methods, that allow the gene to cross the cell membrane. The simplicity of this system has made it a popular non viral system in clinical trials, particularly in vaccination protocols. The major problems with nonviral, plasmid-based systems are as follows: (I) low rates of delivery of the vectors to target-cell nuclei; (II) low rates of integration of transgenes, (III) integration of concatemers copies of the transgene. These problems can be alleviated by using transposons, which showed high integration rate in several cell type.

Nevertheless, most of the current gene therapy applications are based on a second vector category: viral vectors. Many viruses have evolved specific machinery to delivery DNA to cells or to integrate into the cell genome. Viral vectors are the most suitable vehicles for high-efficiency gene transfer and sustained and often high-level expression of the transgene. At present, there are three main vector systems (retroviral, lentiviral and adeno-associated vectors) that can integrate into recipient cells, and at least two

vector types (adenoviral and herpes simplex-based vectors) that are essentially maintained as episomal genomes. Retroviral vectors derived from murine retroviruses were the first vectors used in gene therapy and still remain the most commonly used integrating vectors in clinical trials.

For gene therapy success, an appropriate amount of a therapeutic gene must be delivered into the target tissue without substantial toxicity. For some disorders, long term expression from a relatively small proportion of cells will be sufficient, whereas other pathologies might require high, but transient gene expression. Some forms of gene therapy will also require regulated gene expression. Different strategies for transcriptional targeting have been proposed and validated, some based on the introduction of internal heterologous promoters, some based on the replacement of viral regulatory elements with tissue specific enhancers or promoters[75] [76].

Recently RNA-based therapeutic technologies have been developed[77]; they comprise RNA interference, antisense oligonucleotides and steric-blocking oligonucleotides that bind to a complementary RNA sequence and act through different mechanisms directly on the pre-mRNA and mRNA. Through this mechanism, oligonucleotides can redirect alternative splicing, repair defective RNA, restore protein production or downregulate gene expression. Moreover, they can be extensively chemically modified to acquire more drug-like properties. The ability of RNA-blocking oligonucleotides to restore gene function makes them suited for the treatment of genetic disorders and the positive

results achieved in clinical trials for the treatment of Duchenne muscular dystrophy[78-80] highlight the promising potential of this approach.

Non-viral integrating proficient system

In addition to Sleeping beauty, other efficient transposon/transposase system have been developed: Tol2 and piggyBac.

Tol2

This DNA-TE was discovered in 1996 in the Medaka genome. It belongs to the hobo-Activator-Tam3 (hAT) family and is the first naturally occurring active transposon ever to be isolated from vertebrates.

A notable feature is the ability to mobilize relatively large constructs of foreign DNA that allows Tol2 to carry transgenes of up to 11Kb with minimal loss of transposition efficiency[59, 81]. Recent experiments showed that this transposon system is even able to deliver transgenes of about 70Kb in Zebrafish and mice, proving the exceptional large cargo capability of Tol2.

Also, transposition was found to be directly proportional to transposase protein levels. Therefore, this system does not seem to be affected by Overproduction Inhibition, a phenomenon typical of other transposase systems (e.g. Sleeping Beauty) in which an overproduction of the transposase results in impaired

transposition capability. One of the major drawbacks of the Tol2 system is the relatively low transposition efficiency as compared to other transposons such as piggyBac or Sleeping Beauty[25].
The integration profile of Tol2 transposon resembles that of MLV derived γ-retroviral vectors. In fact, though no consensus sequence was found at target sites, Tol2 transposition is not random and a marked preference for integration near TSSs, DNase I hypersensitive sites and CpG islands was observed[81]. Integration within cancer-related genes was also found to be higher than with other transposon systems. These preferences result in a higher probability to cause insertional mutagenesis as compared to other TEs. In fact, oligoclonal expansion was more evident in T cells transfected with Tol2 than with other systems (e.g. Sleeping Beauty). Moreover, since modification of Tol2 transposase at its N-terminal domain abolishes transposition, retargeting of this system to a safe spot in the genome has proven quite challenging.

PiggyBac

PiggyBac is a naturally occurring type II transposon isolated from the *cabbage looper moth Trichoplusia ni*. Although insect's genome is relatively distant to that of mammals, piggyBac was shown to efficiently transpose in a wide range of mammalian cells. Moreover, transgene silencing upon integration was lower than those of other transposons such as Tol2. Therefore, piggyBac can

achieve efficient stable long-term transgene expression, rendering this transposon-based system one of the most promising to date.
Another important feature is the ability to transpose foreign DNA sequences of up to 9.1Kb without loss of efficiency[81].
PiggyBac target site is characterized by a consensus sequence comprising a TTAA tetranucleotide. The first 100bp flanking this tetranucleotide on either side were also found to be important and, in fact, even minimal changes in these 200bp can result in reduced site preference. Similarly to Tol2, piggyBac's integration profile is not random, showing a marked skewing for CpG islands, TSSs and DNase I hypersensitive sites, arising the risk of insertional mutagenesis[25]. This is expecially troubling if we acts as an enhancer element[82]. However, targeting strategies can be developed in order to promote integration of piggyBac into safe harbors of the host genome. This is possible due to the fact that piggyBac transposase, contrary to those of other systems (e.g.Tol2 and Sleeping Beauty), can be molecularly modified without substantial loss of transposition efficiency. For example, by fusing the Gal4 DNA-binding domain to either the N- or C-terminus of piggyBac transposase, enrichment of insertions at Gal4 recognition sequences (endogenous or exogenous) was observed[83].
Another issue of this transposon system is the fact that the piggyBac family of TEs is abundant in the human genome with about 2000 piggyBac-like elements. Although the evolutionary distance between moths and humans is remarkable, the

possibility that these elements could be mobilized by an exogenous source of piggyBac transposase cannot be excluded.

In a 2007 study was demonstrated the possibility to create inducible forms of the piggyBac system by fusing the N- or C-terminal domain of its transposase to a modified version of the estrogen receptor ligand-binding domain[82]. In the absence of the inducing agent 4-hydroxytamoxifen, transgene expression resulted 800-fold weaker than in the presence of the inductor. This inducible system can be useful in those cases were transgene expression needs to be carefully regulated.

PiggyBac was also found to be a useful tool in the production of induced Pluripotent Stem (iPS) cells [59]. Murine and human embryonic fibroblasts were indeed successfully induced to pluripotency by a piggyBac transposon carrying doxycycline-inducible transcription factors[60]. In another study, Yamanaka's factors (c-Myc, Klf4, Sox2 and Oct4) linked with 2A peptides were provided to mouse and human fibroblasts thanks to a piggyBac vector. After pluripotency the factors were removed thanks to a second administration of a transposase encoding plasmid, leaving no trace of the transposon in the reprogrammed cells' genome. This approach was therefore successful in minimizing genome modifications of iPS cells and in achieving complete removal of possibly dangerous exogenous genes.

It is clear then how piggyBac can be a very useful tool for GT and for biotechnological application.

Retroviral vectors

Retroviruses are lipid-enveloped particles comprising a homodimer of linear, positive-sense, single stranded RNA genomes of 7 to 11 kilobases. Following entry into target cells, the RNA genome is retro-transcribed into linear double-stranded DNA and integrated into the cell's chromatin. This family of viruses includes several varieties being exploited for gene therapy: the mammalian and avian C-type retroviruses (also referred as to *oncoretroviruses*), lentiviruses and spumaviruses. All retroviral genomes have two *long terminal repeat* (LTR) sequences at their ends. LTR and neighboring sequences act *in cis* during viral gene expression, packaging, retro-transcription and integration of the genome. The LTR sequences frame the tandem *gag, pol* and *env* genes encoding the structural proteins, nucleic acid polymerase/integrase and surface glycoprotein, respectively [84]. The location of most *cis*-acting sequences in the terminal regions has enabled simple and effective retrovector design, making them the most widely used vector system in gene therapy trials to date [85]. Most of the retroviral vectors currently in clinical use are based on Moloney Murine Leukemia Virus (MMLV), a well-studied and characterized retrovirus [86, 87]. Some encouraging clinical results of gene therapy have been obtained with these vectors, representing the starting point for the treatment of several diseases such as: Graft-Versus-Host disease (GVHD) [88], Ada-SCID [89, 90], X-SCID [91, 92], CGD [93], and Junctional Epidermolysis Bullosa (JEB) [94].

The production of retroviral vectors requires two components: a *transfer vector* and a *helper construct*. The first contains the gene of interest (up to 8kb in length) and cis-acting sequences located in the LTRs and in the packaging signal (Ψ), while the latter is designed to express viral genes lacking in the vector backbone. To improve the biosafety of the system, the viral genes of the helper construct are expressed from two different plasmids (the *packaging construct* expressing *gag* and *pol,* and the *envelope* construct expressing *env*) and by heterologous transcriptional signals; this minimize the risk of producing Replication Competent Retroviruses (RCR) by recombination events. These three plasmids can be used in transient transfection in 293T cells to achieve rapid propagation of the retroviral vector; alternatively, helper functions can be stably provided from a packaging cell line. The packaging cell line should be chosen with respect to the tropism of the virus generated.

A useful property of retroviral vectors is the ability to integrate efficiently into the chromatin of target cells. Although integration does not guarantee stable expression of the transduced gene, it is an effective way to maintain the genetic information in a self-renewing tissue and in the clonal outgrowth of stem cell. Disruption of the nuclear membrane is required for the preintegration complex to gain access to the chromatin and productive transduction by retroviral vectors is strictly dependent on target cell mitosis shortly after entry. Because only a fraction of cells pass through mitosis at any given time, this severely limits the applications of retroviral vectors in gene

therapy to selected target *ex vivo,* since the cells are isolated and propagated in culture, then infected and transplanted to the patient.

There are mostly three general classes of retroviral vectors. The first type of vector is the LTR- based vector in which the gene of interest is expressed directly from the LTR. The second is an internal promoter vector in which the gene of interest is driven from an internal promoter, and the wild type LTR can be retained in order to express a second transgene or abolished at 3' to generate the Self Inactivating (SIN) vector. Finally, there is a so called reverse orientation vector in which the gene of interest is expressed from its own promoter in reverse orientation. This type of vector is used for insertion of genomic sequences with introns, a process that allows for more appropriate regulation of gene expression.

Lentiviral vectors

Lentiviruses (e.g. human immunodeficiency virus type 1, HIV-1) [95, 96] are complex retroviruses with the ability to infect both proliferating and quiescent cells. Unlike oncoretroviruses, they rely on active transport of preintegration complex through the nucleopore by the nuclear import machinery of the target cell. Thus, a possibly unique application of lentiviral vectors is the transduction of quiescent long-term repopulating stem cells. Like the simple retroviruses, HIV has the three *gag, pol* and *env* genes, but it also carries regulatory genes, tat and rev, and accessory genes, *vpr, vpu, nef* and *vif*.

Replication-defective vectors were originally derived from HIV-1 but hybrid lentiviral vectors have also been derived from non human lentiviruses (for example simian, equine, feline, caprine and bovine).

Vector particles are produced by transient cotransfection of the three constructs (transfer vector construct, packaging construct, and envelope construct) in 293T cells. In a first generation of HIV-derived lentiviral vectors [97], viral particles were generated by expressing the HIV-1 core-proteins, enzymes and accessory factors from heterologous transcriptional signals in the packaging construct and the envelope of another virus, most often the G protein of the vescicular stomatitis virus (VSV-G) from a separate envelope construct. In a second version of the system, the HIV-derived packaging component was reduced to the *gag*, *pol*, *tat*, and *rev* genes [98]. In either case, the vector carries the HIV-derived cis-acting sequences necessary for transcription, encapsidation, reverse transcription, and integration. A representative vector comprises, from the 5' to the 3' end, the HIV 5' LTR, the leader sequence and the 5' splice donor site, approximately 360 bp of the *gag* gene (with the *gag* reading frame closed by a synthetic stop codon), 700 bp of the *env* gene containing the RRE and a splice acceptor site, an internal promoter, the transgene, and the HIV 3' LTR. The third generation lentiviral vector conserves only three of the nine genes of HIV-1 and relies on four separate transcriptional units for the production of transducing particles [99]. This eliminates the possibility that a wild-type virus will be reconstituted through

recombination. The function of Tat becomes dispensable if part of the upstream LTR, in the transfer vector, is replaced by a constitutively active promoter, and *rev* gene can be deleted from the packaging construct and expressed *in trans* from a separate plasmid.

This system would be further improved if the transcriptional elements of HIV were removed from the transfer vector [100], allowing the construction of SIN vectors.

Gene therapy in Epidermolysis Bullosa disease

Epidermolysis bullosa (EB) constitutes a group of dermal genetic diseases, defined genodermatoses, manifesting with fragility of the skin and mucous membranes and presenting with blisters and erosions at birth or shortly thereafter [101, 102]. The spectrum of phenotypic manifestations is broad; in the milder forms there is a life long blistering tendency with no impact on the overall longevity of the affected individual, while in the most severe forms children die during the early postnatal period from metabolic perturbations, dehydration, and sepsis. Some forms are characterised by debilitating scarring with the propensity to early death from aggressive squamous cell carcinomas of the skin. In addition to skin and mucous membrane involvement, there are a number of extracutaneous manifestations. In some cases, abnormalities in the hair, nails, and teeth can be described. In other forms of the disease, the gastrointestinal tract is affected either in the form of oesophageal strictures or congenital pyloric

atresia. A subset of patients manifests with late onset muscular dystrophy. Adding to the heterogeneity of EB is the fact that inheritance can be either autosomal dominant or autosomal recessive. This heterogeneity, has led to suggestions that there may be as many as 30 different subtypes of EB [103]. However, with the molecular technical improvement, now is possible to discriminate EB from mutations in 10 different genes expressed within the cutaneous basement membrane zone (BMZ) at the dermo-epidermal junction [104]. The stratified expression pattern of these genes within the BMZ and in extracutaneous tissues explains the broad phenotypic spectrum of EB. Traditionally, EB has been divided into three broad categories based on the level of tissue separation, determined by diagnostic electron microscopy:
the **simplex form** of EB (EBS) demonstrate tissue separation within the basal keratinocytes at the bottom layer of epidermis; pathogenic mutations reside in the basal keratin genes, K5 and K14. The mutations in these genes are usually inherited in an autosomal dominant manner and result in substitution of conserved amino acid residues, leading to deleterious dominant negative effects on the assembly of keratin intermediate filaments.
The **junctional form** of EB (JEB) display tissue separation within the dermo-epidermal basement membrane, within the lamina lucida; the disease includes family of autosomal recessive mutations. The most abundant is caused by mutation in any of these three genes *LAMA2*, *LAMB3* and *LAMG2* encoding the the laminin 332 heterotrimer.

The **dystrophic form** of EB (DEB) is characterised by tissue separation below the lamina densa within the upper papillary dermis, due to abnormalities in anchoring fibrils; the characteristic genetic lesion is found in the COL7A1 gene which results in reduced production of type VII collagen. This form of EB can be inherited in either autosomal dominant or recessive pattern.

In addition to this traditional classification, a fourth subtype, the **hemidesmosomal variants** of EB (HEB) has been proposed; patients with HEB manifest tissue separation at the basal cell/lamina lucida interface at the level of the hemidesmosomes, caused by mutations in the type XVII collagen and plectin genes [105].

The lack of effective treatments for these diseases represents a challenge for investigators and clinicians in the field. The understanding of the molecular basis of genodermatoses has paved the way for gene therapeutic approaches aiming at functionally correcting the diseases at the genetic level [106-111]. Epidermis is an ideal target for *ex vivo* gene therapy. Skin is easily accessible to manipulations, and consolidated procedures have been established for growing and expanding keratinocytes *in vitro* to form transplantable epithelial sheets that maintain the characteristics of authentic epidermis upon transplantation.

The first successful attempt to introduce and express foreign genes into transplantable human keratinocytes by retroviral-derived transfer vectors dates back to 1987 [112]. Since then, several groups have shown that genetically modified

keratinocytes can express and secrete a variety of transgene products, from apolipoprotein-E to clotting factor IX and VIII and interleukin-6, both *in vitro* and *in vivo* in animal models [113-116]. Successful genetic correction of epidermal skin disorders requires fulfilment of specific criteria, including long-term expression of the therapeutic transgene in the transplanted tissue. The feasibility of *ex vivo* gene therapy applications for EB has been definitely confirmed by the successful results obtained in a pilot, phase-I clinical trial for the *LAMB3*-deficient form of JEB [94]. This trial was aimed at assessing the overall safety of the epidermal transduction/transplantation procedure, analyzing long-term survival of transduced stem cells and persistence of transgene expression, and monitoring humoral and/or cytotoxic immune responses against the genetically modified cells.

