

Published in final edited form as:

Mol Cancer Res. 2013 October ; 11(10): . doi:10.1158/1541-7786.MCR-13-0244.

Cancer gene discovery: exploiting insertional mutagenesis

Marco Ranzani^{1,*}, Stefano Annunziato¹, David J. Adams², and Eugenio Montini^{1,*}

¹San Raffaele Telethon Institute for Gene Therapy, via Olgettina 58, 20132, Milan, Italy

²Experimental Cancer Genetics, Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1HH, United Kingdom

Abstract

Insertional mutagenesis has been utilized as a functional forward genetics screen for the identification of novel genes involved in the pathogenesis of human cancers. Different insertional mutagens have been successfully used to reveal new cancer genes. For example, retroviruses (RVs) are integrating viruses with the capacity to induce the deregulation of genes in the neighborhood of the insertion site. RVs have been employed for more than 30 years to identify cancer genes in the hematopoietic system and mammary gland. Similarly, another tool that has revolutionized cancer gene discovery is the cut-and-paste transposons. These DNA elements have been engineered to contain strong promoters and stop cassettes that may function to perturb gene expression upon integration proximal to genes. In addition, complex mouse models characterized by tissue-restricted activity of transposons have been developed to identify oncogenes and tumor suppressor genes that control the development of a wide range of solid tumor types, extending beyond those tissues accessible using RV-based approaches. Most recently, lentiviral vectors (LVs) have appeared on the scene for use in cancer gene screens. LVs are replication defective integrating vectors that have the advantage of being able to infect non-dividing cells, in a wide range of cell types and tissues. In this review, we describe the various insertional mutagens focusing on their advantages/limitations and we discuss the new and promising tools that will improve the insertional mutagenesis screens of the future.

Keywords

Insertional mutagenesis, Cancer gene discovery; CG Carcinogenesis; CB09 Mapping and cloning of cancer genes; CB12 Functional genomics; VI Viral transformation and carcinogenesis

INTRODUCTION

Insertional mutagenesis is the phenomenon by which an exogenous DNA sequence integrates within the genome of a host organism. This event can result in the deregulation of genes in the neighborhood of the insertion site and can potentially cause a perturbation of cellular phenotype. When insertional mutagenesis induces the deregulation of oncogenes or tumor suppressor genes (TSG), it can cause cell transformation, and indeed insertional mutagenesis has been widely exploited for forward genetics screenings aimed at identifying

*To whom correspondence should be addressed: Marco Ranzani, San Raffaele-Telethon Institute for Gene Therapy, via Olgettina 58, 20132, Milan, Italy. Phone: ++390226439300; fax: ++390226435544, ranzani.marco@hsr.it; Eugenio Montini, San Raffaele-Telethon Institute for Gene Therapy, via Olgettina 58, 20132, Milan, Italy. Phone: ++390226433869; fax: ++390226435544, montini.eugenio@hsr.it.

ranzani.marco@hsr.it; annunziato.stefano@hsr.it; dal@sanger.ac.uk; montini.eugenio@hsr.it

Competing interest:

The authors declare that they do not have any conflict of interest.

novel cancer genes. In this setting, a mobile genetic element (usually a virus or transposon) is mobilized within an animal model system (usually the mouse) and, after a variable time depending on the agent and the animal model, cancer formation ensues. Retrieval of integration sites from tumors and mapping them to the genome allows the identification of cellular genes whose perturbation may have facilitated the promotion of tumor growth. Recurrently targeted sites in independent tumors (defined as Common Insertion Sites [CIS],) are identified as locations hosting candidate cancer genes. Therefore, in forward genetics screens for cancer gene discovery, the integrating mutagen represents both the culprit of cellular transformation and a traceable genetic tag.

In order to perform cancer gene discovery by insertional mutagenesis the requirements are: I) the proper oncogenic agent that efficiently integrates in a random or semi-random fashion into the host genome and deregulates the genes surrounding the insertion site (see the dedicated paragraph on CIS statistics); II) an animal model that is permissive to tumor formation; III) efficient technologies to perform retrieval of integration sites; IV) bioinformatic and statistical tools to map integrations onto the host genome and to identify sites that are enriched for insertions, and hence loci containing candidate cancer drivers. Gene identification represents the first step in the path towards understanding the function of a candidate cancer gene and defining its molecular mechanism of action. Then, once candidate cancer genes have been identified, it is also important to validate them and to define the likelihood that their orthologs drive human cancer or that they act in molecular pathways which are active in human tumors, and hence to prove their relevance to disease outcome in humans. Figure 1 outlines the general scheme of a cancer gene discovery approach by insertional mutagenesis.

Mechanisms of insertional mutagenesis

Integrating vectors can induce cancer by mutating host genes in a number of different ways. They can enhance transcription or translation levels of oncogenes, generate chimeric or truncated transcripts, or inactivate TSG expression.

Enhancer insertions are a common mechanism of mutagenesis, particularly with viruses, and result in the up-regulation of endogenous gene expression (Figure 2A). As regulatory elements may be distal to a gene, enhancer insertions can be located some distance from the proximal promoter and may affect the activity of elements via chromatin loops thus complicating the identification of the affected genes. Enhancer mutations may be independent from their orientation but are classically upstream of a mutated gene in the antisense orientation, or downstream in the sense orientation, especially for retroviruses (1, 2).

Promoter insertions occur when the insertional mutagen integrates in the sense orientation in or close to the proximal promoter region of an endogenous gene, thus uncoupling the cellular gene from its promoter and placing it under the control of elements found within the insertional mutagen (Figure 2B). In this case RNA polymerase transcribes an RNA starting from the promoter of the insertional mutagen that reads-through the introns and exons of the host gene. Frequently, the integrating vector contains splice signals that induce the joining of the first vector-coded exon to those of the host gene. This can result in the translation of high levels of chimeric transcripts (2, 3).

In addition to these modes of action, intragenic insertions can interfere with the splicing of genes into which they integrate (Figure 2C). Many insertional mutagens contain polyA signals, either engineered or endogenous, that may elicit premature termination of gene transcription. Truncation of transcripts in this way may have different consequences on the resulting protein and results in the expression of altered or neomorphic alleles. Viral

insertions in the 3' UTR region of a gene may remove mRNA-destabilizing motifs such as AUUUA hairpins or miRNA target sequences, as has been described to occur for *Pim1* and *N-myc* genes (4-6), resulting in increased levels of a truncated, but wild-type protein-encoding mRNA. Moreover, by the same mechanism insertional mutagens may induce the formation of a 3'-truncated mRNA by removing from the transcript protein coding exons. Alternatively, transcription may start from the integrated promoter, and in this case a 5'-truncated mRNA is transcribed. The resulting C-terminally or N-terminally truncated proteins may possess oncogenic properties and induce tumorigenesis (Figure 2C). The abnormal biological activity of the new mutant protein may be caused for example by a constitutively active kinase domain or a constitutively exposed dimerization domain (2, 7).

Using the same mechanisms, vector insertions can inactivate a gene: landing within a gene may result either in an mRNA encoding an inactive or unstable protein, or in aberrant splicing which abrogates gene function (Figure 2C). Hence, it is possible to detect candidate TSGs as well, even if they are found more rarely than oncogenes in insertional mutagenesis screenings that use retroviruses (5, 8) since usually the loss of both alleles is necessary to induce cellular transformation.

The integrating agents that are efficiently used so far to identify new cancer genes by insertional mutagenesis are retroviruses and transposable elements and are presented in the next sections.

RETROVIRUS-BASED INSERTIONAL MUTAGENESIS

Introduction to retroviruses

Retroviruses (RVs) are a large family of enveloped RNA viruses found in all vertebrates. The retroviral genome is a homodimer of linear, positive-sense, single-stranded RNA of 7 to 11 kilo-bases (Figure 3A), surrounded by a cone-shaped protein core. In the retroviral life-cycle, the genetic information goes from RNA to DNA and exists in two different forms, as genomic RNA when inside the viral particle and as proviral double-stranded DNA when integrated in the host genome. Both ends of the genome contain terminal non-coding sequences, the so-called Long Terminal Repeats (LTRs), composed of 5' and 3' unique sequences (U5 and U3 regions) and of two direct repeats (R) where the transcription start site (TSS) and the polyadenylation signals (polyA) are located (9).

Upon interaction between envelope glycoproteins and cellular receptors, fusion of the virus and host cell plasma membrane occurs, and the viral capsid containing the RNA genome enters the cell. The viral RNA genome is released in the cytoplasm and subsequently retrotranscribed into a double-stranded proviral DNA. After completion of reverse transcription, the proviral DNA is associated with viral proteins, and translocates to the nucleus, where the integrase mediates integration of the provirus into the host cell genome (9) (Figure 3B).