MLV-derived SIN retroviral vectors have instead been developed for the expression of the COL7A1 cDNA [117]. Even though the SIN design improves their safety characteristics, retroviral vectors cannot provide a solution for all EB disorders. In particular, developing a gene transfer strategy for DEB appears to be a formidable challenge. The COL7A1 cDNA exceeds 9kb in length, a size that is hardly accommodated by a retroviral genome; titer and genetic stability due to both the size and the highly repeated nature of the cDNA sequence are persistent problems that hamper a real clinical development of this type of vectors. Woodley and co-workers delivered and expressed full-length type-VII collagen using a SIN lentiviral vector into recessive DEB keratinocytes and fibroblasts, demonstrating correction of cell phenotype and

formation of anchoring fibrils at the dermal-epidermal junction *in vivo* [118]. Similar results were obtained by the same group when the lentiviral vector was intradermically injected in mice [119] or used to transduce DEB fibroblasts, subsequently injected intradermically [120] or intravenously [121] in murine wound healing models. Although these reports showed correct deposition of type-VII collagen at the BMZ and formation of anchoring structures in pre-clinical models, in none of them did the authors investigate rearrangement of the COL7A1 cDNA after integration. As reported by Titeux *et al*, the use of a SIN retroviral vector for transferring the COL7A1 cDNA in DEB keratinocytes and fibroblasts led to correction of the genetic defect in human skin equivalents, but also to the formation of rearranged proviruses in one third of the cases [117]. The presence of rearranged copies of the integrated vector and, consequently, the production of mutant proteins should be taken into account before embarking into clinical studies. Abnormal proteins represent both a serious limitation of the therapeutic effect and a potential hazard. To overcome this problem, COL7A1 "minigenes" have been developed which maintain proper biochemical functions in vitro, and could potentially fit into a retrovirus-based gene transfer vector [122]. However, the efficiency of proteins of reduced size in correcting a collagen deficiency has yet to be demonstrated. Lastly, the recent improved DNA transposon system could solve most of the issues carried by a retroviral vector integrational mutagenesis, sequence rearrangements of COL7A1 cDNA and finally, transfer efficiency rate[106].

Scope of the Thesis

The Sleeping Beauty (SB) transposon-based integration system is a valuable tool for functional genomics in several model organisms and represent a promising vector for gene therapy in humans. The SB transposase was found to bind the inverted repeats of regenerated salmonid transposons in a substrate-specific manner, and to mediate precise cut-and-paste transposition in fish as well as in mouse and human cells. However, a major bottleneck of any transposon-based application is the efficiency of transpositional activity. Therefore, considerable effort has been made to improve the transposition efficiency of SB by modifying its IRs and systematically mutating the transposase gene. The combined effect of these modifications resulted in the generation of the hyperactive SB100X transposase and of the high-capacity "sandwich" transposon (SA). This new system shows an almost 100-fold enhancement of transposition as compared to the first-generation transposon system. I address my work to the molecular characterization of "sandwich" SB-mediated integrants in epithelial cell lines and in primary keratinocytes. The interest in keratinocytes is related to a SB-mediated gene therapy application for the dystrophic form of epidermolysis bullosa (DEB) caused by mutations in the type-VII collagen gene (COL7A1). Delivering the >9 kb COL7A1 cDNA by a retroviral or lentiviral vector is problematic, due to the large size and highly repeated nature of its sequence, which induce genetic

rearrangements during reverse transcription and integration. The Sleeping Beauty transposon-based integration system can potentially overcome these issues.

In chapter I, I compared the “sandwich” (SA) and the conventional (T2) transposon, co-transfecting the SB100X transposase with the pT2 or the pSA transposon plasmid carrying a reporter gene and stuffer DNA to characterize the transposition efficiency of the two constructs with increasing cargo payload.

In chapter II, I deeply investigated the molecular characteristics of T2 and SA construct through cell clonal genomic analyses. In particular I investigated these parameters:

- Persistence of SB100X and transposon carrying plasmids
- Number of integrated transposons per cell
- Correlation between the expression level and the copy number
- Transposon genomic integrity
- Reliability of SA transposition mechanism

In chapter III, the high-resolution genome wide mapping of SA integrations was utilized to evaluate its genotoxicity in comparison to the other DNA transposon and to the commonly used integrating retroviral and lentiviral vector.

Chapter I

Integration efficiency of T2 and SA transposons

The “resurrection” of the Sleeping Beauty transposon/trasposase system, edited by the group of Zoltan Ivics and Zsuzsanna Izsvak[27, 29], was revolutionary. The possibility to manipulate naked or plasmidic DNA and easily integrates it into vertebrate cells open a new field in the biotechnologies. The scientific community had fully understood its potential and worked to ameliorate the activity of this DNA transposon based system. T2 transposon come in the 2002 [28] from observations and experiments done on the Inverted Repeats of the first wild type transposon salmonid-type TcE from Tanichthys albonubes. SA transposon comes later in 2004 [31] from the recreation of a Tc1/mariner Paris transposon structure that was able to transpose element with a size greater than 10 kb.

These two kinds of transposon were never placed in a direct comparison to date.

Transposon fleet

Since our goal is to efficiently transpose a 9kb gene coding for the collagene VII in immortalized and primary keratinocytes, we first characterize the transposition efficiency of the two constructs with increasing cargo payload. In order to retrieve these data, I first cloned an expression cassette in between the T2 and SA IRs. The expression cassette carries: an ubiquitous promoter, designed

CAGGS, resulted from the fusion of the Chicken B-Actin promoter and of the CMV early enhancer element; a reporter yellow fluorescent protein (Venus), GFP derived; and a SV40 poly-Adenilation signal. To increase the transposon cargo payload, I inserted between the two IR/DR increasing size of a stuffer DNA (Figure 7) amplified from the first intron of the HPRT gene. I cloned three sized-different transposon plasmids for both T2 and SA constructs collecting my transposon fleet. For an easier interpretation, I named the plasmid pT2 and pSA with the transposable cassette size expressed in kilo bases (Figure 7):

pT2 3.2kb; pT2 10kb; pT2 14kb

pSA 5.8kb; pSA 9,7kb; pSA 14kb

Epithelial cells transfection rate

The disease model of Epidermolysis Bullosa, based on epidermis impaired adhesion onto derma layer, drove me to use epithelial cells in all of my experiments. HeLa cells, GABEB cells (immortalized keratinocytes derived from patients affected by Generalized Atrophic Benign Epidermolysis Bullosa) and primary keratinocytes, derived from foreskin biopsy of normal donor, were transfected to measure the transposition rate of the SB system in those cells. In spite of HeLa, immortalized GABEB and primary keratinocytes are resistant to any kind of transfection procedures. Nucleofection, CaPO4, and liposomes transfection method were used with very poor results. Only a commercial

liposomes-based reagent from Roche, the Fugene HD®, could efficiently transfect GABEB cells and keratinocytes.

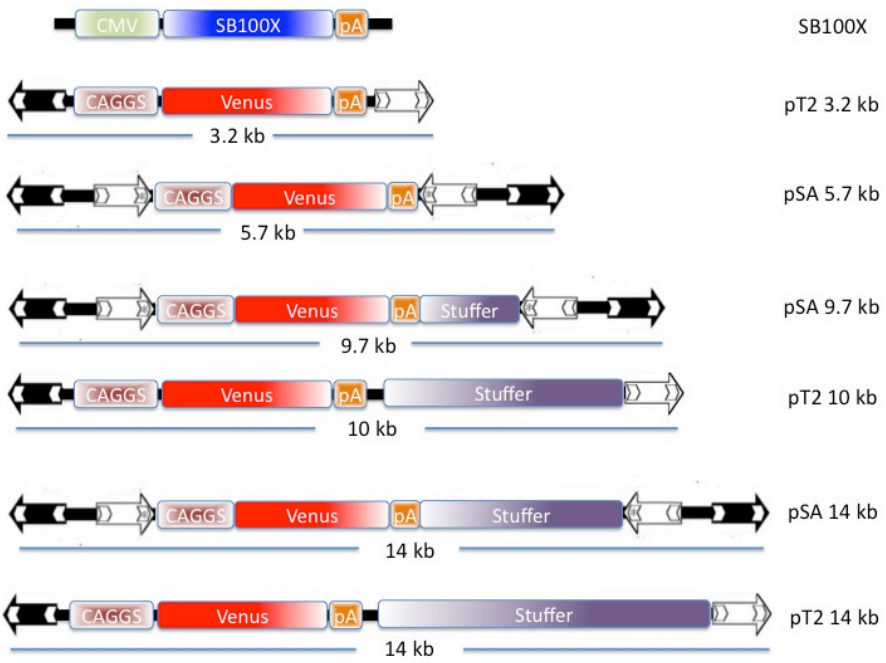


Figure 7. Transposon fleet. The transposons used in this work are represented here. The transposase expression cassette is schematized upper. The SB100X coding sequence is drove by a CMV promoter and followed by a SV40 poly-Adenilation signal. The transposons are characterized by the presence of the IR/DR ends that form a "Sandwich" like conformation in the case of SA engineered construct. All of them possess the expression cassette with the CAGGS promoter, VENUS reporter gene and polyA signal. The stuffer DNA represented has variable increasing size

Transfection of SB100X and transposon plasmids were optimized with several ratio of liposomes/ug of DNA for every cell type. Moreover, the optimization step passed through retrieving the best ratio between the SB100X plasmid and transposon in order to obtain the higher transposition efficiency. The higher

transposition rate was obtained with a stoichiometric ratio of 1:1, or 2:1 (transposon/transposase) in case of transposon plasmids bigger than 9kb. The transposition rate was measured as follow:

I co-transfected HeLa and GABEB cells with SB100X encoding plasmid (from now on referred to as pSB100X) and either pT2 or pSA transposons carrying increasing cargos. In this contest we would better depict how the integration efficiency drop down by increasing the transposon length in both IR backbones.

As control I used: a mock transfected HeLa and GABEB cells and cells transfected with the pT2 Venus or pSA Venus constructs alone. These last controls were needed to verify the role of the transposase SB100X. In fact, without the expression of the transposase, no transposition event should occur and residual reporter gene expression after long (>20 days) periods would only be attributable to noise or to rare random plasmid integration events.

In order to reduce variability due to transfection procedure, I performed 3 independent experiments for each cell type. Transgene expression was measured via flow cytofluorimetry starting from 48 hours post transfection. The Venus positive cells were periodically measured to follow the trend of the signal that persists in presence of SB100X and drops without the transposase as shown in Figure 8.

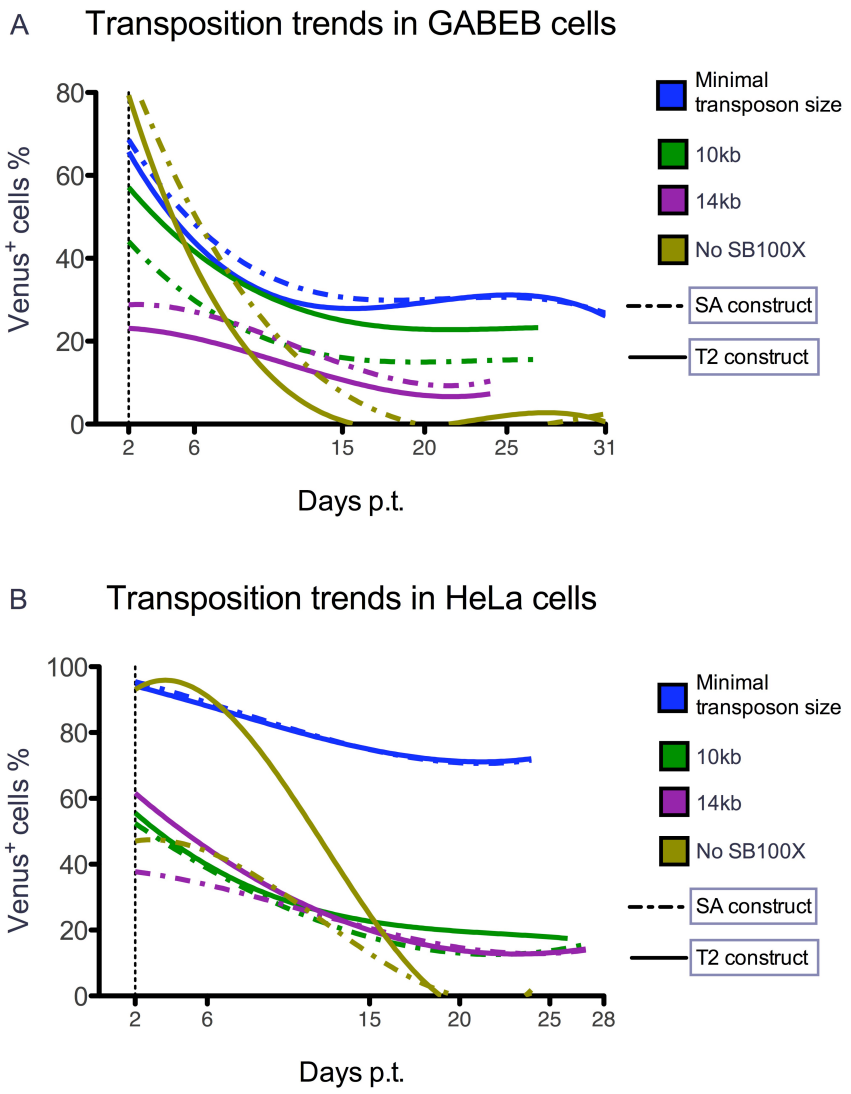


Figure 8. Transposition trend. The days post transfection (p.t) are plotted on the X axis while the percentage of positive Venus cells are represented on Y axis. At the time point of 48 hours p.t. is possible to achieve the maximum expression from a transfected reporter gene placed under a constitutive promoter. In blue (minimal transposon size) are schematized the pretty overlapping samples co-transfected with SB100X and pT2 3.2 kb or the pSA 5.7kb, in green the sample co-transfected with pT210 or pSA 9.7 and pSB100X and purple for samples co-transfected with pT2 14 or pSA14 and pSB100X. In gold are represented the negative controls transfected with pT2 3.2 or pSA5.7 alone, without the SB100X plasmid. A) Trends of transposition in GABEB cells. B) Trends of transposition in HeLa cells.

Looking at the efficiency of transfection and the efficiency of transposition, it's possible to appreciate first of all the lower transfection efficiency in GABAB cells respect to HeLa cells even with pretty short transposons. This evidence is in line with their derivation of primary immortalized keratinocytes. Moreover, as expected, the transfection efficiency is affected by the increased size of transposon plasmid in both cell types. The absence of the SB100X cause the almost complete drop of percentage of Venus+ cells in the control samples suggesting that the transposon plasmid is diluted during each replication cycle. In fact, starting from 6 days post infection there was a remarkable drop in the percentage of reporter gene-expressing cells in all of the samples. Since the plasmids I used did not possess an eukaryotic origin of replication or a Matrix Attachment Region, they couldn't be duplicated during the S phase of the cell cycle. Therefore, during the mitotic event the plasmids contained in the father cell got split between the daughter cells.

For this, during the serial passages, we expected reporter gene-detection to drop to a noise level for the samples transfected with transposon-carrying plasmids alone.

Transposition efficiency

As previously mentioned the transfection efficiency measured 48 hours post transfection is slightly different. Thus, the normalized transposition efficiency is represented by the ratio between the percentage of Venus + cells at the endpoint (21-30 days) and the

percentage of transfected cells 2 days after DNA delivery to the cells. The endpoint of each experiment is determined by measuring the percentage of Venus+ cells in the sample transfected with the transposon alone. Since the transposase coding plasmid is kept constant, the variability of the transposition efficiency is a direct measure of the transposon structure and length. In Figure 9 is reported the transposition rate obtained in HeLa and GABEB cells. As expected, the transposition efficiency was inversely proportional to the transposon size. In HeLa cells the transposition efficiency dropped 2.5 fold by increasing from 3.2kb to 14kb the cargo payload independently from the transposon structure (T2 or SA). Interestingly this size-dependent effect was less evident in GABEB cells. In this cell type the decrement was of 1.6 fold and the transposition rate for 14kb transposon remains approximately 30% compared to the almost 21% in HeLa cells. Moreover, as observed in HeLa cells transposons of similar size but with different IR type behaves in the same way even with a relative large size of 14kb constructs.