Oncogenic retroviruses have been classified in two groups, acute and slow transforming retroviruses. Acute transforming retroviruses induce, with short latency (2-3 weeks), the formation of polyclonal cancers by highly expressing virus-encoded oncogenes, while slow transforming retroviruses drive the formation of oligoclonal tumors with long latency (3-12 months) by inserting into the host genome proximal to oncogenes or TSGs (9).

Retroviruses as insertional mutagens

Insertional mutagenesis screens have for many years used two types of slow transforming ssRNA retroviruses: Moloney murine leukemia virus (MoMLV, a γ -retrovirus) and mouse mammary tumor virus (MMTV, a δ -retrovirus). These viruses cause lymphomas and

mammary tumors, respectively, in mice. The rationale of these screens is infecting newborn mice (post-natal day 1 or 2, when they haven't yet developed an immune system) with a viral supernatant, consequently causing a life-long viraemia: millions of host cells are infected and proviral insertions randomly spread all across their genome. When insertions result in either inactivation of a TSG or activation of an oncogene, the targeted cell may acquire a selective growth advantage and, after successive rounds of mutagenesis, an overt neoplasia may evolve. By cloning proviral insertions sites from tumor cells it is possible to identify candidate genes that have contributed to the development of the malignancy.

Usually, a single genomic integration is not enough to promote tumorigenesis and multiple rounds of proviral insertions within the same genome are required in order for a cancer to develop. Indeed, three or four insertions in or near different oncogenes and TSGs are often found in a single tumoral mass. Moreover, when insertions from single cells within an oligoclonal tumor are cloned, very often they share only some integrations between them, with many integrations being clone-specific: thus, it seems likely that tumorigenesis is the result of multiple rounds of viral insertion during the process of clonal expansion (5). Therefore, cells can undergo multiple rounds of infection and several mutations can accumulate in one cell, recapitulating the multi-step progression of human cancers.

Murine Leukemia viruses in hematopoietic malignancies

Moloney murine leukemia virus (MoMLV) is the prototypical murine leukemia virus (MLV) whose integrations have been extensively studied for the discovery of cancer genes in mice. The U3 promoter of MoMLV recruits the basal transcriptional machinery as it contains a TATA box and GC-rich sequences, while the enhancer element within the virus has binding sites for B- and T-cell specific transcription factors, such as ETS, NF1, RUNX and MYB. This probably accounts for the tropism of MoMLV and its propensity for lymphomagenesis: efficient viral replication and, therefore, gene deregulation, is restricted to cells providing transcription factors required for a high level of viral transcription(2).

MoMLV mainly induces T- and B-cell leukemia/lymphoma in mice. By analyzing MoMLV integrations in murine tumors, several oncogenes have been identified, including *c-Myc*, *Pim1*, *Pim2*, *Pvt1* (10). Many studies have created recombinant MLVs by substituting LTRs from other viruses and have tested them in mice: for example Starkey et al.(11) replaced the MoMLV LTRs with those of feline leukemia virus. This chimeric virus, named MoFe2-MuLV, mainly induced T-cell lymphoma. Interestingly, analysis of MoFe2-MuLV integrations showed that this virus did not target the same sites as MoMLV, and indeed led to the identification of new cancer genes, such as *Mf8t*, *Jundm2*, *Ahi1*, *Rras2* (12).

Different spontaneous mutants as well as engineered versions of MoMLV displayed the capability to induce other type of hematopoietic malignancies. The SL3-3 murine leukemia virus strain typically induces T-cell lymphomas, but a range of LTR mutations, such as mutation of RUNX binding sites, extends cancer latency and skews tumor formation causing myeloid, B-lymphoid and erythroid tumors (13). Other MLVs are specific for different hematopoietic compartments, for instance the erythroleukemia-inducing Friend virus is more active in erythroid cells, while the Graffi-1.4 virus induces mainly acute myeloid leukemia (AML)(14). A different approach is to use a specific mouse model to skew the tumor spectrum. For instance, using MoMLV in transgenic *Eμ-Myc* mice induced the development of B-cell lymphoma and allowed the identification of additional cancer genes that collaborate with MYC overexpression in this tumor (15).

Instead of injecting a viral supernatant in newborn mice, other studies identified CIS in tumors derived from the recombinant inbred strains BXH2 and AKXD. In BXH2 mice, the MLV is not inherited through the germline, rather is horizontally transmitted via

transplacental infection of implanted embryos (16). In the AKXD strains the MLV is inherited through the germline but then somatic activation of additional recombinant viruses occurs (17). Therefore the virus starts to replicate and integrates into the genome during early gestation with mutagenesis continuing throughout the life of the animal causing a life-long viraemia. Cells of the immune system, showing a high proliferative rate during early postnatal development, are the preferred compartment for viral propagation. BXH2 and AKXD mice spontaneously develop at high frequency myeloid and B-lymphoid tumors respectively, so they represent valuable models to identify cancer genes involved in these malignancies. Indeed, studying and comparing the CIS retrieved from these mice allowed identifying lineage-specific cancer genes involved in hematopoietic malignancies that are also relevant in human tumors (18, 19).

MMTV in mammary cancer

Mouse mammary tumor virus (MMTV) is a retrovirus with a similar tumorigenic behavior as MoMLV but with a specific tropism for mammary cells. Since carcinomas represent the most frequent human malignancies, the ability of MMTV to hit the epithelial compartment is valuable and has resulted in the identification of many clinically relevant cancer genes. The MMTV life cycle usually begins with the ingestion of infected milk by pups from their viraemic mother. After a few days, the virus infects B cells in lymphoid tissue of the gut, e.g. Peyer's patches. Here, by stimulating a T cells antigen response, a pool of B and T MMTV-infected cells is generated as a reservoir of virus, from which MMTV spreads into different tissues ultimately resulting in transformation of mammary epithelial cells. MMTV-infected mice develop hormone-dependent preneoplastic lesions, which eventually evolve into hormone-independent mammary adenomas and carcinomas.

MMTV has been employed for forward genetic studies since 1982, when Nusse and Varmus identified *Wnt1* as a frequent insertion target (20). Different studies performed with MMTV-induced tumors led to the identification of 12 candidate cancer genes, mainly belonging to *Wnt* and *Fgf* pathways (21). However, proportionally fewer studies have employed MMTV than MoMLV, but recently Theodorou et al. performed the first large-scale MMTV insertional mutagenesis screen in the mouse mammary gland (22). 160 mammary tumors were screened identifying 33 CIS, 21 of which were previously unknown MMTV targeted genes. More recently, thanks to technological improvements, a deep analysis of integrations has widened the spectrum of MMTV-identified CIS and allowed studying tumor heterogeneity (23, 24).

Retroviruses for solid tumors

Besides MMTV in mammary cancer, there is another solid tumor-causing slow-transforming retrovirus which has been employed for insertional mutagenesis screens, the myeloblastosis-associated virus type 2 (MAV-2). MAV-2 is an avian replication-competent non-acute oncogenic retrovirus. In chickens infected as embryos or early after hatching with this avian retrovirus, nephroblastomas form (25). Nephroblastoma is a type of kidney tumor similar to the Wilms tumor, the most frequent human renal neoplasia of childhood. These studies showed that the *Twist* gene, a transcription factor that is thought to inhibit p53-dependent apoptosis, was the predominant common site of proviral insertion in tumor cells. Moreover, 18 candidate cancer genes were found, several of which were deregulated in human pediatric kidney tumors or other type of human tumors (26, 27).

Other lines of research have used a modified MoMLV to induce tumors in non-hematopoietic cell types, by combining tissue-specific viral delivery and oncogene transfer. Autocrine activation of platelet-derived growth factor B-chain (PDGF β) receptor is a key step in human glioblastoma formation. By intracerebral injection of a replication defective

MoMLV encoding PDGF in newborn mice, Uhrbom et al. induced the development of gliomas. Viral propagation was enabled by co-infection of a replication-competent MoMLV (28). Exploiting this model of glioma, Johansson and colleagues were able to identify 66 candidate cancer genes. Some of these genes probably synergize with PDGF signaling in the induction of gliomagenesis. As for MMTV results, the genes targeted in this study only partially overlapped with the ones previously identified with native MoMLV in leukemia. The spectrum of MoMLV-induced brain tumors mimics the kinetic and the molecular lesions of human glioblastoma multiforme and oligodendroglioma. Importantly, some of the newly identified cancer genes were found to be deregulated in human brain tumors (29, 30).

Despite their efficiency in inducing multiple rounds of oncogenic integrations in the hematopoietic system and mammary gland, allowing the identification of novel cancer genes, the limited tissue tropism of retroviruses has significantly limited their applicability at other solid tumor sites.

TRANSPOSON-MEDIATED INSERTIONAL MUTAGENESIS

Transposons are DNA sequences that can move from one location on the genome to another. There are two general classes of transposons: retrotransposons and DNA transposons. Retrotransposons move by a “copy and paste” process (replicative transposition), transposing via an RNA intermediates which is then converted to DNA by retrotranscription, which then inserts in new locations in the genome. Some attempts to exploit retrotransposons as insertional mutagens in mice have been performed (31, 32), but the low integration efficiency, the integration of incomplete retrotranscribed elements and the concomitant induction of chromosomal aberrations are still limiting their applicability to cancer gene screenings(33, 34).

DNA transposons are a class of “cut and paste” transposons that rely on a transposase enzyme, which recognizes specific DNA sequences and “cuts” the DNA between them. The excised DNA is then re-integrated at another site in the genome. DNA transposon-mediated insertional mutagenesis screenings have been developed in the last decade and allow the generation of tumors in a wider spectrum of tissues than the ones that are accessible using retroviruses (8, 35-38).

The *Sleeping Beauty* system

Since DNA transposons are actively mobile only in plants and invertebrates, the *Sleeping Beauty* (SB) transposon system was generated by reverse engineering in 1997, based on sequence comparison between multiple salmonid species of non-functional transposase genes of the *Tc1/mariner* family that had accumulated inactivating mutations (39). The SB system is composed of two elements: a transposase, i.e. the enzyme responsible for mobilization, and the transposon, i.e. the actual mobilized sequence of DNA, which is flanked by the binding sites for the SB transposase, the so-called inverted repeats/direct repeats (IR/DRs) (Figure 4A). Transposition occurs when the transposase binds two sites in each IR/DR, and the closer the IR/DRs, the more efficient transposition will be (40). When the transposase excises a transposon, it leaves behind a three base footprint. The transposon can then reinsert at any location in the genome where a TA dinucleotide is present (there are more than 300 million TA sites in the genome). During integration, the TA is duplicated (Figure 4B). The cargo of the transposon (the segment contained between the two IR/DRs) can be any sequence of choice, but transposition efficiency decreases with increased cargo sizes (40). By introducing point mutations in the transposase gene, it has been possible to significantly increase the level of mobilization (38, 41, 42).

Transposition may be controlled by separating the transposase from the transposon, thus establishing a non-autonomous bi-partite transposon system. In such a bipartite system, the transposon can be mobilized only when the transposase protein is expressed. This may be achieved by generating two transgenic mouse strains: the first is called the “jumpstarter” strain expressing the transposase gene; the second is the “mutator” strain carrying the non-autonomous transposon. Upon crossing the jumpstarter strain with the mutator strain, the transposon is mobilized in the soma of the resulting offspring and insertional mutagenesis ensues.

Cancer gene discovery using the *Sleeping Beauty* system

The mechanisms by which transposons can induce insertional mutagenesis are essentially the same as described for retroviruses. For cancer gene discovery, the SB cargos which have been employed are mutagenic elements designed to mimic and potentiate those present in retroviruses. The “oncogene trap” of SB transposons can induce loss-of-function mutations in TSGs, as well as gain-of-function mutations in proto-oncogenes. Since SB is a cut and paste transposon and there is only a 40-50% probability that an excised transposon will reintegrate into the genome, once transposition is initiated the number of transposons integrated in the genome will decrease over the time. In order to mutate enough cooperating cancer genes to induce cancer before transposons disappear from the genome, a large number of transposable elements is required. Therefore, in mutator mouse strains, the transposons are arranged in a reservoir of head-to-tail concatemered copies resident at an inert chromosomal locus. As described above, transposase activity mobilizes the transposons from the concatemer and as transposons reintegrate they elicit cancer gene mutations (Fig. 5A).

The utility of SB for oncogenic insertional mutagenesis was first demonstrated by two groups in 2005 (43, 44). In both studies, transposons were introduced into mice by pronuclear injection of a linear plasmid bearing a copy of the SB transposon, which integrates in a genome locus in the form of a multicopy concatemer. In order to mobilize the transposons, the resulting mouse strain was crossed to a mouse strain expressing the transposase gene from a ubiquitous promoter. Both groups used essentially the same transposon, named T2/Onc (44) and T2/Onc2 (43). The study by Collier et al. used a transgenic mouse carrying the “original” SB10 transposase under the control of the CAGGS promoter (composed of sequences from the chicken β -actin and human cytomegalovirus immediate early promoters) to mobilize T2/Onc from a low-copy (~25 copies) concatemer. The study by Dupuy et al used a knock-in mouse in which the more active SB11 transposase had been knocked into the endogenous *Rosa26* locus by homologous recombination in mouse embryonic cells and these mice were crossed with a strain carrying a high-copy (150-350 copies) concatemer of the T2/Onc2 transposon. Both T2/Onc and T2/Onc2 transposons are “oncogene traps” engineered with *engrailed-2* (*En-2*) SA sites followed by the simian virus 40 (SV40) pA sequence in both orientations, to intercept upstream SD after intronic insertions and prematurely terminate transcripts running into the transposon. Moreover, they contain the murine stem cell virus (MSCV) 5 LTR followed by the *FoxF2* SD site, in order to create fusion transcripts after splicing with downstream endogenous exons (Figure 5A). The only difference between T2/Onc and T2/Onc2 is that T2/Onc2 contains a larger fragment of a SA sequence and is flanked by optimized SB transposase binding sites that increase SB transposition. Substantially, both transposons are capable of activating oncogenes by enhancer and promoter insertions and by 5' and 3' truncations, and are capable to inactivate TSGs. However, unlike retroviruses, there is little evidence that T2/Onc2 is able to induce enhancer insertion, even if it contains MSCV sequences. Therefore, SB insertions targeting oncogenes are almost always oriented in the same transcriptional orientation of the host gene, usually near 5' end or in an intron. Conversely, insertions in

TSGs show little directional bias and are scattered throughout the coding region (8) (Fig. 5A).

Doubly-transgenic mice (for transposon and transposase) generated by Dupuy et al. on a wild-type background were subject to a high level of embryonic lethality. Only 24 mice survived until birth, but all succumbed to cancer within 17 weeks of age. Most tumors were T-cell lymphomas, but there were also other hematopoietic malignancies and, in a few cases, medulloblastomas (solid tumors of the cerebellum) and duodenal and pituitary neoplasias. In contrast, in the study by Collier et al., double transgenic mice did not develop tumors spontaneously on a wild-type background, but sarcomagenesis was accelerated in mice deficient for the TSG *p19Arf*. The majority of T2/Onc;CAGGS-SB10; *Arf*^{-/-} mice developed soft tissue sarcomas (similar to what had been observed previously in *Arf*^{-/-} mice (45)), but osteosarcomas, lymphomas, myeloid leukemias, malignant meningiomas and a pulmonary adenocarcinoma were also observed. Both studies allowed the identification of several candidate cancer genes, including new cancer genes relevant for the human disease.

The T2/Onc and T2/Onc2 both contain the 5 LTR of the murine stem cell virus (MSCV) derived from a cloned MoMLV and therefore promoting oncogene activation preferably in the hematopoietic compartment. Dupuy et al. (46) recently generated a novel mutagenic transposon, T2/Onc3, whose structure is identical to that of T2/Onc2 except that the MSCV LTR is replaced with the cytomegalovirus immediate early enhancer fused to the CAG promoter. The CAG promoter was chosen as it had been shown to strongly drive transcription in epithelial cells but was weak in the hematopoietic lineage. Double transgenic mice developed a wide array of carcinomas (in the liver, skin, lung, colon, ovary, parathyroid, mammary gland, salivary gland and at many other sites). Liver adenomas, squamous cell carcinomas of the skin, hepatocellular carcinomas and lung adenomas were the most common tumors. Integration site analysis was performed on 17 squamous cell carcinomas of the skin and 11 hepatocellular carcinomas and led to the identification of three and two CIS respectively. This approach was capable to shift the tumor spectrum from mostly lymphomas to primarily carcinomas. However, in these experimental settings, tumors from different tissues compete in causing lethality of the host. Therefore, tumors characterized by slow progression and late onset may be elusive with this approach. Moreover, analysis of a large collection of tumors of the same type may be more informative in understanding tissue specific pathways of carcinogenesis.