This result suggests that the transposase could be favoured by some cellular factor more abundant in GABEB cells or, more in general, in primary immortalized human cells.

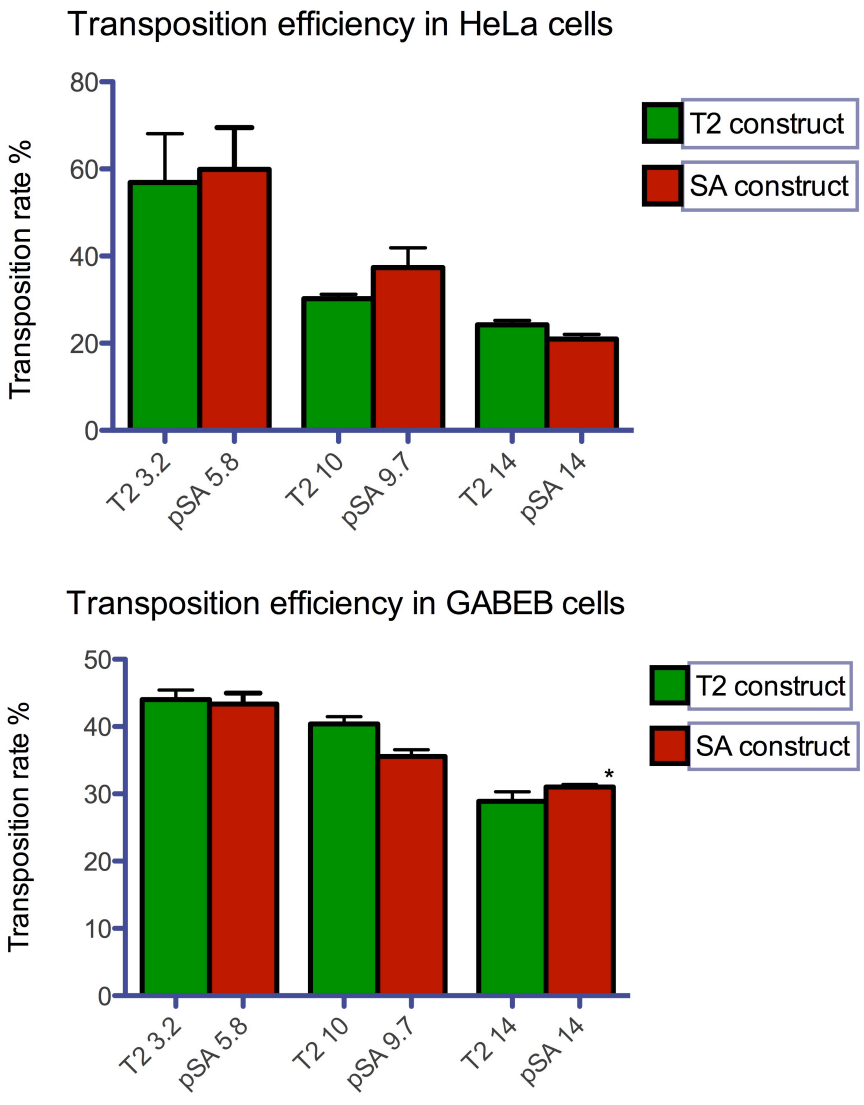


Figure 9. Transposition efficiency. HeLa and GABEB cells were co-transfected with the transposon and transposase carrying plasmids. The transposition rate represented in the Y axis is the ratio between venus + cells at ca. 20 days p.t. and 2 day post transfection. Error bars derive from SEM analysis with n=4; with asterisk n=2.

Chapter II

Clonal molecular analysis

I deeply investigated the molecular characteristics of T2 and Sandwich (SA) construct through cell clonal genomic analyses. HeLa and GABEB bulk populations of Venus$^+$ cells, were limited diluted and seeded into a 96-well plates to obtain a single cell expansion per well. In this way I could analysed the single integration event in each expanded clone. In particular I investigated these parameters:

- Persistence of SB100X and transposon carrying plasmids
- Number of integrated transposons per cell
- Correlation between the expression of reporter gene and the copy number
- Transposon integrity
- Reliability of SA transposition mechanism

Persistance of SB100X and transposon carrying plasmids

I have first investigated if cloned cells harbour the transposon or transposase plasmidic backbone. Therefore, to verify the presence of the transposon backbone, a PCR analysis of the ampicillin sequence was performed and the amplified products were loaded on agarose gel (Figure 10)

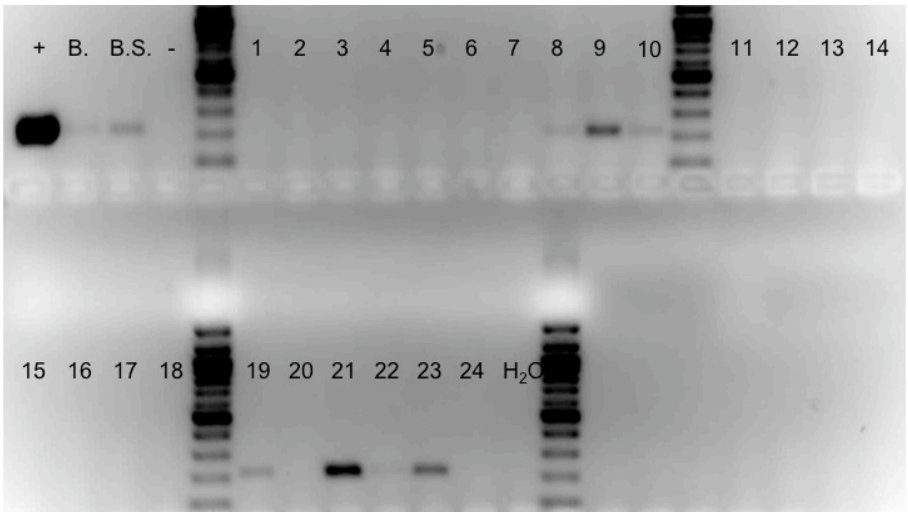


Figure 10. PCR detection of transposon backbone. Genomic DNA of GABEB pT2 10 transposed clones were used in a PCR reaction to specifically amplify an ampicillin sequence. (+) is positive control derived from amplification of 100pg of pT2 10kb plasmid. (–) represent the untreated cell population. **B.** is bulk population of pT2 10 transposed GABEB while **B.S.** is the bulk population sorted for Venus signal.

The Figure 10 shows that six (8, 9, 10, 19, 21, 23) out of 24 clones (25%) were positive for the presence of the ampicillin sequence. In the vast majority of the clones (75%) the Ampicillin was not amplified suggesting the transposon carrying plasmid was lost during cells doublings. Interestingly this proportion (25% positive and 75% negative for plasmid backbone) was found in all the analysed clones deriving from T2 and SA transposons carrying smaller or larger IR cassette. Likewise, primers specific for the transposase sequence were used to verify the persistence of the SB100X plasmid backbone into the isolated clones (Figure 11). This control is crucial for safety reasons. Aberrant integration of the transposase coding sequence could be detrimental due to its

reported [123] toxicity in stable retroviral transduced HeLa cells and to its mobilization activity on integrated transposons.

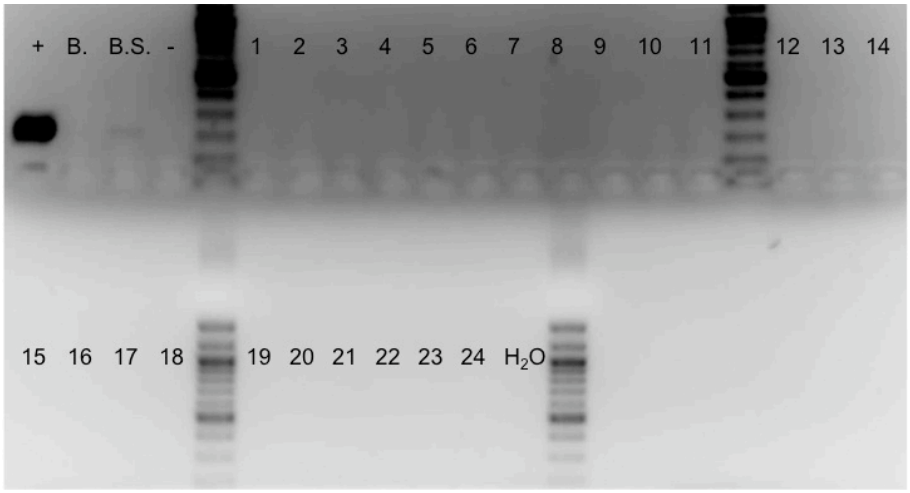


Figure 11. PCR detection of transposase SB100X sequence. Genomic DNA of GABEB pT2 10 transposed clones were used in a PCR reaction to specifically amplify the SB100X sequence. (+) is positive control derived from amplification of 100pg of pSB100X plasmid. (–) represent the untreated cell population. **B.** is bulk population of pT2 10 transposed GABEB while **B.S.** is the bulk population sorted for Venus signal.

As shown in Figure 11 any transposase specific amplicons were detected, indicating that aberrant integration events of the SB100X plasmid did not occurred.

Number of integrated transposon per cell

I than assessed the number of integrated SA transposon per cell. To answer this question I avail the Southern Blot technique. I digested the genomic DNA (gDNA), extracted from HeLa and GABEB clones, with a Restriction Enzyme (RE) that fulfils this three requirement: single cutter enzyme in the transposon

cassette (Figure 12-A); frequent cutter enzyme in the human genome (one RE site every 2-5kb) in order to release and to distinguish fragments of different size of transposon-DNA/gDNA; enzyme not sensible to the CpG methilation to avoid bias digestion. In all blots I used a Venus-specific probe radiolabelled with dCTP-P32.

Southern blot analysis in Figure 12-B shows the copy number in GABEB clones transposed with pSA 9.7kb. Most of the samples, 13 out of 16, carry a single integrated transposon, only 1 clone has three copies and 2 out of 16 clones show two transposed events resulting in an average copy number of 1.3. GABEB clones obtained with pSA 5.7kb harbour 1 to 5 copies with an average of 3 integrated events per clone (Figure 12-C). HeLa clones transposed with pSA 9.7 show an average copy number to 2.3 with a wider standard deviation due to some clones harbouring up to 7 copies (Figure 12-D). Removing this outlier clone, the average reach 1.8, slightly higher than reported in GABEB cells. Unpublished data, obtained in our lab, confirm that the copy number of a T2 10Kb transposon is on average 1.5 in HeLa and in immortalized keratinocytes .

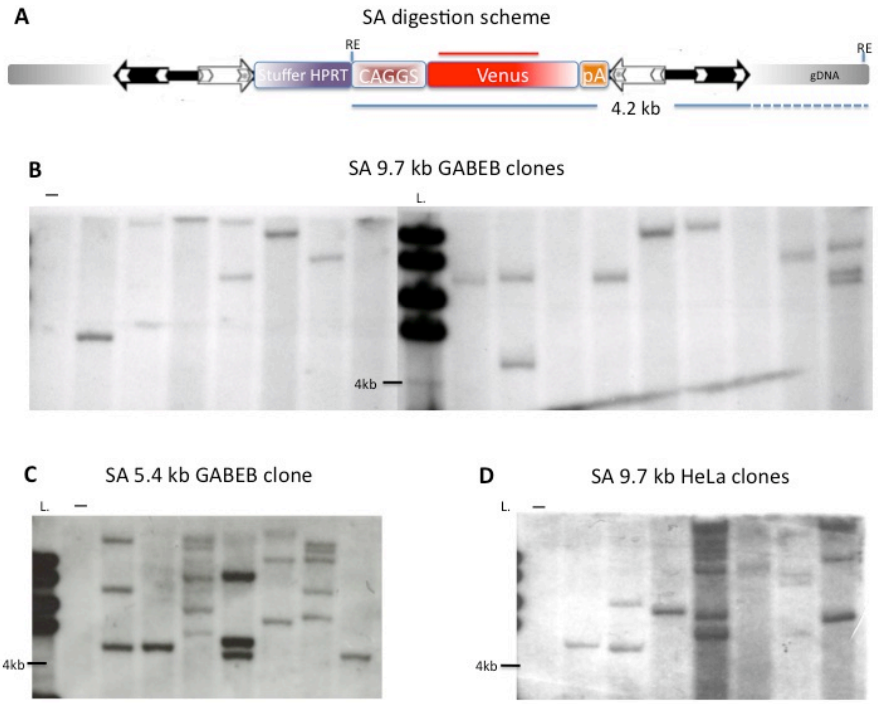


Figure 12. Southern Blot analysis of SA contruct. A) Schematic representation of integrated transposon digested with a Restriction Enzyme (RE) that cuts once in the transposon sequence releasing fragment sizes bigger than 4.2 kb. The red line represents the radioactive Venus-specific probe. B) GABEB clones copy number transposed with pSA 9.7. C) GABEB clones copy number transposed with pSA 5.7. D) HeLa clones transposed with pSA 9.7. 1Kb molecular weight marker is depicted by (L). Negative control of untreated cells is depicted with the sign (-).

Correlation between expression level of the reporter gene and the number of integrated transposons

Once the number of transposed cassette was determined by Southern Blot analysis, I investigated whether the expression of Venus fluorescent protein could be proportional to the copies hosted by every clone. A cytofluorimetric analysis was performed for each clone and the Mean Fluorescence Intensities (M.F.I.),

derived by the mean of the arbitrary unit of fluorescence, were compared to the copies number previously counted by Southern Blot (Figure 13).

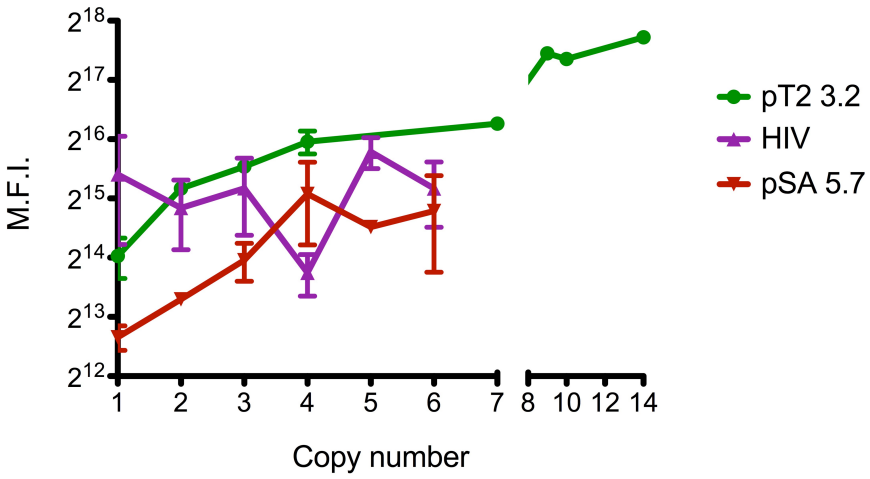


Figure 13. M.F.I.-Copy number correlation. GABEB clones transposed with pT2 are represented in green; Red for clones transposed with pSA 5.7 and purple for cloned transduced with HIV vector carrying a GFP reporter gene.

Figure 13 shows a direct correlation between MFI and copy number in T2 and SA transposed clones. Conversely, expression of the reporter gene in lentiviral (HIV-based) integrants does not correlate with copy number. The direct correlation observed in the case of transposition events could be used to infer the copy number of a given clone by cytofluorimetric analysis.

Transposon sequence stability

Retroviral vectors are widely used for gene therapy application. Tons of papers reported that these vectors could cause

rearrangements during the retrotranscription step resulting in a partially deleted integrated proviruses. This phenomenon is particularly frequent if the vector carries transgene with repetitive sequences. I investigated if the Sleeping Beauty system based on DNA integration without any intermediate step could avoid the risk of rearrangments during the integration process. Genomic DNA extracted from clones were analysed by Southern blot after restriction with a RE which cut twice inside the transposon realising a diagnostic fragment of the expected size consistent with the size of the transposons used in this study. Depending on the RE present throughout the transposon, a 1.8 and 2.4kb fragment was released in the pSA 5.7 and pT2 3.2 construct respectively (Figure 14-A, B). Hybridization with a Venus-specific probe showed the expected bands in all the clones indicating the integrity of transpososed cassettes. In the right panel of Figure 14-B was possible to observe one clone with a band higher than expected (red asterisk). For this integration event is possible to hypothesize that a small rearrangement in the sequence of one of the two RE sites occurred, therefore the fragment is the result of a restriction occurring in the transposon at the right or left end and elsewhere in the genome.