Tissue-specific Cre-inducible SB mutagenesis

In order to direct mutagenesis to specific tissues, a strategy based on Cre-mediated activation of the SB transposase has been developed. In this setting, the SB transposase cDNA is preceded by a stop cassette surrounded by loxP sites; recombination of the loxP sites by Cre places the transposase under the control of the *Rosa26* promoter. When combined with Cre expressed in a target tissue, this approach allows inducing tissue-specific insertional mutagenesis. Importantly many tissue-specific Cre-transgenic mouse strains are available providing almost endless opportunities for tissue directed mutagenesis. In 2009, two papers were published describing transposon-based genetic screens in which transposition was confined to specific cell types using a conditional SB allele and a tissue-specific Cre-recombinase (47, 48). One of these studies (47) was performed using triple-transgenic mice carrying the canonical T2/Onc concatemer, the *Rosa26*-SB11 transposase cDNA preceded by a loxP-flanked stop cassette and a Cre transgene under the control of the Villin promoter. The Villin-Cre elicited transposase expression in epithelial cells of the gastrointestinal tract and a high proportion of mice died before 18 months, harboring intraepithelial neoplasias, adenomas and carcinomas in the small and large intestines. Integration analysis led to the identification of 77 CIS from 135 intestinal tumors. Two following studies (49, 50) used the same SB system in mice heterozygous for loss-of-

function mutations in *Apc* in the germ-line or somatically induced by Cre recombination. Another study (48) used triple-transgenic mice with the same three transgenes in order to limit transposition to the liver, but in that case Cre was under the control of the hepatospecific Albumin promoter. As mutations in *P53* are the most frequent lesions in human HCC, quadruple-transgenic mice were also bred to carry a conditional dominant negative *Trp53* allele. Mice from this model developed different types of liver lesions, ranging from preneoplastic nodules to hepatic adenomas and HCCs with lung metastases. Collectively these studies identified several candidate cancer genes of relevance to human HCC and colorectal cancer.

Similar approaches have been used to model other types of human malignancies. Using a transgene in which the Cre cDNA was fused to the initiation codon of the activation-induced cytidine deaminase (AID) gene, conditional transposase activation was used to drive B-cell malignancies, including diffuse large cell lymphomas and follicular lymphomas through insertional mutagenesis (46). Using different T cell differentiation stage-specific promoters, SB insertional mutagenesis has also been used to investigate the impact of the cell of origin in the development of T acute lymphoblastic leukemia (T-ALL) (51). In a similar way a recent study exploited inducible tissue-specific activation of SB transposition in a mouse model of melanoma (52). Here a tissue-specific tyrosinase promoter (active in melanocytes) was used to drive a tamoxifen inducible Cre. Topical administration of tamoxifen induced melanocyte-specific transposition that when combined with a conditional *Braf*^{V619E} allele (analogous to human *BRAF*^{V600E}) induced melanogenesis. This screen led to the discovery of several competing endogenous RNAs (ceRNAs) behaving as tumor suppressors by acting as “sponges” for microRNAs targeting *Pten*.

In another screen Vassiliou and colleagues used a sophisticated system to study the genes that cooperate with mutation in nucleophosmin (*NPM1*) to induce acute myeloid leukemia (53). Since a dominant mutation in *NPM1* is one of the most frequent mutations in human AML, they generated a mouse strain that expresses a humanized mutated *Npm1* allele upon Cre activation. The transposon used was similar to T2/Onc but containing the LTR of Graff1.4 MLV that preferentially promotes myeloid leukemia rather than lymphoid leukemia. In this model the expression of mutant *Npm1* and the SB transposase was induced using the hematopoietic specific *Mx1*Cre allele. Administration of polyinosinic:polycytidylic acid activated the Cre and resulted both in the expression of the mutated *Npm1* and the activation of SB insertional mutagenesis.

Following a different approach, a model of B cell precursor acute lymphoblastic leukemia (BCP-ALL) was created by combining the expression of the oncogenic fusion product *Etv6-RUNX1* with SB insertional mutagenesis (54). *ETV6-RUNX1* is the result of a translocation and it is present in around 25% of BCP-ALL; however, its overexpression in mice is unable to recapitulate the human disease since it requires co-operating mutations to cause overt leukemia. Therefore, van der Weyden and colleagues generated a knock-in mouse model that expresses *Etv6-RUNX1* fusion product and a hyperactive version of SB transposase (HSB5) under the control of *Etv6* promoter (and separated by an Internal Ribosome Entry Site sequence). Crossing this mouse with a strain transgenic for T2/Onc resulted in insertional mutagenesis restricted only to the cells that express the oncogenic fusion product *Etv6-RUNX1*. These mice developed leukemias with an early onset and a phenotype significantly skewed to BCP-ALL compared to mice carrying the *Etv6-RUNX1* alone.

Recently SB insertional mutagenesis has been widely employed to screen for cancer genes also in other tissues, including pancreas (55, 56), central (57-59) and peripheral nervous system (52, 60), thus widening the spectrum of tumors that can be studied.

SB insertional mutagenesis to characterize the process of metastasis

SB insertional mutagenesis may also be exploited to track and characterize the process of metastasis. Indeed, in contrast to retroviruses that can spread widely within the host organism, transposons undergo multiple cycles of mobilization and re-integration without an extracellular phase. Thus the profile of insertions found within a metastatic tumor denotes the primary from which it originated. Such an approach was used to track the origin of individual lung metastasis that had originated from SB-induced HCCs (48). As transposition continues as tumors disseminate, drivers of disease spread may be identified. In a recent study insertional mutagenesis was performed by activating SB in the cerebellar progenitor compartment of *Ptch*^{+/-} or *Trp53*-mutated mice which caused a significant increase in the incidence of medulloblastoma formation and of spread of these tumors to distal sites. A comparative analysis of SB integrations from these tumors made it possible to identify primary and metastatic tumor pairs, and revealed that advanced disease originates from specific subclones within the primary tumor that undergo further evolution with metastasis (57).

The PiggyBac system for insertional mutagenesis and cancer gene discovery

Other transposable elements have been developed and exploited in mice, such as Minos (from *Drosophila hydei*) and Tol2 (from *Oryzias latipes*, the Medaka fish). However, to date the only transposon that has been efficiently exploited for cancer gene discovery as an alternative to SB is the *PiggyBac* system. *PiggyBac* (PB) is a DNA transposon from the cabbage looper moth *Trichoplusia ni* that has been developed to be active in mammalian cells, including mice (61). The main differences between SB and PB are: PB system can efficiently mobilize larger cargos (up to several hundred Kb); PB integrates at a TTAA tetranucleotide sequence; upon excision PB does not leave any footprint, although imprecise excision events may damage the genome at the mobilization site. Importantly, PB displays a bias to integrate inside genes, a potential advantage in screens for cancer genes (62).

The main work that used PB insertional mutagenesis to identify cancer genes in mice was published in 2010 in *Science* (63) and deployed a bi-functional transposon containing 5' and 3' consensus sequences for both the SB and PB transposases. The design of the integrating cassettes used was similar to T2/Onc transposon since they contained a promoter, SD, SA, and polyA sites to generate gain-of-function and loss-of-function mutations. Three different types of transposons were developed each containing a different enhancer/promoter: MSCV, CAG, or the phosphoglycerate kinase promoter (PGK) (Figure 5B). Nineteen different transposon lines were produced from these three different transposons each containing a different number of copies resident at different chromosomal loci. Fourteen of these lines were crossed with mice that constitutively express PB transposase from the *Rosa26* locus. As for SB (43), they observed a high rate of embryonic lethality in mouse strains that contain a high copy number of transposons. However, from strains with intermediate to low copy number transposon arrays tumors formed. The tumor latency and the tumor type were found to be profoundly affected by the type of transposon. Mobilization of the MSCV transposon resulted in >90% hematopoietic tumors. Conversely, the CAG transposon mainly induced solid tumors, including sarcomas and various carcinomas with poor differentiation, and in some cases metastasis was observed.

Mice carrying mobilized PGK transposons developed hematopoietic and solid tumors, with several mice bearing both types of malignancies. Analysis of 5590 non-redundant transposon integration events resulted in the identification of 72 CIS. Analogous to SB, gain-of-function events were found near oncogenes, while TSGs contained intragenic loss of function mutations. Remarkably, 42% of the candidate cancer genes identified by PB insertional mutagenesis were not significantly hit in previous studies performed using

retroviruses or SB transposons, suggesting that PB targets a unique spectrum of loci. To restrict the activity of PB transposase, Cre regulated systems have been developed that are essentially identical to those developed for SB (64). Additionally, the versatility of the PB transposase has allowed the generation of a fusion PB-estrogen receptor protein allowing for temporal control of transposition (65, 66). The optimization of PB-based systems for cancer gene discovery in mice is still ongoing at the time of writing of this review.

LENTIVIRAL VECTOR-BASED INSERTIONAL MUTAGENESIS

Introduction to lentiviral vectors

Lentiviral vectors (LVs) have been developed by engineering the genome of the HIV virus. The replication cycle of lentiviruses is essentially similar to that of retroviruses, despite the fact that the lentiviral pre-integration complex is capable of integrating into cells that are non-replicating. Exploiting this feature of LVs has allowed them to be used as efficient tools for gene transfer for gene therapy applications.

Since LVs have been developed for gene therapy purposes, they have been designed for gene transfer without the requirement for the full viral life-cycle, which has significant advantages in terms of their biosafety for clinical applications. The production of replication defective LVs involves the construction of a transfer vector that contains a promoter driving a transgene of interest and the sequences that provide signals for vector packaging, reverse transcription, and integration (9). Thus LVs include: (1) a promoter and polyadenylation signal for the viral genome; (2) a viral packaging signal (Ψ) to direct incorporation of vector RNA into virions; (3) signals required for reverse transcription, including a transfer RNA-binding site (PBS), a polypurine tract (PPT) for the initiation of first- and second-strand DNA synthesis, a repeated (R) region at both ends of the viral RNA required for transfer of DNA synthesis between templates, and a central polypurine tract (cPTT) from the polymerase open reading frame (ORF) that supports reverse transcription and nuclear translocation of the proviral DNA (67, 68). All the other sequences that encode for structural proteins of HIV have been eliminated from the vector genome. Packaging of LVs is achieved by providing the structural and regulatory proteins *in trans* by transfection into a packaging cell line, together with the vector genome (Figure 6).

In the last years several improvements have been made to LV packaging systems by modifying either the packaging construct in order to increase their biosafety, or by making the vector genome self-inactivating (SIN), so that upon integration LVs lose the transcriptional capacity of their LTRs. Such modification is achieved by deleting the enhancers and promoter sequences within the LTRs of the viral genome, thus generating a transcriptionally inactive integrated provirus that lacks enhancer and promoter sequences in both the LTRs (9, 69). This modification prevents transcription of the full vector genome and thus a possible mobilization of the vector. However, any promoter internal to the LTRs in such vectors will still be active to drive the expression of the gene of interest (67, 70, 71).

Another important aspect for LV development is the viral cellular tropism, which is dictated by the interactions of the viral envelope glycoproteins with cellular receptors on the target cells. LV have the ability to incorporate in their envelope proteins from both related and unrelated viruses when they are provided *in trans* during LV production, in a process referred as pseudotyping. The use of a different envelope can expand or selectively restrict the host-range of the vectors. A common vector pseudotype employed is the G glycoprotein of the vesicular stomatitis virus (VSV-G). It confers pantropism to the vector, since it mediates fusion of the viral envelope with the cellular membrane and increases particle stability, thus allowing vector concentration by ultracentrifugation.

Production of LVs is achieved by the transient transfection into 293T cells (commonly used as a transient packaging cell line) of 4 plasmids which encode respectively: I) the structural protein of the virions, II) a regulatory protein Rev essential for efficient vector production, III) the envelope protein that confers transduction specificity; IV) the vector genome (Figure 6). After transfection, viral particles containing RNA sequences encoded by the transfer plasmid are released into the medium by producer cells and can be collected for centrifugation and storage (68).

Since their RNA genome does not encode for the structural proteins that are essential for the assembly of the viral particles, LVs are devoid of any replication potential. The resulting VSV-G pseudotyped LVs are therefore capable to efficiently transduce *in vivo* replicating and non-replicating cells from different tissues, thus providing an efficient single round of integration.

Specifically modified lentiviral vectors can induce insertional mutagenesis

Since LVs represent a powerful tool for gene therapy and recent clinical trials using γ -retroviral vectors (γ RVs) have raised several concerns over the potential of these vectors to act as insertional mutagens, the biosafety profile of different γ RV and LV designs has been compared (72, 73). In an attempt to identify the molecular features that have a major impact on vector genotoxicity, a panel of vectors was generated by swapping sequences from γ RVs to LVs and moving enhancer/promoter sequences from LTRs to a position within the vector genome. The genotoxicity of these vectors was tested in a sensitive model of *ex vivo* transduction and transplantation using *Cdkn2a*^{-/-} hematopoietic stem cells (HSC). Acceleration of tumor onset was achieved by comparing to mice that received mock-treated cells. Integration profiles were generated for each integrating vector following transduction. Using this approach, it was shown that transcriptionally active LTRs are major determinants of genotoxicity even when reconstituted in LVs, and that self-inactivating LTRs enhance the safety of γ RVs. By comparing the genotoxicity of vectors with matched design, γ RVs displayed a significantly higher genotoxicity compared to LVs, which is probably due to the specific biases of integration that are typical of γ -retroviruses (73). Therefore, despite their significantly improved biosafety profile when used with the SIN design for gene therapy applications (72-81), specifically tailored LVs carrying strong enhancer/promoter sequences in the LTRs displayed a detectable genotoxicity and could be used to induce insertional mutagenesis, recapitulating the mechanisms previously described for retroviruses (72, 73).

Besides the capability to induce insertional mutagenesis in tumor prone HSC, LVs have other features that make them promising tools for cancer gene discovery. Previous works showed that LVs can induce insertional mutagenesis by both inducing gain-of-function (73) and loss-of-function mutations (82). The LV integration pattern has been extensively characterized: they have a strong bias for integration within actively transcribed genes (83-85). Moreover, LVs can interact with the splicing machinery of the host gene resulting in the generation of truncated transcripts (86, 87). Factors that are likely to increase the chances of deregulating gene function. Moreover, VSV-G pseudotyped LVs are very stable and pantropic and have the intrinsic capability of integrating in non-replicating cells (70), thus allowing efficient *in vivo* transduction of different tissues. Additionally, since they have been extensively engineered for gene therapy, it is possible to design LVs to be active in different specific tissues by simply swapping the enhancer/promoter sequences.

Lentiviral vector-based insertional mutagenesis for cancer gene discovery

We have recently developed a LV-based insertional mutagen and exploited it to perform a screening for new liver cancer genes(88). We constructed a transgene-less LV with highly-active hepatospecific enhancer-promoter sequences (Enhanced Transthyretin (89)) in the

LTR in order to activate genes upon integration in hepatocytes while avoiding unwanted effects in non-parenchymal cells. By a single injection of this LV into newborn mice, we efficiently induced HCCs in three different mouse models and generated a collection of 30 HCCs covering different grades of HCC (from G1 to G3) and displaying gene expression signatures reminiscent of those found in human HCCs.

Although the retrieval of vector integration resulted in low number of univocally mapped integrations sites (172 on 30 HCCs) compared to previous studies with continuously integrating vectors, we identified 4 CIS. All the 4 newly identified liver cancer genes were validated *in vivo* in mice. Importantly, all the 4 novel oncogenes were found to be significantly deregulated and/or amplified or deleted in human HCCs. Moreover, when we stratified human HCC patients according to the expression level of the novel cancer genes or to the gene expression signatures induced by their up-regulation (as characterized in murine tumors), we identified groups of patients characterized by a differential survival. These findings indicate that the newly identified liver cancer genes may represent new prognostic markers and therapeutic targets for human HCC.

In this LV-based screening we observed a strong and unprecedented enrichment for tumors bearing a single CIS integration (83%), when compared to other insertional mutagenesis systems. This high efficiency of CIS retrieval is likely the result of the combination of the non-replicative nature of LVs, the high efficiency of LV transduction in hepatocytes, and the efficient coverage of genes by integrations. The LV integrations are produced only in a short time window after injection and before the *in vivo* selection of transformed clones occurs. These characteristics result in a decreased number of total integrations that may reduce the incidence of tumor induction and the total yield of identified cancer genes, which may represent a limit of LVs when compared to transposons and retroviruses. On the other hand, this exclusive peculiarity of LVs may facilitate the identification of early-mutated genes in carcinogenesis by eliminating the bystander and progression-related integrations that are accumulated during the entire life-time of the mouse with the autonomously replicating systems used so far.

Although they have been used just recently as insertional mutagens, the intrinsic versatility, the wide tissue tropism, and the high *in vivo* transduction efficiency illustrates the great potential of LVs for cancer gene discovery *in vivo*.