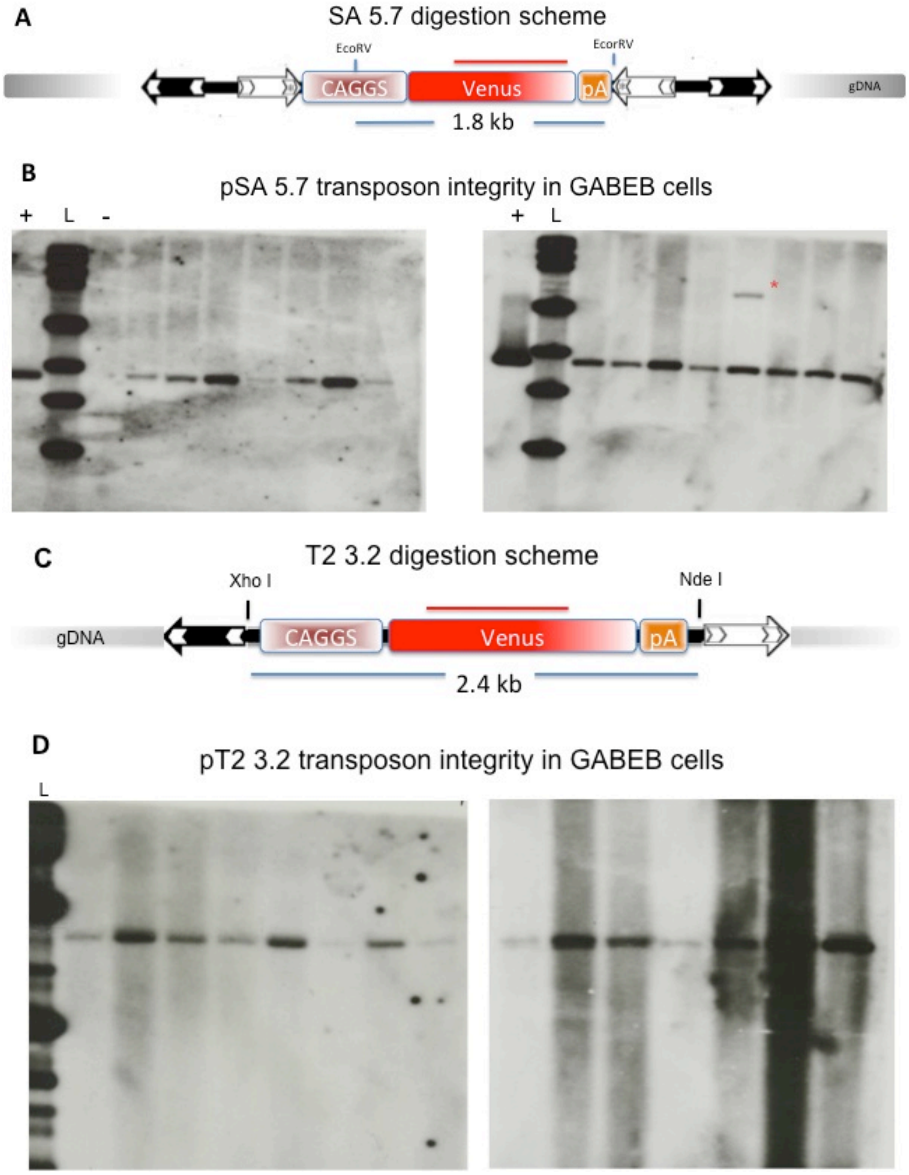
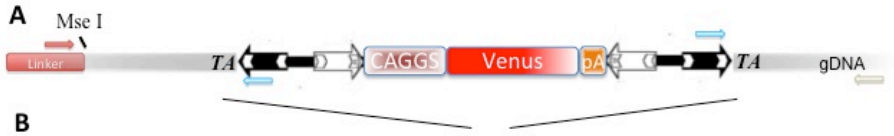


Figure 14 Integrity of transposed cassette. A) EcoRV was used to digest the pSA 5.7 transposon, releasing a fragment of 1.8kb. Red bar shows the radioactive probe used for the hybridization. **B)** Southern blot analysis on pSA 5.7 transposed GABEB clones. The positive control (+) is represented by EcorV fragment release by plasmid pSA 5.7. Mock trnasfected GABEB cells were loaded as negative control (-). **C)** schematic representation of XhoI and NdeI digestion of the pT2 3.2 transposon that release a 2.4kb fragment.. **D)** Southern Blot analysis in pT2 transposed GABEB clones.

Reliability of SA transposition mechanism

Sandwich transposon has an entire IR/DR sequence presented at both ends. Two base pair mutations within the internal DR sequences disable the transposase to use that substrate for the catalysis. That means that for every transposon there are 8 transposase recognised sequence, therefore, different transposition chiasm conformations could arise and modify the integration activity or causes chromosomal aberrations. Integration events mediated by SA transposon were mapped bidirectionally in the human genome due to the repeated IR/DR structure at both ends. Transposed GABEB clones were used to sequence and map on the human genome the single integration event thus confirming the genuine cut-and-paste activity of SB100X. Figure 15-A illustrates the scheme of the modified LM-PCR [124, 125] used to map the left IR-genome junctions. Briefly, 14 days after transfection, Venus-expressing cells were cloned and expanded. Genomic DNA was extracted and digested with *Mse*I. Synthetic adapter (linker) was ligated and the linker-genome-IR fragments were amplified by PCR using a couple of IR and linker-specific primers (blue and red arrows, respectively). The genuine right IR–genome junctions were confirmed by PCR using a forward primer specific to the IR and reverse primers specific for the genomic sequence amplified by LM PCR (blue and grey arrows, respectively). Figure 15-B shows 25 integration

events, belonging to 10 single clones, mapped on the human genome.



B

# Integrants	Clone	Genomic sequence mapped	Chr.
1	Clone 1	GTATGGAAGA**TA**ATACATGT	Ch.X
2	Clone 1	TATTATTGAT**TA**TCTAAACA	Ch.10
3	Clone 5	AAAGACATATA**TA**TATGTCTTTATT	Ch.3
4	Clone 5	TTTCCTGGTAC**TA**TCTTCCTGTTC	Ch.11
5	Clone 12	GAAACCTATA**TA**TAAACATA	Ch.2
6	Clone 12	CATAATCCTG**TA**TGTCAGCA	Ch.21
7	Clone 15	TTAGGTCATA**TA**ATTAGTTT	Ch.14
8	Clone 15	TCCGTAGGTAGG**TA**TGTGTGTGTGT	Unk
9	Clone 15	CACCATTACA**TA**CCTCCCTT	Ch.7
10	Clone 15	TACATTGCTA**TA**AATACCTG	Ch.6
11	Clone 36	ATTCCATCATA**TA**AATAATACGTA	Ch.20
12	Clone 36	GTGGCTAATATA**TA**GAACCTTCTTCA	Ch.3
13	Clone 38	TGACAACAGAATT**TA**TAATAGGGTTT	Ch.18
14	Clone 38	TTATAGTAGTAGG**TA**ACTTTGGCCAA	Ch.20
15	Clone 38	ATTATATAGCCTA**TA**TGAAATGGAA	Ch.22
16	Clone 46	AGCTAAAAATA**TA**GTATATATTAGA	Ch.8
17	Clone 56	ATCAGGGTAAA**TA**GGATATCCATCAT	Ch.5
18	Clone 56	CTCCAGCAAAA**TA**TCTAAATGTTCA	Ch.11
19	Clone 60	CAGCTAAGAATA**TA**GAAGGTACACT	Ch.15
20	Clone 34	TGAGATGATG**TA**CACGTATA	Ch.3
21	Clone 34	ATATTTGACA**TA**TATATGTG	Ch.13
22	Clone 34	GTGACAGAGC**TA**GCTAGACT	Ch.7
23	Clone 34	GTGCTTAAGC**TA**TGAAATAC	Ch.5
24	Clone 34	CCTCAAAAGGT**TA**AACATAGAAT	Ch.1
25	Clone 34	AAGGGGGCTTTT**TA**AACAAGCAG	Ch.2
26	Clone 34	.TAAGTAGCATCACTGGTAGATC	Ch.2
27	Clone 34	.TAATCAAAATTTTACTGAACTTA	Ch.7
28	Clone 34	.TAACATTTTCATCAACTTACAAA	Ch.14

Figure 15. Integration site analysis. A) Schematic representation of techniques used to map and confirm both the transposon junction with gDNA.

LM-PCR to sequence the 5' junction used a couple of primers annealing on synthetic linker and transposon IR sequences. The genome specific PCR mapping the 3' junction was carried by a primer designed on the LM PCR-mapped genomic sequence and a transposon specific primer. The table (panel **B)** summarize the 28 integrations that were examined. TA dinucleotides (depicted in bold) are the target site duplicated after transposition, The three integrations that were not bidirectionally sequenced are reported in red.

The presence of the TA dinucleotide, in bold, correctly duplicated during the integration, represent the target site of the *Sleeping Beauty* transposase. 3' and 5' vector-genome junctions were amplified demonstrating that the integration events occurred without genomic rearrangements, deletion or insertion, in the target sites. Three integrations were mapped by LM-PCR only in the 3' end of the transposon. Unfortunately the genuine integration events was not confirmed in the 5' end suggesting possible rearrangement in the sequence proximal to the target site or chromosomal aberration in that genomic region of GABEB cells.

Chapter III

Integration pattern analysis

Scientist community has endeavoured to study the integration pattern of this new DNA transposon system to evaluate its genotoxicity in comparison to the commonly used integrating retroviral and lentiviral vectors. In the last few years some paper came out describing the integration profile of the T2 transposon [25, 48-51, 126], nobody paid attention to the SA construct.

Therefore, I studied the integration profile of the pSA 5.7 upon SB100X delivery. I address this issue into immortalized GABEB cells for 2 reasons: 1) they are patient derived keratinocyte immortalized with SV40 thus resembling the genomic landascape of primary cells; 2) transfection efficiency in primary keratinocytes is very poor, impairing the possibility to get clones to analyze.

To achieve a considerable amount of integration events I transfected $20x10^6$ GABEB cells with pSA 5.7 transposon and SB100X carrying plasmid. The 20% of Venus positive cells were sorted three days after transfection to enrich the population expressing the reporter gene and expand only the transposed cell. One million and half cells were sorted out having, theoretically, at least an equal number of events. This theoretical number needs to be halved because LM-PCR technique is affected by the availability of *MseI* restriction enzyme in the human genome, that was estimated to allow a genomic sequences coverage of about the

50%. Therefore, the GABEB cell library collects, at least, $7.5x10^5$ events that are more than enough for my purposes.

The 454 pyrosequencing returned 30,000 row reads that were trimmed for the presence of SA IR specific linker with gDNA. Non-perfectly redundant reads were than mapped onto the human genome (UCSC genome browser hg19), requiring the alignment to start within the first three nucleotides and to possess a minimum of 90% identity. Sequences were discarded as mapping to multiple sites when they had more than one match on the human genome differing in identity less than 2%. All UCSC Known Genes having their transcription start site (TSS) at ±50 kb from an integration or random site were annotated as targets. In case of multiple transcript variants, we arbitrarily chose the isoform with the nearest TSS to an integration or random site. 4,020 unique blasted sequences were found and associated to some features related to the integration sites. These features include exons, introns and distance of transposon integration sites within a window of ±2,5kb from the transcription start sites (TSSs) of genes. This dataset (A) was used to analyze the transposon preference for repetitive elements and for hot spots. A more restricted dataset (B) was generated to study distribution by chromosome of the transposon integrants and its preference respect exon, intron, TSS and intergenic region.

The dataset B was generated as follow: 30,000 raw sequences were trimmed and blasted on human genome by nBLAST (NCBI genome brower) according to these criteria: >90% identity and E-

value <0.05. This trimming returned 2,020 unique integration events.

As control we used 10,000 random genomic sites generated in silico taking into account the biases introduced by the LM-PCR and the availability of the TA dinucleotides in the genome. Integration sites were annotated as TSS-proximal when mapping in the ±2.5-kb window around a TSS, intragenic when mapping within a transcription unit >2.5 kb downstream a TSS, and intergenic in all other cases (Figure 16A). More than 50% of the integrations were intergenic unlike retroviral or lentiviral integrants.

In general, the distribution of the SB integrants resembles the composition of the human genome showing 2% of exons , 40% of introns and 3% of TSS proximal regions most likely representing the protein coding genes. Than, I evaluated the distribution of the integrants in each chromosome respect to the 10,000 random sequences. As shown in panel B of Figure 16, the SA integration site follow a flat distribution as the random site with the exception of Y chromosome where a 2.4 fold enrichment was observed. Looking carefully to this unexpected result I argue that the integrants belonging to Chromosome Y are only 4 out of 2,020 instead in the random control 8 out of 10,000 land in chromosome Y (p=0.3). Since the 2-sample test for equality of proportions resulted weakly convincing I would not claim that the SA transposon has a preference for Y chromosome. To better address this issue a larger amount of integrants, probably 1 log more, will be required.

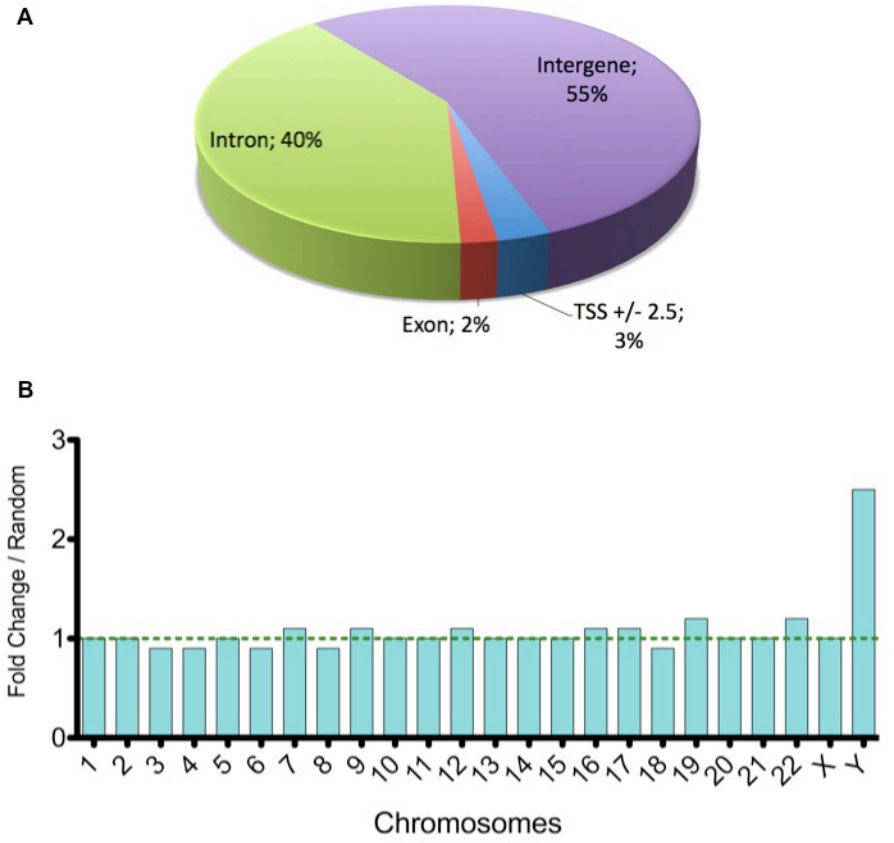


Figure 16. SA integrants distribution. SA "sandwich" transposon integrants (library B) were analysed to highlight their distribution throughout the human genome. **A**) The Pie-graph depicts the distribution of 2,020 integrants in main genomic features: Exon, Intron, gene Trascription Start Sites (TSS) proximity defined at +/- 2,5kb. All the sequences that not fell in any of these features were defined as intergenic. **B**) Integrants chromosome distribution normalized with random sequences. The Y axis indicates the fold change of SA integrants respect to the random control indicated by the green dotted line.