METHODS FOR THE RETRIEVAL AND ANALYSIS OF INTEGRATIONS

Retrieval of the integrations sites

The isolation of insertion sites from tumors generated using viral or transposon vectors is a vital step in the process of identifying new cancer drivers. Classically splinkerette PCR (S-PCR)(90) and linear amplification mediated PCR (LAM-PCR) (91) -based techniques have been used to amplify the junction between the end of the transposon or provirus and the surrounding cellular genome via ligation of a linker-cassette sequence by a restriction enzyme (RE) site, with the products of these reactions being sequenced on capillary(92) or next-generation sequencing platforms(91, 93). Since tumors generated using insertional mutagens are frequently poly or oligoclonal, variants of the classical S-PCR or LAM-PCR protocols that avoid biases due to distribution of RE sites have been developed. These approaches utilize polymerase-based strategy(94) or DNA shearing(24, 95) in order to provide a semi-quantitative estimate of the number of cells that contain a given insertion, thus allowing to study the genetic complexity of tumors and to infer the likelihood that an insertion targeting a candidate cancer gene has contributed to the initiation or progression of the cancer.

Since tumors may be composed of many subclones, sequencing depth is a key variable determining the yield of insertion sites. For this reason, Illumina sequencing approaches, where millions of insertion site reads may be analyzed, have recently been applied to integration retrieval(60). While PCR-based methods are a proven tool for insertion site analysis, it is likely that in the near future they will be replaced by whole-genome sequencing, and software tools for isolating insertion sites from these data have already been developed(96).

Statistical analyses for the identification of Common Insertion Sites

The basic principal used for the statistical analysis of insertion sites is to identify regions in the genome that carry more insertions than what would be expected by chance, which are then defined as CIS and likely host candidate cancer genes. Several strategies have been used. These include kernel convolution based methods(97), Poisson distribution statistics(98), and Monte Carlo based methods(48). The overlap between different methods on the same data is in the order of 60-80%(50).

The ideal insertional mutagen would be endowed with a random integration pattern that can broadly cover the whole genome. As each of the currently available insertional mutagens has its own insertion biases, methods for CIS identification have been tweaked to correct for such biases. For instance, in order to account for the preference of SB transposon to integrate in TA dinucleotides, frequency of integration was corrected for the TA content of each chromosome (48). Since LV integrations cluster in large chromosomal areas giving rise to false positive calls in CIS stats, tests to calculate if a gene represents an outlier for integration frequency within a region were applied to screen for CIS that are really due to the process of selection(74). Other factors such as chromatin structure and GC content are also likely to influence the likelihood that an insertional mutagen will insert into the genome; therefore the generation of a “background” unselected dataset to filter for insertion site hotspots is warranted.

LIMITS AND ADVANTAGES OF INSERTIONAL MUTAGENESIS SYSTEMS

Comparing insertional mutagenesis with other cancer gene discovery screenings

Different high throughput experimental approaches have been developed to discover novel cancer genes. The recent advance in sequencing technologies has boosted the analysis of cancer genomes at different levels, including identification of point mutations, chromosomal aberrations and epimutations. However these approaches require hundreds or thousands of tumors to be analyzed to implicate in cell transformation those genes that are less frequently mutated (99, 100), thus making data from functional approaches, such as insertional mutagenesis, a very important complement to other high throughput experimental approaches.

High-throughput screening with cDNA libraries or short hairpin RNA (shRNA) libraries have also been exploited to identify oncogenes and TSGs (101, 102), respectively. Despite the great advances provided by these approaches, there are also some caveats. For example cDNA libraries may not express a gene product at an appropriate level to induce tumorigenesis, while shRNAs are notorious for off-target effects and incomplete silencing. Additionally, they rely on the composition of predefined libraries, while the genome wide distribution of insertional mutagens allows to identify even non-annotated genes by mapping clusters of integrations within a chromosomal region. Therefore insertional mutagenesis represents a powerful approach for functional studies of novel cancer genes.

Pros and Cons of the different insertional mutagens

Despite the large amount of oncogenes (and, less frequently, TSGs) that was identified by insertional mutagenesis studies, the current approaches still suffer from several biases and limitations.

A very important issue that has to be considered when performing screenings for cancer genes by insertional mutagenesis is that each insertional mutagen has its own intrinsic integration biases. Transposons, and especially SB mobilized transposons, are characterized by marked “local hopping”, which is the phenomenon by which transposon integrations frequently occur at loci close to their donor loci (8). As a consequence, genes in the region surrounding donor loci are usually mutated at a very high frequency which potentially results in false-positive CIS calls, i.e. CIS that do not contribute to clonal expansion of the tumor. A simple way to overcome this issue is to exclude from further analysis all the integrations coming from the donor chromosome. Some transposon screens have used two different mouse strains harboring different transposon concatemers, which in part overcomes this issue.

Different insertional mutagens have different insertion site preferences. For example PB preferably integrates within genes, and both PB and SB transposons preferably integrate at actively transcribed loci (62, 103). The different integration site selection of different subtypes of retroviruses has been extensively studied and depends largely on different factors, including the specificity of the integrase enzyme, the sequence of the LTRs, the chromatin status of the cell and several host cell factors. Indeed, MLVs display a strong bias for integration in the neighborhood of the TSS of genes and in CpG islands. Moreover, -retroviruses preferably integrate within transcriptional units and in actively transcribed genes (83-85). Conversely, MMTV displays a more random distribution of integrations throughout the genome (104). LVs are characterized by a strong bias of integration within genes and specifically within actively transcribed genes (83-85). Although some of LV, MLV or PB biases can favor the deregulation of genes and therefore increase the chances of transformation, they could complicate the correct identification of CIS and limit the spectrum of targeted cancer genes. These problems can be overcome only characterizing the integration biases of each specific insertional mutagen in the specific tissue of interest in non-selective conditions. Then, the use of statistical approaches that take into account the specific biases of a vector when calculating the significance of a CIS may be a solution. Importantly, by virtue of their different integration biases, the use of different insertional mutagens to screen for cancer genes in the same tissue may significantly widen the spectrum of candidate cancer genes that can be identified.

The limited tissue tropism and the inability to integrate in non-proliferating cells typical of - and -retroviruses have been overcome by transposable elements. However, the development of specific knock-in mouse strains for the activation of transposon insertional mutagenesis in several different tissues is still labor-intensive and time-consuming. Since it is estimated that about 200 different types of cancer from 60 different organs affect human beings (8), it could be a hard challenge to generate all the models for tissue specific insertional mutagenesis. The availability of more flexible and versatile mutagens could accelerate the development of cancer gene screening in different solid tissues. In this regard, the potential versatility of LVs may be helpful to model insertional mutagenesis in different tissues, exploiting also the wide array of regulatory sequences that have been developed for gene therapy applications.

The continuous excision of transposons during the progression of the tumor may un-tag transformation events that were relevant for the tumor induction. Moreover, SB leaves a footprint that could continue to interfere with the normal function of the targeted gene even

after the transposon has been excised. Therefore, some cancer genes could be underestimated in transposon screening and could be identified only increasing the number of tumors and integrations that are analyzed. Retroviruses are devoid of this problem, but they also have intrinsic limitations. Besides their limited tropism, it was shown that, since retroviruses frequently act as transcriptional enhancers, they often integrate distal to the gene they are affecting, a factor that may hamper the identification of the real cancer gene that is upregulated by integrations within a CIS (105).

Retroviruses and LVs are characterized by a preference to drive the up-regulation of oncogenes, possibly meaning that many TSGs are missed in screens using these mutagens (5, 8). The fact that biallelic inactivation of TSGs is required, which is unlikely to readily occur with insertional mutagens, is also likely to be a contributory factor in the paucity of TSGs discovered from insertional mutagenesis screens. Although haploinsufficient or dominant negative alleles of TSGs may be generated, TSGs are probably still underrepresented. A strategy to partially alleviate this bias exploited a mouse model of Bloom syndrome, characterized by an increased rate of mitotic recombination and tumor predisposition. In this model, the retrovirus integrates into one allele of a TSG resulting in a loss-of-function event that is homozygosed, thus leading to the complete loss of the TSG. Retroviral screens using this approach have been performed and allowed the identification of TSGs, although the majority of the cancer genes identified were still oncogenes (106).

Transposons and retroviruses continuously undergo rounds of integration throughout the life of the host mouse. This feature offers several advantages, such as a high density of integrations throughout the genome and the capacity to mimic the accumulation of different mutations during the progression of tumors. As mentioned above, these continuously occurring integrations open the possibility to study tumor progression and metastasis. Moreover, besides increasing the probability to retrieve CIS genes, the high density of integration within each genome has allowed the identification of co-occurring or mutually exclusive CIS and genetic interactions between genes. However, the continuous acquisition of integrations is likely to significantly increase the number of passenger insertions potentially increasing noise and decreasing the signal to noise ratio of the screen. Statistical methods for CIS identification are helpful to filter out these bystander integration events but ultimately independent biological validation is required.