Integration pattern comparison

Several integrating viral vectors were used, to date, for gene addition approaches in gene therapy. However, high-throughput

techniques' revealed integration preferences for active transcribed genes for retro and lentiviral vectors that may result in insertional mutagenesis [45-47]. To propose DNA transposons as gene delivery vectors, the scientific community redirects their effort to assess their genotoxicity. I utilized T2, Tol2 and PB libraries, kindly provided by Zoltan Ivics and Zsuzsanna Izsvak, published by Ammar et al. [51] and SIN-HIV and MLV datasets from Cavazza et al. [127], to compare the SA library generated in epithelial cells. Integration analysis of SA and T2 transposons results in a close to random distribution. T2 and MLV vectors slightly disfavored the exons respect to the random and to the SA libraries. Conversely, Tol2, PB and MLV integrations markedly over-target TSSs defined in 5kb windows (Figure 17). Self Inactivating-HIV, instead, does not prefer TSSs but has a significant preference for gene body.

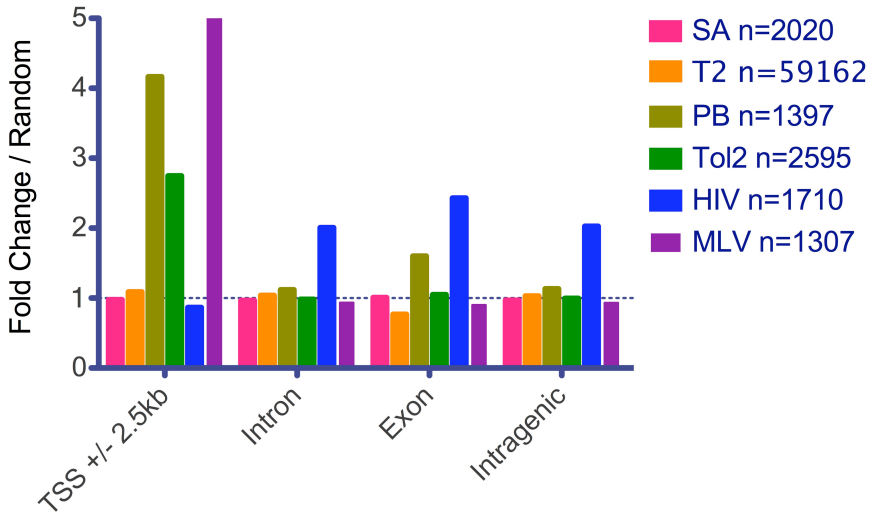


Figure 17. Comparative analysis of integration pattern. The fold change values (Y axis) are calculated as the percentage of integration events divided by the percentage of random events. Evaluated the integration features within Genes(Intragenic), Dotted line represent the random distribution.

SA integration analysis in repetitive element

Human genome is composed by almost 65% of repetitive element [128]. SA integration pattern results in a random distribution in genetic features and hence should follow this trend also in repetitive element. To confirm this distribution I used the library A of 4,020 hits. Due to more stringent criteria, library B could trim out sequences landing in repetitive elements These sequences were assigned to features retrieved by RepeatMaskers database. RepeatMasker is a program that screens DNA sequences for interspersed repeats and low complexity DNA sequences. On average, almost 50% of a human genomic DNA sequence currently will be covered by this program. Unexpectedly SA integrations are not following a random distribution. SA integrations are over-

represented in LINE, SINE, satellite and simple repeats compared to random sequences. The Figure 18 shows a six fold enrichment in Satellite repeated element, three fold in simple repeats and two fold in LINE. Conversely, SINE and LTR were under represented.

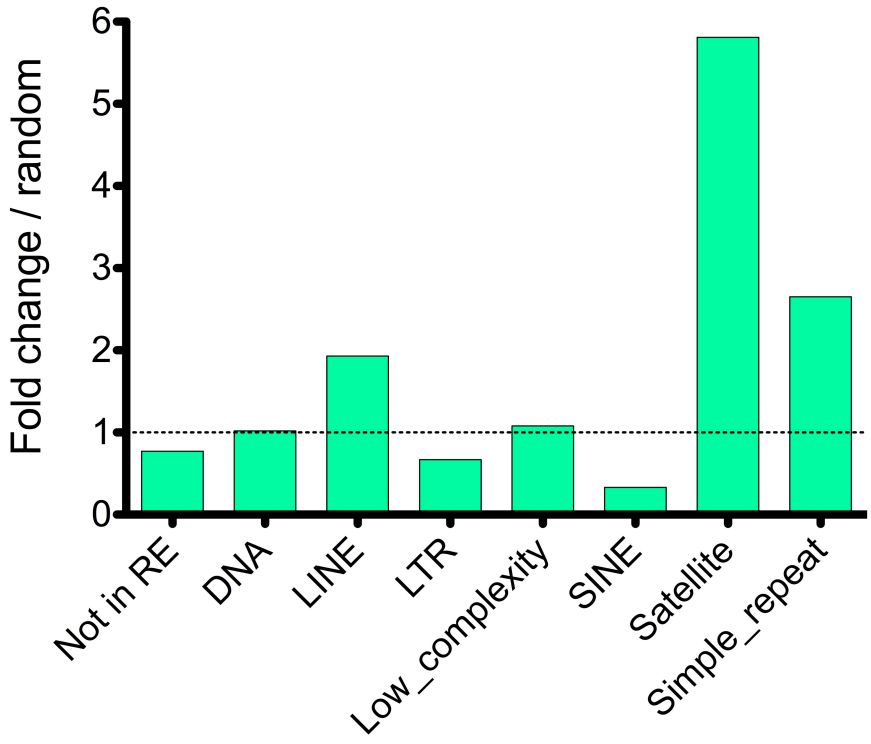


Figure 18. SA integrants distribution in repetitive element. SA "sandwich" transposon integrants (library A) were analysed to highlight their distribution in the human genome repetitive elements. On X axes are reported the main repetitive elements: DNA, LINE, LTR, Low_complexity, SINE, Satellite and Simple_repeat. All the sequences that not fell in any of these features were defined as not in Repetitive Element (Not in RE). Integrants distribution is normalized with random sequences. The Y axis indicates the fold change of SA integrants respect to the random control indicated by the green dotted line.

Analysis of the hot spots

I next evaluated the presence of recurrent insertion sites (hot spots) for SA and T2 transposon. Even though, macroscopically, the integration pattern is close to random, a deeper analysis showed some integrations site clustered in few kilobases suggesting specific mechanism that lead the transposase. On the same line in 2005 was reported that HIV-1 integrase is tethered by housekeeping protein LEDGF/p75 (lens derived epithelium-derived grow factor) to gene/genomic region bound directly or indirectly by p75 [129]. Also MLV integrase has several partners among the Transcription Factor Binding Sites involved in its TSS tethering and tropism for actively transcribed genes[130, 131].

For this analysis I considered the not deeply filtered SA library A of 4,020 unique blasted integration in order to not collapse hits coming from independent events mapped in a windows of few bases. Integration clusters were defined by at least two consecutive genuine hits in a statistically defined 3.5kb window.

Surprisingly, with both SA and T2 libraries was possible to score some hot spots that account for 1,5% and 2,1% of the total hits sequenced, respectively. On the contrary, the hot spots of MLV integration account for more than 21% of the integration events, while for LV vector the hot spots represent ca. 8%, of the total hits [45, 130]. Although SB system shows lower amount of integration involved in clusters, it is possible to distinguish some heavily targeted genomic regions hosting 16 or 12 integrations.

Furthermore, some of these integrations forming clusters are constrained in a space of 1,000 base pairs that was never observed for any transposons.

Hits/cluster	SA clusters	SA cluster length (bp)	T2 clusters	T2 cluster length (bp)
16	1	8917		
14	1	1019		
12			1	857
7	1	1396		
6			2	5738*
5	1	9474	8	6264*
4	2	2798*	35	3579*
3	4	1134*	351	3108*
Sum	10		397	
Hits involved in clusters	62		1257	
Tot hits	4,020		59,170	
Clustered Hits/tot hits	1,5%		2,1%	

Table 1. Clusters summary of T2 and SA transposon. Hits/cluster column indicate the amount of integrants that generate a cluster. SA and T2 cluster indentify how many cluster are in each of the "Hit/cluster" feature reported in the first column. Below are reported the sum of clusters, the amount of integrants (hits) involved in clusters, the total amount of hits per library and percent of the total hits that are involved in cluster. *Average length

Investigating the hot spots of the SA and T2 transposon I noticed that more than 20% of them fell in a specific repeated element, the peri-centromeric ALR/alpha satellite. The other clusters were enriched in other repetitive genomic elements such as sub-

telomeric regions BSR/Beta Satellite, L1 LINE and ribosomal small subunit RNA SSU-rRNA_Hsa. Moreover, two clusters placed on chromosome 1 and 6 centromeric region and one in telomeric region of chromosome 18 were scored in both T2 and SA libraries. I deeply analysed the cluster on chr. 1 centromeric region to have a whole picture of the cluster (Figure 19). Highlighted in green the TA targeted by the SA transposon and with the arrows the direction of the sequenced junction starting from the IR. The direction of the arrows could help us to discriminate the origin of some events, close each other, if effectively comes from independent integration events and are not produced by a single mis-transposition that could cuts out some bases. A clear example comes from #2, 3, 4 integrations. They are 3 independents hits spanned in only 12 bases of distance and orientated in the same direction confirming an independent transposition event for all of them. These data indicate that specific mechanisms could tether the Sleeping Beauty transposase into hot spots how similarly happen with viral integrases. This genetic characteristic needs to be deeply analysed.

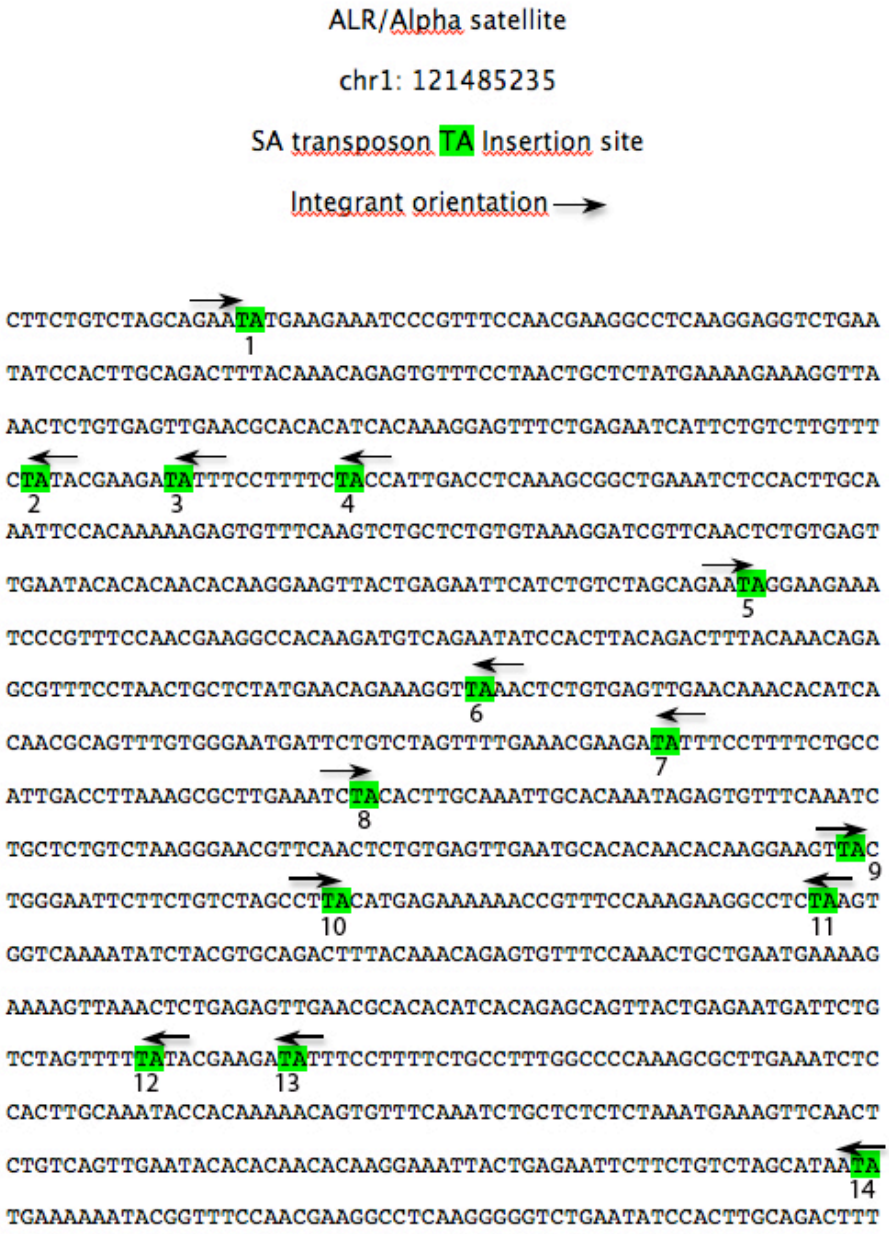


Figure 19. Scheme of SA cluster on Chr1: 121485235 ALR/alpha Satellite.

Discussion

Transposable elements were first discovered by Barbara McClintock in 1942 through the maize kernels observation and characterization. Since that year, it was possible to amplify the transposon knowledge with the description of different kinds of transposons, different mechanism of transposition and, with genomes sequencing project, also their abundance in the human genome. Several studies report that only few retrotransposon are still active in human genome while DNA transposon lost their capacity to be mobilized via cut-and-paste mechanism. The Sleeping Beauty (SB) transposon found in fish genomes belongs to the Tc1/mariner transposon superfamily and, since its molecular reconstruction from inactive fossil sequences in 1997 [27], it has been extensively studied and employed to stably incorporate genes of interest into genomes for several applications. The "resurrection" of an active vertebrate DNA transposon based system, aroused the interest of the scientific community for its enormous potential as a tool in mammalian cellular biology. This interest was converted in a broad engineering effort in order to ameliorate the transposition activity modifying both components of the SB system, the transposon IR and the transposase enzyme. Mates et al. in 2009 [27, 36] synthesizes a hyperactive SB transposase, designed SB100X, that was shown to integrate efficiently in vertebrate cells and also in human hematopoietic stem cell. Rewarded as "molecular of the year", SB100X started to be considered as an effective alternative to retroviral vectors.

Contemporarily, SB transposon IR were mutated to improve their capacity to be mobilized and, to date, there is not a direct comparison that define genetic characteristics of the T2 and SA IRs. SA construct was defined as "sandwich" transposon due to its structure in which two SB elements, modified so that mobilization of single components results impaired, are placed in a head-to-head orientation flanking a genetic cargo. This transposon was poorly characterized till its publication in 2004 [31]. In this work I analyzed different molecular features of SA transposon comparing them to the T2 one in order to assess safety, efficiency and genomic stability for their potential applications in gene therapy. For this purposes I embraced the disease model called Dystrophic Epidermolysis Bullosa (DEB), a rare skin disease, that is caused by mutation in COLVII coding sequence which impair the epidermis adhesion onto the derma layer. The COLVII gene has a peculiar sequence with several repeated domain that contribute to structure the 9kb cDNA. These characteristics make the COLVII gene unstable in retroviral and lentiviral vector resulting in a recombined integrated sequence once transduced in target cells. Sleeping Beauty system could overcome this problem because has no retro-transcription step and integrate efficiently with a long term expression. Efficient, non toxic and safe delivery of the Collagen VII cDNA into clonogenic keratinocytes isolated from patient's biopsies could potentially restore the collagen VII production in a autologous transplantable skin implant. I started from the molecular characterization of T2 and SA transposon in combination with hyperactive SB100X. For this purpose, I used

epithelial cells that are targets for gene addiction in DEB. HeLa, GABEB cells (immortalized keratinocytes coming from a patient affected by Generalized Atrophic Benign Epidermolysis Bullosa) and primary keratinocytes were transfected with a panel of T2 or SA transposon carrying size-increasing Venus expression cassette in combination with SB100X plasmid. I tracked the expression of the Venus reporter gene in treated cells to estimate the transposition efficiency normalizing the data with the transfection rate. Basically, the transposition efficiency is derived by the ratio between the percentage of Venus + cells at the endpoint (21-30 days) and the percentage of transfected cells 2 days after DNA delivery to the cells. Transfection rate result higher in HeLa, less pronounced in GABEB cells, very low in primary keratinocytes and was shown to be negatively affected by the increasing size of transposon plasmid. Primary keratinocytes were poorly transfected with >10kb transposon and I cannot retrieve fully satisfactory data except the transposition efficiency with T2 3.2kb and SA 5.7kb that were comparable to that obtained in GABEB cells. As expected, the transposition efficiency was inversely proportional to the transposon size. Interestingly this size-dependent effect was less pronounced in GABEB cells observing a decrement of 1.6 fold instead of 2.5 fold obtained in HeLa cells. Finally, with 14kb cassette, GABEB and Hela cells achieve a transposition efficiency of 30% and 21% respectively. This could be influenced by intervention of some cellular factors more abundant in GABEB cells acting to favour the transposase activity.