To overcome this issue, recently a self-inactivating piggyBac transposon has been developed for transient insertional mutagenesis (65, 66). In this system the transposon and the transposase are one next to the other in a configuration in which the PB transposase is under the control of the constitutive CAGGS promoter that is inside the transposon. The PB transposase is fused to a domain of the estrogen receptor that keeps the protein inactive in the cytoplasm until the ligand is supplied. Upon administration of tamoxifen, the PB transposase is translocated to the nucleus where it catalyzes the excision of the transposon and its insertion elsewhere in the genome. Consequently, once the transposon has been excised from its original position next to the transposase, the enzyme is no more transcribed and the transposition is arrested. This system seems promising, but until now it has just been exploited to induce transformation of cell lines.

Remarkably, in this regard the replication deficient design of LVs may allow to temporally restrict the integration events at the moment of vector administration, and therefore could significantly enrich for early events that occur in the process of the neoplastic transformation. LVs represent a promising tool for insertional mutagenesis, but their extensive exploitation in the cancer gene discovery arena has just started.

Nonetheless, LVs may suffer from some potential limitations too. Indeed, since LVs are replication deficient, an extensive transduction of the targeted organ is required to obtain significant levels of mutagenesis and eventually cell transformation. Therefore, LV-mediated insertional mutagenesis may be inefficient in organs that are difficult to access. LV integrations are produced in a short time window after injection and before the *in vivo* selection of transformed clones occurs. This results in a lower total number of integrations than γ -retroviruses and transposons produce, which may reduce the incidence of tumor induction and the total yield of identified cancer genes. On the other hand, this characteristic may also facilitate the identification of genes that are mutated early in carcinogenesis because it eliminates bystander and progression-related integrations.

Conclusions and perspectives

Overall, the different insertional mutagens are characterized by intrinsic peculiarities, which are associated to advantages and disadvantages in their use and make them complementary to each other. Importantly, the recent studies with PB transposons and LVs clearly showed that using novel tools allowed the identification of novel culprits of cell transformation which were elusive in the previous screenings. Their extensive application in different mouse models may further increase our knowledge on the molecular culprits of cell transformation. Moreover, γ -retroviral vectors have been recently developed for gene therapy applications (107, 108). Since they have a different integration profile compared to other integrating vectors, their modification to generate efficient insertional mutagens could provide a new tool for cancer gene discovery screens, as we did with LVs.

Remarkably, insertional mutagenesis plays also a role in human disease. Recent studies have shown that spontaneous retrotransposon integration has a role in the pathogenesis of different human tumors (109-111). If future studies confirm that it is a common phenomenon in human cancers, mapping retrotransposon integrations may be helpful to identify novel cancer genes, thus providing a kind of spontaneous insertional mutagenesis screening. Insertional mutagenesis has also been demonstrated to occur in some gene therapy trials with RVs (112) triggering great concern in this field. Therefore further studies aimed at unraveling the basic mechanisms of insertional mutagenesis will be instrumental for the development of safer gene therapy vectors.

The continuous development of new insertional mutagenesis tools, together with the improvement of sequencing technologies for the retrieval of integration sites, will continuously boost these forward genetics screenings that promise to significantly improve the knowledge of the molecular pathogenesis of cancer.

Acknowledgments

We would like to thank Pierangela Gallina, Fabrizio Benedicenti, Monica Volpin and Andrea Calabria for the contribution to LV studies and Daniela Cesana for the contribution to LV studies and for the revision of the manuscript. We would like to thank the Telethon Foundation and the Association for International Cancer Research for supporting our research.

LIST OF ABBREVIATIONS

5' / 3' UTR	5' / 3' untranslated region
ALL	Acute lymphoblastic leukemia
AML	Acute myeloid leukemia
CIS	Common insertion site

RV	-retroviral vector
HCC	Hepatocellular carcinoma
HSC	Hematopoietic stem cell
IR/DR	Inverted repeats/direct repeats
MAV-2	Myeloblastosis-associated virus type 2
MMTV	Mouse mammary tumor virus
MLV	murine leukemia virus
MoMLV	Moloney murine leukemia virus
MSCV	Murine stem cell virus
LTR	Long terminal repeats
LV	Lentiviral vector
ORF	Open reading frame
PB	PiggyBac
polyA	polyadenylation signal
RE	restriction enzyme
RV	Retrovirus
SA	Splice acceptor
SB	Sleeping Beauty
SD	Splice donor
SIN	Self-inactivating
TSG	Tumor suppressor gene
TSS	Transcription start site
VSV-G	G glycoprotein of the vesicular stomatitis virus

REFERENCES

1. Selten G, Cuypers HT, Zijlstra M, Melief C, Berns A. Involvement of c-myc in MuLV-induced T cell lymphomas in mice: frequency and mechanisms of activation. *The EMBO journal*. 1984; 3:3215–22. [PubMed: 6098468]
2. Uren AG, Kool J, Berns A, van Lohuizen M. Retroviral insertional mutagenesis: past, present and future. *Oncogene*. 2005; 24:7656–72. [PubMed: 16299527]
3. Morishita K, Parker DS, Mucenski ML, Jenkins NA, Copeland NG, Ihle JN. Retroviral activation of a novel gene encoding a zinc finger protein in IL-3-dependent myeloid leukemia cell lines. *Cell*. 1988; 54:831–40. [PubMed: 2842066]
4. van Lohuizen M, Breuer M, Berns A. N-myc is frequently activated by proviral insertion in MuLV-induced T cell lymphomas. *The EMBO journal*. 1989; 8:133–6. [PubMed: 2653809]
5. Kool J, Berns A. High-throughput insertional mutagenesis screens in mice to identify oncogenic networks. *Nature reviews Cancer*. 2009; 9:389–99.
6. Selten G, Cuypers HT, Berns A. Proviral activation of the putative oncogene Pim-1 in MuLV induced T-cell lymphomas. *The EMBO journal*. 1985; 4:1793–8. [PubMed: 2992942]
7. Ceci JD, Patriotis CP, Tsatsanis C, Makris AM, Kovatch R, Swing DA, et al. Tpl-2 is an oncogenic kinase that is activated by carboxy-terminal truncation. *Genes & development*. 1997; 11:688–700. [PubMed: 9087424]