In general, T2 and SA transposons of similar size behave in the same way even with a relative large size of 14kb cargos in both cell type. Transposed population of these cells were subjected to limiting dilution to obtain a single cell derived expansion. The clones derived were employed to characterize several molecular parameters. First of all, for safety reasons I verified the absence of SB100X sequence in clones through a PCR amplification. Retroviral-mediated permanent expression of the transposase was demonstrate, by Galla et al. [123], to induce citotoxicity in stable transduced HeLa cells. Moreover the stable persistence of transposase could favour a re-mobilization of integrated transposon. On the contrary, PCR detection of transposon backbone sequence shows a different scenario. 25% of T2 10kb transposed GABEB clones result positive for ampicillin sequence present in transposons backbone. Similar unpredicted plasmid integrations were found in other cloned population treated with pSA 5.7kb, pT2 3.2kb and pSA 10kb. It is possible to hypothesize that the plasmid backbone carrying the transposon could have some advantage to remain episomal or integrate in the genome. The transposon excision step from the plasmid leaves the backbone with a double strand break that could be recruited from the endogenous repair proteins and introduced into the cell genome. Then, I planned to verify the average copy number of SA transposon. This parameter was previously reported for T2 construct in different papers [71, 126, 132] and it varies from the amount of plasmid provided at the transfection step. What I analysed is the SA copies harboured in each clone via Southern

blot. In SA 9.7kb transposed GABEB clones is possible to note 13 samples out of 16 hosting a single integrated copy, two of them harbour two copies and only one possess three copies resulting in an average copy number of 1.3. With SA 5.4kb transposed GABEB the average rose to 3 events per clone, while HeLa cells treated with SA 9.7kb show an average copy number of 2.3 with a wider standard deviation due to some clones with 7 copies. Removed from this outlier, the average copy number goes to 1.8. Unpublished data available in the laboratory confirm the average of 1.5 copies in T2 10kb treated HeLa and in immortalized keratinocytes while show an higher average of 4.5 in T2 3.2kb transposed GABEB cells. This scenario suggests, at first instance, that the hyperactive transposase's activity is subjected to the transposon size more than the cell or the IR type. The transfection procedure, even if pondered to introduce similar stoichiometric amount of plasmids, cannot efficiently deliver constructs exceeding 12kb. However, it is possible to fine tune this parameter playing with ratios of plasmids used for transfection, or as demonstrated by Yant et al. [61], bypass the transfection procedure through the generation of Adenoviral vector carrying transposase and transposon.

Once assessed the copies of transposon integrated in each clone. I could associate them to the expression level of the Venus fluorescence gene in the corresponding clone. I observed a direct proportion that links the copy number to the expression level. This feature could be used also to infer the copy number of a given clones by cytofluorimetric analysis. Conversely, expression of the

reporter gene in lentiviral (HIV-based) integrants does not correlate with copy number. HIV integrated cassettes are probably more subjected to the activity of surrounding areas. In fact, HIV integrase overtarget intragenic sequences and the exogenous expression cassette could be perturbed by the activity of the specific gene hit by, for example, aberrant splicing events [133]. Moreover, Retroviral and Lentiviral vectors could cause rearrangements during retrotranscription step resulting in a partially deleted integrated proviruses, a frequent occurrence in transgene hosting repetitive sequences. Therefore, I investigated if SB system could avoid this recombination during the transposition process. GABEB clones' genomic DNA were digested with enzymes that cut twice internally the transposon sequence and excise a diagnostic fragment. Through the Southern blot techniques was possible to detect the integrity of all the transposed sequences in both T2 and SA constructs. Only one clone, shows a band higher than expected that more likely is due to micro-rearrangments in one of the two restriction enzyme sites thus releasing a fragment that represent a transposon-genome junction. Afterwards, I investigated the Sandwich reliability in transposition process. This query derive from the double transposon element flanking the transposable cassette. This structure could form with the transposase units different transposition chiasm that could cause aberrant integration or chromosomal alterations. I bidirectionally mapped the SA integration sites in GABEB clones to assess the genuine cut and paste mechanism. Twenty five integrations, out of 28, were

validate for a canonical transposition events signed by the TA dinucleotide duplication at both ends. Three of them, coming from the same clone, were not confirmed at both 5' and 3' ends. PCR amplification of 2 of these 3 insertion sites resulted in several unspecific amplicons and for the last hit was completely negative. It's possible to provide three hypothesis for these events: 1) the sequences in which the integration occur are awkward to amplify for structural conformations or for redundant sequences that gives unexpected amplicons; 2) GABEB genomic regions bear micro or macro alteration close to the SB integration site and therefore, there were not available PCR templates; 3) integration causes a chromosomal aberration integrating the two IR/DR ends in different loci.

LM-PCR was also employed together with pyrosequencing to derive a high-definition map of SA/SB100X integration sites in the genome of GABEB cells. This analysis is critical for the integrating vector because allows to evaluate genotoxicity, to understand molecular mechanism recruited by the system and to compare the integration landscape to the commonly used and well studied retroviral and lentiviral vectors. Recently, some published works trace the integration profile of Tol2, piggyBac systems and T2 transposon upon SB100X delivery [25, 48-51, 126]. These datasets were collected together for a direct comparison with SA construct. First analysis of 2,020 integrations showed a random distributions of transposed events in genetic features: 2% were mapped in exons, 3% in the surrounding areas of a Transcription Start Site (TSS ±2.5kb), 40% in introns and 55% were intergenic.

As control we used 10,000 random genomic sites generated in silico "weighted" for TA minimal SB requirement and for MseI sites used for LM-PCR reaction. Therefore, values coming from SA library were normalized upon the random control values. In this way I could evaluate the SA integrants distribution in each chromosome that follows the random dataset values with exception of a 2.5 fold enrichment in Y chromosome. Looking carefully to this unexpected result I argue that the integrants belonging to Chromosome Y are only 4 out of 2,020 while 8 out of 10,000 land in chromosome Y in the random control ($p=0.3$). Since the 2-sample test for equality of proportions resulted weakly convincing I would not claim that the SA transposon has a preference for Y chromosome. To better address this issue a larger amount of integrants, probably 1 log more, will be required. SA distribution in genetic features was compared with the T2 IR integration sites. These latter follow the random distribution likely shown in SA analysis with a slightly under-representation in exons. Other than Sleeping Beauty, Tol2 and piggyBac DNA transposons were evaluated for the integration profile comparison. Tol2 and piggyBac over-target TSS-proximal sequences which indicates a cellular co-factor tethering. In particular, piggyBac has a preference for exon and, together with the TSS, denote a broad preference for open chromatin[51, 83]. Lentiviral and retroviral integration patterns were extensively studied and were shown to influence the targeted gene [127, 130, 131, 133]. Indeed, Self-Inactivating HIV derived integrations confirm a pronounced preferences in active gene body and avoid

TSS surrounding sequences [129]. Conversely, MLV favours integration near TSS bearing the risk of transactivating endogenous gene[46, 125]. Chromosomal integration of SB transposons are directed into TA dinucleotide sites, and influenced by additional factors including DNA structure and TA content. Despite the requirement for TA sites and its duplication during the integration, SB's insertion profile is considered fairly random at the genomic level, and thus might be safer as compared with integrative viral vector or other transposon system.

Another integration analysis aimed to confirm a SA random distribution in repetitive element. Unexpectedly SA integrations are over-represented in LINE, satellite and simple repeats compared to the random control while SINE and LTR were under targeted. Is possible to hypothesize a most favourable DNA structure in over-targeted element or other mechanism that actively work on the transposase/transposon chiasm of transposition. I would underline that in these data is possible to have a negative bias related to repetitive elements because of the blast that discharge not-unique mapping hits. Repetitive element are present several times throughout the human genome and therefore are most likely excluded during the alignment. Therefore, I deeply investigated the SA and T2 integration preferences to evaluate the presence of recurrent insertion sites. Even though the distribution of integrated transposon was shown to be macroscopically random, I could observe some events clustered in a restricted compartment suggesting a possible tethering mechanism for the SB transposase. As already reported

for pathogenic/oncogenic virus such as HIV and MLV, the integration site selection is not random but it is drove by biochemical interaction between integrase and cellular cofactors. In particular, LEDGF/p75 protein affects the choice of target sites for HIV-1 integrase in transcription units. However, it has been shown that additional factors may be involved in targeting HIV DNA integration [134]. Similarly, MLV integrase has several partners among the transcription factor binding sites clarifying its TSS predilection. 8% and 21% of HIV and MLV integrations, respectively are clustered in genomic region showing defined genetic characteristics. MLV integrations cluster in regulatory region and transcribed gene, while HIV prefer actively expressed gene and frequently occurred in chromosomes harbouring gene-dense regions, in particular over targeted gene are reported in $CD34^+$ cells employed in pre and post-transplantation in adrenoleukuodystrophy (ALD) clinical trial [135, 136].

In the SA and T2 libraries was possible to score some hot spots that account for 1,5% and 2,1% of the total hits, respectively. Despite the relative lower percents respect to RV and LV, these hot spots were found heavily targeted hosting even 14 or 12 integrations in a window of 1,000 base pairs. Among the hot spots of both T2 and SA transposon was possible to observe more than 20% of them falling in a peri-centromeric repetitive element ALR/alpha satellite. This data corroborate the 6-fold enrichment noted in repetitive element distribution analysis. Surprisingly, three clusters, two placed on chromosome 1 and 6 centromeric region and one in telomeric region of chromosome 18, were

scored in both T2 and SA libraries. These data indicate that specific mechanisms could tether the Sleeping Beauty transposase to the hot spots how similarly happen with viral integrases or also could be still possible to hypothesize a particular DNA conformation in TA sites [137] that are perfectly compliant with the transposon/transposase complex of integration. These characteristics rise an interesting matter that need to be deeply investigated.

The SB transposon system has certainly enriched the field of genomic manipulations aimed at several purposes including those imposed by gene therapy. One of the first preclinical studies aimed to restore the fumarylacetoacetate hydrolase (FAH) enzyme into mouse model for the tyrosinemia type 1 disease[68]. In 2000, Yant group introduce the factor IX gene sequence through SB11 in haemophilic mice and demonstrate a long term expression of the therapeutic gene [44]. Therapeutic potential of the SB system was also reported in the treatment of mucopolysaccharidosis (MPS) types 1 and VII [71], Huntington disease [69], sickle cell disease [138] and pulmonary hypertension [139]. The SB transposon system was also employed for cancer gene therapy. This goal was pursued introducing a chimeric antigen receptor (CAR) into T cell directed against the CD19 antigen expressed on malignant B cells. This studies lead to the first clinical trial in patients affected B malignancy utilizing the SB system [66]. These studies are remarkable and will provide significant alternatives to the current therapies and to the application of the LV and RV vectors in gene addition protocols.

The genotoxicity of the RNA vectors came out in the short term follow up of the firsts clinical trial on X-linked severe combined immunodeficiency (X-SCID) [140]. Following studies were implemented in other on-going clinical trial to assess the genotoxicity [136, 141, 142] as chronic granulomatous disease (CGD) [143], adenosine deaminase immunodeficiency (ADA-SCID) [144] and Wiskott-Aldrich syndrome (WAS) [141, 145]. This thesis on Sleeping Beauty transposons traces some of the experimental studies done to understand RV genotoxicity [45-47] and assess some safety issues required for a long and stable gene transfer.

Material and methods

Plasmid constructs

Plasmids carrying the Sandwich IR contructs, T2 IR CAGGS Venus and SB100X were kindly provided by Zoltan Ivics. CAGGS Venus expression cassette was Dra III excised from pT2 3.2 and introduced into EcoRV digested pSA to obtain pSA 5.7.

pT2 3.2 and pD28 were digested with XbaI to clone pT2 CAGGS Venus plus a C346 stuffer DNA of 2,7 kb.

Two pieces of HPRT stuffer were PCR amplified and cloned into the pCR 2,1 (TOPO® cloning kit) plasmid. pT2 10 was cloned ligating the pT2 CAGGS Venus SpeI with Stuffer HPRT NHE fragments. Finally, the pT2 14 plasmid derive from pT2 10 ClaI and the second piece of Stuffer HPRT Not digested and ligated with the former. Digestions with NheI RE provide the fragments pSA 5.7 NheI and Stuffer NheI to clone the pSA 9.7. Then the pSA 9.7 was digested with PmeI enzyme and ligated with Stuffer HPRT PvuII to obtain the pSA 14.

Cell culture

Different cell lines or primary cells were used in this work. HeLa cells were cultured using DMEM medium (Lonza) added with 10% Fetal Bovine Serum (FBS), 1% L-Glutamine (L-Gln) and 1% Penicillin-Streptomycin (Pen/Strep). For each experiment, an aliquot of cryo-preserved HeLa cells was thawed and plated on 8cm dishes. Upon reaching 80-90% of confluency, they were re-plated on 6-wells

culture plates at a concentration of $3x10^5$ cells/well. After 24h, cultures in each well were at 70-80% confluency, ready to be transfected.

Mouse NIH3T3 fibroblast cell line were maintained in Dulbecco's Modified Eagle's medium (Euroclone), supplemented with 10% bovine serum.

Primary keratinocytes and keratinocytes from homozygous generalized atrophic benign epidermolysis bullosa (GABEB) patients immortalized with SV40 (kindly provided by J.W. Bauer) were cultivated in EpiLife medium supplemented with human keratinocyte growth supplement (HKGS) (Invitrogen, US). Epilife is a serum-free keratinocyte culture medium with a low calcium (0.06 mM) concentration supplemented with HKGS which results in a final concentration of 0.2% (v/v) BPE, 5 lg/mL bovine insulin, 0.18 *lg/mL* hydrocortisone, 5 lg/mL bovine transferrin and 0.2 ng/mL human EGF. Upon reaching 80-90% of confluency, they were re-plated on 6-wells culture plates at a concentration of $2,3x10^5$ cells/well. After 24h, cultures in each well were at 70-80% confluency, ready to be transfected.

Transfection based transposition efficiency

HeLa and GABEB cells were both transfected with FugeneHD transfection reagent (Roche). First, for each sample 2 μg of DNA were added to 100μl of either DMEM (for HeLa) or EpiLife (for GABEB). The mediums used for this transfection reaction mix were not added with FBS, L-Gln or Pen/Strep.

The transposon/transposase amounts of plasmid DNA were calculated to respect the stoichiometric ratio of 1:1 or 2:1 for transposon >10kb, in a total quantity of 2μg. 2μg of only transposon plasmid were used for non-transposed control.

Each transfection reaction mix was additioned with 6μl of FugeneHD and subsequently mixed by pulse-vortexing for a few seconds. The mixes were therefore left at room temperature for 10' in order to allow the formation of lipoplexes. After the 10' had expired, each mix was added drop-by-drop to a cell culture sample, which was subsequently incubated at 37°C.

The percentage of Venus + cells were determined 2 and 20-30 days post-transfection via flow cytometry and the values were used for calculation of the transposition efficiency: Venus+ cells at last count (20-30 days post transfection) divided by Venus+ cells at Day 2 and multiplied by 100. Cells that were only transfected with the transposon plasmid served as control for background integration events.

Treated clone were analysed via flow cytometry to determine presence of doublets and the venus+ mean fluorescence.

Limiting Dilutions

GABEB cells were limiting diluted to obtain a concentration of 0.5 cell/well, plated onto lethally irradiated NIH3T3 cells and cultured in keratinocyte growth medium, a DMEM and Ham's F12 media mixture (2:1) containing FCS (10%), penicillin-streptomycin (1%), glutamine (2%), insulin(5 μg/ml), adenine

(0.18 mM), hydrocortisone (0.4 μg/ml), cholera toxin (0.1 nM), triiodothyronine (2 nM). After 1 week, the medium was replaced by EpiLife medium supplemented with HKGS. After 1 week, the medium was replaced by EpiLife medium supplemented with HKGS. After 2 weeks GABEB cells were were trypsinised at subconfluence and re-plated without the NIH 3T3 feeder-layer in EpiLife HKGS medium.

HeLa cells were limiting diluted to obtain a concentration of 0.3 cells/well and plated in 96 well plate in DMEM medium complemented with 10% FBS.

Southern blot analysis

Genomic DNA was extracted from 1-5x10^6 GABEB cells by a QIAmp DNA Mini kit (Quiagen), digested overnight with: NheI (copy number pSA 9.7); EcoRV (integrity cassette pSA 5.7); XhoI with NdeI (integrity cassette pT2 3.2). Digested gDNA were ran in 10ug aliquots on a 0,8% agarose gel, transferred to a nylon membrane (Duralon, Stratagene) by Southern capillary transfer and probed with 2x10^7 cpm ^{32}P-labeled Venus probe according to standard techniques [146].

PCR episomal amplification

About 100ng of template gDNA were used in a PCR reaction. Primers capable to amplify the Amp resistance gene (Amp For 5'-GGTTAGCTCCTTCGGTCCTG-3'; Amp Rev 5'-

GCTATGTGGCGCGGTATTAT-3') or the SB100X transposase (SB For 5'-GCCACTCAGCAAGGAAGAAG-3'; SB Rev 3'-GTGTTGGAAGACCCATTTGC-3') were used to detect genomic integrations of pSA 9,7 backbone and pSB100X respectively. The PCR reaction was started with 5' of incubation at 95°C and followed by 30 cycles comprising 30" at 94°C, 30" at 58°C and 30" at 72°C.

LM-PCR and bioinformatic analysis

Integration sites were amplified by linker-mediated PCR (LM-PCR), as described[124]. Briefly, genomic DNA was extracted from 0.5-5x10^6 transposed cells and digested with *Mse*I and a XhoI enzyme to prevent amplification from internal mutated IR fragments. An *Mse*I double-stranded linker was then ligated and LM-PCR performed with nested primers specific for the linker and SA IR/DR (TAGpSAIR:5'-CGTATCGCCTCCCTCGCGCCATCAGCGTAGTGTATGTAAACTTCCGACTTC-3' and LinkerTAG: 5'-CTATGCGCCTTGCCAGCCCGCTCAGAGGGCTCCGCTTAAGGGAC-3'). PCR products were shotgun-cloned (TOPO TA cloning kit, Invitrogen; Carlsbad, CA) into libraries of integration junctions for GABEB clones, which were then sequenced. A valid integration contained the TAGpSAIR nested primer, the entire SA IR/DR sequence up to a TA dinucleotide and the linker nested primer. Sequences between the TA and the linker primers were mapped onto the human genome by the BLAT genome browser (UCSC

Human Genome hg19). Sequences featuring a unique best hit with ≥95% identity to the human genome were considered genuine integration sites.

LM-PCR derived amplicons were ran on a Roche/454 GS FLX using titanium chemistries by GATC Biotech AG Next Gen Lab (Germany). 30,000 raw sequences were trimmed with Python scripts for the presence of SA IR sequence and TA dinucleotide. The retrieved sequences were mapped as follows:

-First map all reads to the human genome (hg19) using NCBI BLAST (nblast).

- Only keep hits with at least 90% sequence identity and an E-value of at most 0.05

- Only keep reads heaving a single hit following the above criteria

- Only keep reads which could be mapped from their beginning, because otherwise the exact mapping position could not be determined

- Collapse all reads to the same position, also if they have different strands (to make reads non-redundant).

10,000 control sites were in-silico created taking in count bias introduced by LM-PCR techniques. The control was "weighted" for MseI site distance and TA dinucleotide presence and it follows the same mapping criteria used for the SA sites.

Insertion sites and experimental control sequences were annotated according following criteria: insertion sites were classified as "TSS-proximal" when occurring at a distance of ±2.5 kb from the TSS of any Known Gene, "intragenic" when occurring within the transcribed portion, in "exon" when fell in noted gene

exon, in “intron” when fell in known gene intron and “intergenic” in all other cases.

The analysis of repetitive element was performed with the Galaxy genomic hyperbrowser (https://main.g2.bx.psu.edu/) crossing the genome browser table Repeat Masker with genomic integration coordinates.

Clusters analysis were statistically defined as described in Cattoglio et al. [130].

References

1. McClintock, B., *The origin and behavior of mutable loci in maize.* Proceedings of the National Academy of Sciences of the United States of America, 1950. **36**(6): p. 344-355.
2. McClintock, B., *Induction of Instability at Selected Loci in Maize.* Genetics, 1953. **38**(6): p. 579-599.
3. Hartl, D., et al., *What restricts the activity of mariner-like transposable elements.* Trends in genetics : TIG, 1997. **13**(5): p. 197-201.
4. Lander, E., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.
5. Morrish, T., et al., *DNA repair mediated by endonuclease-independent LINE-1 retrotransposition.* Nature genetics, 2002. **31**(2): p. 159-165.
6. Pritham, E.J., T. Putliwala, and C.d. Feschotte, *Mavericks, a novel class of giant transposable elements widespread in eukaryotes and related to DNA viruses.* Gene, 2007. **390**(1,Äì2): p. 3-17.
7. Voytas, D.F. and J.D. Boeke, *Ty1 and Ty5 of Saccharomyces cerevisiae.* Mobile DNA II, 2002. **2**.
8. Havecker, E., X. Gao, and D. Voytas, *The diversity of LTR retrotransposons.* Genome biology, 2004. **5**(6): p. 225.
9. Han, J., *Non-long terminal repeat (non-LTR) retrotransposons: mechanisms, recent developments, and unanswered questions.* Mobile DNA. **1**(1): p. 15.
10. Kurzynska-Kokorniak, A., et al., *DNA-directed DNA polymerase and strand displacement activity of the reverse transcriptase encoded by the R2 retrotransposon.* Journal of molecular biology, 2007. **374**(2): p. 322-333.
11. Kapitonov, V. and J. Jurka, *Rolling-circle transposons in eukaryotes.* Proceedings of the National Academy of Sciences of the United States of America, 2001. **98**(15): p. 8714-8719.
12. Feschotte, C. and E. Pritham, *DNA transposons and the evolution of eukaryotic genomes.* Annual review of genetics, 2007.

13. Cordaux, R. and M. Batzer, *The impact of retrotransposons on human genome evolution.* Nature reviews. Genetics, 2009. **10**(10): p. 691-703.
14. Gellert, M., *V(D)J recombination: RAG proteins, repair factors, and regulation.* Annual review of biochemistry, 2002. **71, 101–32**.
15. Agrawal, A., Q. Eastman, and D. Schatz, *Transposition mediated by RAG1 and RAG2 and its implications for the evolution of the immune system.* Nature, 1998. **394**(6695): p. 744-751.
16. Kapitonov, V. and J. Jurka, *RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons.* PLoS biology, 2005. **3**(6).
17. Nakamura, T. and T. Cech, *Reversing time: origin of telomerase.* Cell, 1998. **92**(5): p. 587-590.
18. Tang, W., et al., *Secreted and membrane attractin result from alternative splicing of the human ATRN gene.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(11): p. 6025-6030.
19. Muotri, A., et al., *Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition.* Nature, 2005. **435**(7044): p. 903-910.
20. Cordaux, R., et al., *Estimating the retrotransposition rate of human Alu elements.* Gene, 2006. **373**: p. 134-137.
21. Huang, C., et al., *Mobile interspersed repeats are major structural variants in the human genome.* Cell. **141**(7): p. 1171-1182.
22. Izsvak, Z., et al., *Involvement of a bifunctional, paired-like DNA-binding domain and a transpositional enhancer in Sleeping Beauty transposition.* The Journal of biological chemistry, 2002. **277**(37): p. 34581-34588.
23. Plasterk, R., Z. Izsv√°k, and Z. Ivics, *Resident aliens: the Tc1/mariner superfamily of transposable elements.* Trends in genetics : TIG, 1999. **15**(8): p. 326-332.
24. Vigdal, T., et al., *Common physical properties of DNA affecting target site selection of sleeping beauty and other Tc1/mariner transposable elements.* Journal of molecular biology, 2002. **323**(3): p. 441-452.
25. Huang, X., et al., *Gene transfer efficiency and genome-wide integration profiling of Sleeping Beauty, Tol2, and piggyBac*

transposons in human primary T cells. Molecular therapy : the journal of the American Society of Gene Therapy. **18**(10): p. 1803-1813.

26. Cooley, L., R. Kelley, and A. Spradling, *Insertional mutagenesis of the Drosophila genome with single P elements.* Science (New York, N.Y.), 1988. **239**(4844): p. 1121-1128.
27. Ivics, Z., et al., *Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells.* Cell, 1997. **91**(4): p. 501-510.
28. Cui, Z., et al., *Structure-function analysis of the inverted terminal repeats of the sleeping beauty transposon.* Journal of molecular biology, 2002. **318**(5): p. 1221-1235.
29. Ivics, Z., et al., *Identification of functional domains and evolution of Tc1-like transposable elements.* Proceedings of the National Academy of Sciences of the United States of America, 1996. **93**(10): p. 5008-5013.
30. Izsvak, Z., *Involvement of a Bifunctional, Paired-like DNA-binding Domain and a Transpositional Enhancer in Sleeping Beauty Transposition.* Journal of Biological Chemistry, 2002. **277**.
31. Zayed, H., et al., *Development of hyperactive sleeping beauty transposon vectors by mutational analysis.* Molecular therapy : the journal of the American Society of Gene Therapy, 2004. **9**(2): p. 292-304.
32. Petrov, D., et al., *Diverse transposable elements are mobilized in hybrid dysgenesis in Drosophila virilis.* Proceedings of the National Academy of Sciences of the United States of America, 1995. **92**(17): p. 8050-8054.
33. Geurts, A., et al., *Gene transfer into genomes of human cells by the sleeping beauty transposon system.* Molecular therapy : the journal of the American Society of Gene Therapy, 2003. **8**(1): p. 108-117.
34. Yant, S., et al., *Mutational analysis of the N-terminal DNA-binding domain of sleeping beauty transposase: critical residues for DNA binding and hyperactivity in mammalian cells.* Molecular and cellular biology, 2004. **24**(20): p. 9239-9247.
35. Baus, J., et al., *Hyperactive transposase mutants of the Sleeping Beauty transposon.* Molecular therapy : the journal

of the American Society of Gene Therapy, 2005. **12**(6): p. 1148-1156.
36. Mates, L., et al., *Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates.* Nature genetics, 2009. **41**(6): p. 753-761.
37. Izsvak, Z. and Z. Ivics, *Sleeping beauty transposition: biology and applications for molecular therapy.* Molecular therapy : the journal of the American Society of Gene Therapy, 2004. **9**(2): p. 147-156.
38. Ivics, Z.n., et al., *The Sleeping Beauty transposable element: evolution, regulation and genetic applications.* Current issues in molecular biology, 2004. **6**(1): p. 43-55.
39. Zayed, H., et al., *The DNA-bending protein HMGB1 is a cellular cofactor of Sleeping Beauty transposition.* Nucleic acids research, 2003. **31**(9): p. 2313-2322.
40. Liu, G., et al., *Excision of Sleeping Beauty transposons: parameters and applications to gene therapy.* The Journal of Gene Medicine, 2004. **6**(5): p. 574-583.
41. Walisko, O., et al., *Sleeping Beauty transposase modulates cell-cycle progression through interaction with Miz-1.* Proceedings of the National Academy of Sciences of the United States of America, 2006. **103**(11): p. 4062-4067.
42. Izsvak, Z., et al., *Healing the wounds inflicted by sleeping beauty transposition by double-strand break repair in mammalian somatic cells.* Molecular cell, 2004. **13**(2): p. 279-290.
43. Izsvak, Z., Z. Ivics, and R. Plasterk, *Sleeping Beauty, a wide host-range transposon vector for genetic transformation in vertebrates.* Journal of molecular biology, 2000. **302**(1): p. 93-102.
44. Yant, S., et al., *Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system.* Nature genetics, 2000. **25**(1): p. 35-41.
45. Cattoglio, C., et al., *Hot spots of retroviral integration in human CD34+ hematopoietic cells.* Blood, 2007. **110**(6): p. 1770-1778.
46. Aiuti, A., et al., *Multilineage hematopoietic reconstitution without clonal selection in ADA-SCID patients treated with*

stem cell gene therapy. The Journal of clinical investigation, 2007. **117**(8): p. 2233-2240.
47. Bushman, F., *Retroviral integration and human gene therapy.* The Journal of clinical investigation, 2007. **117**(8): p. 2083-2086.
48. Yant, S., et al., *High-resolution genome-wide mapping of transposon integration in mammals.* Molecular and cellular biology, 2005. **25**(6): p. 2085-2094.
49. Hackett, P., et al., *Evaluating risks of insertional mutagenesis by DNA transposons in gene therapy.* Translational research : the journal of laboratory and clinical medicine.
50. Voigt, K., et al., *Retargeting sleeping beauty transposon insertions by engineered zinc finger DNA-binding domains.* Molecular therapy : the journal of the American Society of Gene Therapy. **20**(10): p. 1852-1862.
51. Ammar, I., et al., *Retargeting transposon insertions by the adeno-associated virus Rep protein.* Nucleic acids research. **40**(14): p. 6693-6712.
52. Liang, Q., et al., *Chromosomal mobilization and reintegration of Sleeping Beauty and PiggyBac transposons.* Genesis (New York, N.Y. : 2000), 2009. **47**(6): p. 404-408.
53. Dupuy, A., et al., *Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system.* Nature, 2005. **436**(7048): p. 221-226.
54. Geurts, A., et al., *Conditional gene expression in the mouse using a Sleeping Beauty gene-trap transposon.* BMC biotechnology, 2006. **6**: p. 30.
55. Hansen, G., et al., *Large-scale gene trapping in C57BL/6N mouse embryonic stem cells.* Genome research, 2008. **18**(10): p. 1670-1679.
56. Dupuy, A., et al., *A modified sleeping beauty transposon system that can be used to model a wide variety of human cancers in mice.* Cancer research, 2009. **69**(20): p. 8150-8156.
57. Copeland, N. and N. Jenkins, *Harnessing transposons for cancer gene discovery.* Nature reviews. Cancer. **10**(10): p. 696-706.
58. Tamhane, M. and R. Akkina, *Stable gene transfer of CCR5 and CXCR4 siRNAs by sleeping beauty transposon system to*

confer HIV-1 resistance. AIDS research and therapy, 2008. **5**: p. 16.

59. VandenDriessche, T., et al., *Emerging potential of transposons for gene therapy and generation of induced pluripotent stem cells.* Blood, 2009. **114**(8): p. 1461-1468.
60. Woltjen, K., et al., *piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells.* Nature, 2009. **458**(7239): p. 766-770.
61. Yant, S., et al., *Transposition from a gutless adeno-transposon vector stabilizes transgene expression in vivo.* Nature biotechnology, 2002. **20**(10): p. 999-1005.
62. Ehrhardt, A., et al., *A direct comparison of two nonviral gene therapy vectors for somatic integration: in vivo evaluation of the bacteriophage integrase phiC31 and the Sleeping Beauty transposase.* Molecular therapy : the journal of the American Society of Gene Therapy, 2005. **11**(5): p. 695-706.
63. Ohlfest, J., et al., *Combinatorial antiangiogenic gene therapy by nonviral gene transfer using the sleeping beauty transposon causes tumor regression and improves survival in mice bearing intracranial human glioblastoma.* Molecular therapy : the journal of the American Society of Gene Therapy, 2005. **12**(5): p. 778-788.
64. Liu, H., et al., *Sleeping Beauty-based gene therapy with indoleamine 2,3-dioxygenase inhibits lung allograft fibrosis.* FASEB journal : official publication of the Federation of American Societies for Experimental Biology, 2006. **20**(13): p. 2384-2386.
65. Hackett, P., D. Largaespada, and L. Cooper, *A transposon and transposase system for human application.* Molecular therapy : the journal of the American Society of Gene Therapy. **18**(4): p. 674-683.
66. Singh, H., et al., *Redirecting specificity of T-cell populations for CD19 using the Sleeping Beauty system.* Cancer research, 2008. **68**(8): p. 2961-2971.
67. Maiti, S., et al., *Sleeping beauty system to redirect T-cell specificity for human applications.* Journal of immunotherapy (Hagerstown, Md. : 1997). **36**(2): p. 112-123.

68. Montini, E., et al., *In vivo correction of murine tyrosinemia type I by DNA-mediated transposition.* Molecular therapy : the journal of the American Society of Gene Therapy, 2002. **6**(6): p. 759-769.
69. Chen, Z., et al., *Sleeping Beauty-mediated down-regulation of huntingtin expression by RNA interference.* Biochemical and biophysical research communications, 2005. **329**(2): p. 646-652.
70. Zhu, J., et al., *Erythroid-specific expression of beta-globin by the sleeping beauty transposon for Sickle cell disease.* Biochemistry, 2007. **46**(23): p. 6844-6858.
71. Aronovich, E., et al., *Prolonged expression of a lysosomal enzyme in mouse liver after Sleeping Beauty transposon-mediated gene delivery: implications for non-viral gene therapy of mucopolysaccharidoses.* The Journal of Gene Medicine, 2007. **9**(5): p. 403-415.
72. He, C.-X., et al., *Insulin expression in livers of diabetic mice mediated by hydrodynamics-based administration.* World journal of gastroenterology : WJG, 2004. **10**(4): p. 567-572.
73. Robbins, P.D.a.S.C.G., Viral vectors for gene therapy. Pharmacol Ther, 1998. 80(1): p. 35-47.
74. Verma, I.M. and N. Somia, *Gene therapy -- promises, problems and prospects.* Nature, 1997. **389**(6648): p. 239-42.
75. Kay, M.A. and H. Nakai, *Looking into the safety of AAV vectors.* Nature, 2003. **424**(6946): p. 251.
76. Clackson, T., *Regulated gene expression systems.* Gene Ther, 2000. **7**: p. 120-125.
77. Kole, R., A.R. Krainer, and S. Altman, *RNA therapeutics: beyond RNA interference and antisense oligonucleotides.* Nat Rev Drug Discov, 2012. **11**(2): p. 125-40.
78. Cirak, S.e.a.E.s.a.d.r.i.p.w.D.m.d.a.s.p.m.o.t.a.o.-l., phase 2, dose-escalation study. Lancet 378, 595-605, doi:10.1016/S0140-6736(11)60756-3 (2011).
79. Cirak, S., et al., *Exon skipping and dystrophin restoration in patients with Duchenne muscular dystrophy after systemic phosphorodiamidate morpholino oligomer treatment: an open-label, phase 2, dose-escalation study.* Lancet, 2011. **378**(9791): p. 595-605.

80. Goemans, N.M., et al., *Systemic administration of PRO051 in Duchenne's muscular dystrophy.* N Engl J Med, 2011. **364**(16): p. 1513-22.
81. Meir, Y.-J. and S. Wu, *Transposon-based vector systems for gene therapy clinical trials: challenges and considerations.* Chang Gung medical journal. **34**(6): p. 565-579.
82. Cadinanos, J. and A. Bradley, *Generation of an inducible and optimized piggyBac transposon system.* Nucleic acids research, 2007. **35**(12).
83. Owens, J., et al., *Chimeric piggyBac transposases for genomic targeting in human cells.* Nucleic acids research. **40**(14): p. 6978-6991.
84. Coffin, J., Hughes, S.H., Varmus, H.E., *Retroviruses*, ed. C.S.H.L. Press. 2000: Plainview.
85. Rosenberg, S.A., et al., *Human gene marker/therapy clinical protocols.* Hum Gene Ther, 2000. **11**(6): p. 919-79.
86. Miller, D.G., et al., *Large-scale analysis of adeno-associated virus vector integration sites in normal human cells.* J Virol, 2005. **79**(17): p. 11434-42.
87. Riviere, I., K. Brose, and R.C. Mulligan, *Effects of retroviral vector design on expression of human adenosine deaminase in murine bone marrow transplant recipients engrafted with genetically modified cells.* Proc Natl Acad Sci U S A, 1995. **92**(15): p. 6733-7.
88. Bonini, C., et al., *HSV-TK gene transfer into donor lymphocytes for control of allogeneic graft-versus-leukemia.* Science, 1997. **276**(5319): p. 1719-24.
89. Bordignon, C., et al., *Gene therapy in peripheral blood lymphocytes and bone marrow for ADA- immunodeficient patients.* Science, 1995. **270**(5235): p. 470-5.
90. Aiuti, A., et al., *Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning.* Science, 2002. **296**(5577): p. 2410-3.
91. Cavazzana-Calvo, M., et al., *Gene therapy of human severe combined immunodeficiency (SCID)-X1 disease.* Science, 2000. **288**(5466): p. 669-72.
92. Gaspar, H.B., et al., *Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector.* Lancet, 2004. **364**(9452): p. 2181-7.

93. Ott, M.G., et al., *Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1.* Nat Med, 2006. **12**(4): p. 401-9.
94. Mavilio, F., et al., *Correction of junctional epidermolysis bullosa by transplantation of genetically modified epidermal stem cells.* Nat Med, 2006. **12**(12): p. 1397-402.
95. Gardlik, R., et al., *Vectors and delivery systems in gene therapy.* Med Sci Monit, 2005. **11**(4): p. RA110-21.
96. Vigna, E. and L. Naldini, *Lentiviral vectors: excellent tools for experimental gene transfer and promising candidates for gene therapy.* J Gene Med, 2000. **2**(5): p. 308-16.
97. Naldini, L., et al., *In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector.* Science, 1996. **272**(5259): p. 263-7.
98. Zufferey, R., et al., *Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo.* Nat Biotechnol, 1997. **15**(9): p. 871-5.
99. Dull, T., et al., *A third-generation lentivirus vector with a conditional packaging system.* J Virol, 1998. **72**(11): p. 8463-71.
100. Zufferey, R., et al., *Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery.* J Virol, 1998. **72**(12): p. 9873-80.
101. Uitto, J.a.G.R., Progress in epidermolysis bullosa: from eponyms to molecular genetic classification. Clin Dermatol, 2005. 23(1): p. 33-40.
102. Uitto, J. and G. Richard, *Progress in epidermolysis bullosa: from eponyms to molecular genetic classification.* Clin Dermatol, 2005. **23**(1): p. 33-40.
103. Koss-Harnes, D., et al., *A site-specific plectin mutation causes dominant epidermolysis bullosa simplex Ogna: two identical de novo mutations.* J Invest Dermatol, 2002. **118**(1): p. 87-93.
104. Pulkkinen, L. and J. Uitto, *Hemidesmosomal variants of epidermolysis bullosa. Mutations in the alpha6beta4 integrin and the 180-kD bullous pemphigoid antigen/type XVII collagen genes.* Exp Dermatol, 1998. **7**(2-3): p. 46-64.

105. Varki, R., et al., *Epidermolysis bullosa. I. Molecular genetics of the junctional and hemidesmosomal variants.* J Med Genet, 2006. **43**(8): p. 641-52.
106. Cavazza, A. and F. Mavilio, *Gene therapy of skin adhesion disorders (mini review).* Current pharmaceutical biotechnology. **13**(10): p. 1868-1876.
107. Uitto, J.P., L., The genodermatoses: candidate diseases for gene therapy. Hum Gene Ther 2000, 11, (16), 2267-75.
108. Uitto, J. and L. Pulkkinen, *The genodermatoses: candidate diseases for gene therapy.* Hum Gene Ther, 2000. **11**(16): p. 2267-75.
109. Hengge, U.R., *Gene therapy progress and prospects: the skin--easily accessible, but still far away.* Gene Ther, 2006. **13**(22): p. 1555-63.
110. Uitto, J., *Progress in heritable skin diseases: translational implications of mutation analysis and prospects of molecular therapies*.* Acta Derm Venereol, 2009. **89**(3): p. 228-35.
111. Tamai, K., Y. Kaneda, and J. Uitto, *Molecular therapies for heritable blistering diseases.* Trends Mol Med, 2009. **15**(7): p. 285-92.
112. Morgan, J.R., et al., *Expression of an exogenous growth hormone gene by transplantable human epidermal cells.* Science, 1987. **237**: p. 1476-1479.
113. Fakharzadeh, S.S., et al., *Correction of the coagulation defect in hemophilia A mice through factor VIII expression in skin.* Blood, 2000. **95**(9): p. 2799-805.
114. Garlick, J.A., et al., *Retrovirus-mediated transduction of cultured epidermal keratinocytes.* J. Invest. Dermatol., 1991. **97**: p. 824-829.
115. Mathor, M.B., et al., *Clonal analysis of stably transduced human epidermal stem cells in culture.* Proc. Natl. Acad. Sci. USA, 1996. **93**: p. 10371-10376.
116. White, S.J., et al., *Long-term expression of human clotting factor IX from retrovirally transduced primary human keratinocytes in vivo.* Hum. Gene Ther., 1998. **9**: p. 1187-1195.
117. Titeux, M., et al., *SIN retroviral vectors expressing COL7A1 under human promoters for ex vivo gene therapy of recessive dystrophic epidermolysis bullosa.* Mol Ther, 2010. **18**(8): p. 1509-18.

118. Chen, M., et al., *Restoration of type VII collagen expression and function in dystrophic epidermolysis bullosa.* Nat Genet, 2002. **32**(4): p. 670-5.
119. Woodley, D.T., et al., *Injection of recombinant human type VII collagen restores collagen function in dystrophic epidermolysis bullosa.* Nat Med, 2004. **10**(7): p. 693-5.
120. Woodley, D.T., et al., *Normal and gene-corrected dystrophic epidermolysis bullosa fibroblasts alone can produce type VII collagen at the basement membrane zone.* J Invest Dermatol, 2003. **121**(5): p. 1021-8.
121. Woodley, D.T., et al., *Intravenously injected human fibroblasts home to skin wounds, deliver type VII collagen, and promote wound healing.* Mol Ther, 2007. **15**(3): p. 628-35.
122. Chen, M., et al., *Development and characterization of a recombinant truncated type VII collagen "minigene". Implication for gene therapy of dystrophic epidermolysis bullosa.* J Biol Chem, 2000. **275**(32): p. 24429-35.
123. Galla, M., et al., *Avoiding cytotoxicity of transposases by dose-controlled mRNA delivery.* Nucleic acids research. **39**(16): p. 7147-7160.
124. Schmidt, M., et al., *Detection and direct genomic sequencing of multiple rare unknown flanking DNA in highly complex samples.* Human Gene Therapy, 2001. **12**(7): p. 743-749.
125. Wu, X., et al., *Transcription start regions in the human genome are favored targets for MLV integration.* Science (New York, N.Y.), 2003. **300**(5626): p. 1749-1751.
126. Huang, X., et al., *Stable gene transfer and expression in human primary T cells by the Sleeping Beauty transposon system.* Blood, 2006. **107**(2): p. 483-491.
127. Cavazza, A., et al., *Self-inactivating MLV vectors have a reduced genotoxic profile in human epidermal keratinocytes.* Human Gene Therapy, 2013.
128. Burgess, D.J., *Genomics: Repetitive elements underestimated?* Nature Reviews Genetics. **13**(2): p. 74-74.
129. Emiliani S, M.A., Busschots K,, *Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome*

tethering and HIV-1 replication. Journal of Biological Chemistry, 2005(280): p. 6.

130. Cattoglio, C., et al., *High-definition mapping of retroviral integration sites defines the fate of allogeneic T cells after donor lymphocyte infusion.* PloS one. **5**(12).
131. Bushman, F., et al., *Genome-wide analysis of retroviral DNA integration.* Nature reviews. Microbiology, 2005. **3**(11): p. 848-858.
132. Xue, X., et al., *Stable gene transfer and expression in cord blood-derived CD34+ hematopoietic stem and progenitor cells by a hyperactive Sleeping Beauty transposon system.* Blood, 2009. **114**(7): p. 1319-1330.
133. Moiani, A., et al., *Lentiviral vector integration in the human genome induces alternative splicing and generates aberrant transcripts.* The Journal of clinical investigation. **122**(5): p. 1653-1666.
134. Ciuffi, A., et al., *A role for LEDGF/p75 in targeting HIV DNA integration.* Nature medicine, 2005. **11**(12): p. 1287-1289.
135. Cartier, N., et al., *Hematopoietic stem cell gene therapy with a lentiviral vector in X-linked adrenoleukodystrophy.* Science (New York, N.Y.), 2009. **326**(5954): p. 818-823.
136. Biffi, A., et al., *Lentiviral vector common integration sites in preclinical models and a clinical trial reflect a benign integration bias and not oncogenic selection.* Blood. **117**(20): p. 5332-5339.
137. Olson, W. and V. Zhurkin, *Working the kinks out of nucleosomal DNA.* Current opinion in structural biology. **21**(3): p. 348-357.
138. Belcher, J., et al., *Heme oxygenase-1 gene delivery by Sleeping Beauty inhibits vascular stasis in a murine model of sickle cell disease.* Journal of molecular medicine (Berlin, Germany). **88**(7): p. 665-675.
139. Liu, L., et al., *Sleeping Beauty-mediated eNOS gene therapy attenuates monocrotaline-induced pulmonary hypertension in rats.* FASEB journal : official publication of the Federation of American Societies for Experimental Biology, 2006. **20**(14): p. 2594-2596.
140. Thrasher, A., et al., *Gene therapy: X-SCID transgene leukaemogenicity.* Nature, 2006. **443**(7109).
141. Scaramuzza, S., et al., *Preclinical safety and efficacy of human CD34(+) cells transduced with lentiviral vector for the treatment of Wiskott-Aldrich syndrome.* Molecular

therapy : the journal of the American Society of Gene Therapy. **21**(1): p. 175-184.
142. Montini, E., et al., *Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration.* Nature biotechnology, 2006. **24**(6): p. 687-696.
143. Ott, M., et al., *Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1.* Nature medicine, 2006. **12**(4): p. 401-409.
144. Aiuti, A., et al., *Correction of ADA-SCID by stem cell gene therapy combined with nonmyeloablative conditioning.* Science (New York, N.Y.), 2002. **296**(5577): p. 2410-2413.
145. Avedillo Diezz, I., et al., *Development of novel efficient SIN vectors with improved safety features for Wiskott-Aldrich syndrome stem cell based gene therapy.* Molecular pharmaceutics. **8**(5): p. 1525-1537.
146. Sambrook, J. and D.W. Russell, *Molecular cloning: a laboratory manual.* CSHL press, 2001. **1**.

Acknowledgments

In questi tre anni di lavoro ho avuto il supporto di molte persone. Ringrazio il Professor Fulvio Mavilio per la fiducia riposta in me, per le sue indicazioni e incoraggiamenti. Ringrazio la Professoressa Recchia per il suo aiuto nelle questioni di laboratorio, per l'aiuto nel redigere questo lavoro, per la sua disponibilità e semplicità. Ringrazio i colleghi di laboratorio per i loro insegnamenti e per le risate insieme.
Ringrazio tutti i collaboratori esterni al laboratorio che hanno partecipato a questo lavoro: il laboratorio di Zoltan Ivics e Zsuzsanna Izsvak che mi ha ospitato nell'esperienza Berlinese, i Bioinformatici Cristian Taccioli per il trimming delle sequenze e Andreas Doring per il Blast.

Oltre che in queste occasioni formali, spero di riuscire ad esprimere la mia gratitudine quotidianamente alle persone a me vicine.
Gli amici "modenesi" sono stati una piacevolissima sorpresa e sono diventati famiglia per me. Ringrazio anche gli amici "milanesi" che mi hanno sostenuto nella Gomorra italiana rendendola addirittura piacevole. Ringrazio anche le mie sorelle per essersi sposate ma anche per l'aiuto nei momenti più difficili. Queste persone mi hanno sempre fatto pensare positivamente e anche se non se ne accorgono le osservo per prenderne esempio.
A voi tutti i miei più profondi acknowledgment di cuore.