8. Copeland NG, Jenkins NA. Harnessing transposons for cancer gene discovery. *Nature reviews Cancer*. 2010; 10:696–706.
9. Coffin, JM.; Hughes, SH.; Varmus, H. *Retroviruses*. Cold Spring Harbor Laboratory Press; Plainview, N.Y.: 1997.
10. Fan H. Leukemogenesis by Moloney murine leukemia virus: a multistep process. *Trends Microbiol*. 1997; 5:74–82. [PubMed: 9108934]
11. Starkey CR, Lobelle-Rich PA, Granger SW, Brightman BK, Fan H, Levy LS. Tumorigenic potential of a recombinant retrovirus containing sequences from Moloney murine leukemia virus and feline leukemia virus. *J Virol*. 1998; 72:1078–84. [PubMed: 9445002]
12. Johnson C, Lobelle-Rich PA, Puetter A, Levy LS. Substitution of feline leukemia virus long terminal repeat sequences into murine leukemia virus alters the pattern of insertional activation and identifies new common insertion sites. *J Virol*. 2005; 79:57–66. [PubMed: 15596801]
13. Sorensen KD, Quintanilla-Martinez L, Kunder S, Schmidt J, Pedersen FS. Mutation of all Runx (AML1/core) sites in the enhancer of T-lymphomagenic SL3-3 murine leukemia virus unmasks a significant potential for myeloid leukemia induction and favors enhancer evolution toward induction of other disease patterns. *J Virol*. 2004; 78:13216–31. [PubMed: 15542674]
14. Touw IP, Erkeland SJ. Retroviral insertion mutagenesis in mice as a comparative oncogenomics tool to identify disease genes in human leukemia. *Mol Ther*. 2007; 15:13–9. [PubMed: 17164770]
15. van Lohuizen M, Verbeek S, Scheijen B, Wientjens E, van der Gulden H, Berns A. Identification of cooperating oncogenes in E mu-myc transgenic mice by provirus tagging. *Cell*. 1991; 65:737–52. [PubMed: 1904008]
16. Shaughnessy JD Jr, Largaespada DA, Tian E, Fletcher CF, Cho BC, Vyas P, et al. Mrvi1, a common MRV integration site in BXH2 myeloid leukemias, encodes a protein with homology to a lymphoid-restricted membrane protein Jaw1. *Oncogene*. 1999; 18:2069–84. [PubMed: 10321731]
17. Gilbert DJ, Neumann PE, Taylor BA, Jenkins NA, Copeland NG. Susceptibility of AKXD recombinant inbred mouse strains to lymphomas. *J Virol*. 1993; 67:2083–90. [PubMed: 8383230]
18. Li J, Shen H, Himmel KL, Dupuy AJ, Largaespada DA, Nakamura T, et al. Leukaemia disease genes: large-scale cloning and pathway predictions. *Nat Genet*. 1999; 23:348–53. [PubMed: 10610183]
19. Suzuki T, Shen H, Akagi K, Morse HC, Malley JD, Naiman DQ, et al. New genes involved in cancer identified by retroviral tagging. *Nat Genet*. 2002; 32:166–74. [PubMed: 12185365]
20. Nusse R, Varmus HE. Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. *Cell*. 1982; 31:99–109. [PubMed: 6297757]
21. Hilkens J. Recent translational research: oncogene discovery by insertional mutagenesis gets a new boost. *Breast Cancer Res*. 2006; 8:102. [PubMed: 16469121]
22. Theodorou V, Kimm MA, Boer M, Wessels L, Theelen W, Jonkers J, et al. MMTV insertional mutagenesis identifies genes, gene families and pathways involved in mammary cancer. *Nat Genet*. 2007; 39:759–69. [PubMed: 17468756]
23. Klijn C, Koudijs MJ, Kool J, Ten Hoeve J, Boer M, de Moes J, et al. Analysis of Tumor Heterogeneity and Cancer Gene Networks Using Deep Sequencing of MMTV-Induced Mouse Mammary Tumors. *PLoS One*. 2013; 8:e62113. [PubMed: 23690930]
24. Koudijs MJ, Klijn C, van der Weyden L, Kool J, ten Hoeve J, Sie D, et al. High-throughput semiquantitative analysis of insertional mutations in heterogeneous tumors. *Genome Res*. 2011; 21:2181–9. [PubMed: 21852388]
25. Pecenka V, Dvorak M, Karafiat V, Sloncova E, Hlozaneck I, Travnicek M, et al. Avian nephroblastomas induced by a retrovirus (MAV-2) lacking oncogene. II. Search for common sites of proviral integration in tumour DNA. *Folia Biol (Praha)*. 1988; 34:147–69. [PubMed: 2849568]
26. Pajer P, Pecenka V, Kralova J, Karafiat V, Prukova D, Zemanova Z, et al. Identification of potential human oncogenes by mapping the common viral integration sites in avian nephroblastoma. *Cancer Res*. 2006; 66:78–86. [PubMed: 16397219]
27. Pajer P, Pecenka V, Karafiat V, Kralova J, Horejsi Z, Dvorak M. The twist gene is a common target of retroviral integration and transcriptional deregulation in experimental nephroblastoma. *Oncogene*. 2003; 22:665–73. [PubMed: 12569359]

28. Uhrbom L, Hesselager G, Nister M, Westermark B. Induction of brain tumors in mice using a recombinant platelet-derived growth factor B-chain retrovirus. *Cancer Res.* 1998; 58:5275–9. [PubMed: 9850047]
29. Johansson FK, Brodd J, Eklof C, Ferletta M, Hesselager G, Tiger CF, et al. Identification of candidate cancer-causing genes in mouse brain tumors by retroviral tagging. *Proc Natl Acad Sci U S A.* 2004; 101:11334–7. [PubMed: 15273287]
30. Johansson FK, Goransson H, Westermark B. Expression analysis of genes involved in brain tumor progression driven by retroviral insertional mutagenesis in mice. *Oncogene.* 2005; 24:3896–905. [PubMed: 15750623]
31. An W, Han JS, Wheelan SJ, Davis ES, Coombes CE, Ye P, et al. Active retrotransposition by a synthetic L1 element in mice. *Proc Natl Acad Sci U S A.* 2006; 103:18662–7. [PubMed: 17124176]
32. Ostertag EM, DeBerardinis RJ, Goodier JL, Zhang Y, Yang N, Gerton GL, et al. A mouse model of human L1 retrotransposition. *Nat Genet.* 2002; 32:655–60. [PubMed: 12415270]
33. Babushok DV, Ostertag EM, Courtney CE, Choi JM, Kazazian HH Jr. L1 integration in a transgenic mouse model. *Genome Res.* 2006; 16:240–50. [PubMed: 16365384]
34. Gilbert N, Lutz-Prigge S, Moran JV. Genomic deletions created upon LINE-1 retrotransposition. *Cell.* 2002; 110:315–25. [PubMed: 12176319]
35. Mattison J, van der Weyden L, Hubbard T, Adams DJ. Cancer gene discovery in mouse and man. *Biochim Biophys Acta.* 2009; 1796:140–61. [PubMed: 19285540]
36. Starr TK, Largaespada DA. Cancer gene discovery using the Sleeping Beauty transposon. *Cell Cycle.* 2005; 4:1744–8. [PubMed: 16294016]
37. Collier LS, Largaespada DA. Hopping around the tumor genome: transposons for cancer gene discovery. *Cancer Res.* 2005; 65:9607–10. [PubMed: 16266976]
38. Mates L, Chuah MK, Belay E, Jerchow B, Manoj N, Acosta-Sanchez A, et al. Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates. *Nat Genet.* 2009; 41:753–61. [PubMed: 19412179]
39. Ivics Z, Hackett PB, Plasterk RH, Izsvak Z. Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell.* 1997; 91:501–10. [PubMed: 9390559]
40. Izsvak Z, Ivics Z, Plasterk RH. Sleeping Beauty, a wide host-range transposon vector for genetic transformation in vertebrates. *Journal of molecular biology.* 2000; 302:93–102. [PubMed: 10964563]
41. Geurts AM, Yang Y, Clark KJ, Liu G, Cui Z, Dupuy AJ, et al. Gene transfer into genomes of human cells by the sleeping beauty transposon system. *Mol Ther.* 2003; 8:108–17. [PubMed: 12842434]
42. Zayed H, Izsvak Z, Walisko O, Ivics Z. Development of hyperactive sleeping beauty transposon vectors by mutational analysis. *Mol Ther.* 2004; 9:292–304. [PubMed: 14759813]
43. Dupuy AJ, Akagi K, Largaespada DA, Copeland NG, Jenkins NA. Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. *Nature.* 2005; 436:221–6. [PubMed: 16015321]
44. Collier LS, Carlson CM, Ravimohan S, Dupuy AJ, Largaespada DA. Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. *Nature.* 2005; 436:272–6. [PubMed: 16015333]
45. Kamijo T, Bodner S, van de Kamp E, Randle DH, Sherr CJ. Tumor spectrum in ARF-deficient mice. *Cancer Res.* 1999; 59:2217–22. [PubMed: 10232611]
46. Dupuy AJ, Rogers LM, Kim J, Nannapaneni K, Starr TK, Liu P, 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; 69:8150–6. [PubMed: 19808965]
47. Starr TK, Allaei R, Silverstein KA, Staggs RA, Sarver AL, Bergemann TL, et al. A Transposon-Based Genetic Screen in Mice Identifies Genes Altered in Colorectal Cancer. *Science.* 2009
48. Keng VW, Villanueva A, Chiang DY, Dupuy AJ, Ryan BJ, Matise I, et al. A conditional transposon-based insertional mutagenesis screen for genes associated with mouse hepatocellular carcinoma. *Nat Biotechnol.* 2009; 27:264–74. [PubMed: 19234449]

49. Starr TK, Scott PM, Marsh BM, Zhao L, Than BL, O'Sullivan MG, et al. A Sleeping Beauty transposon-mediated screen identifies murine susceptibility genes for adenomatous polyposis coli (Apc)-dependent intestinal tumorigenesis. *Proc Natl Acad Sci U S A*. 2011; 108:5765–70. [PubMed: 21436051]
50. March HN, Rust AG, Wright NA, ten Hoeve J, de Ridder J, Eldridge M, et al. Insertional mutagenesis identifies multiple networks of cooperating genes driving intestinal tumorigenesis. *Nat Genet*. 2011; 43:1202–9. [PubMed: 22057237]
51. Berquam-Vrieze KE, Nannapaneni K, Brett BT, Holmfeldt L, Ma J, Zagorodna O, et al. Cell of origin strongly influences genetic selection in a mouse model of T-ALL. *Blood*. 2011; 118:4646–56. [PubMed: 21828136]
52. Karreth FA, Tay Y, Perna D, Ala U, Tan SM, Rust AG, et al. In vivo identification of tumor-suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. *Cell*. 2011; 147:382–95. [PubMed: 22000016]
53. Vassiliou GS, Cooper JL, Rad R, Li J, Rice S, Uren A, et al. Mutant nucleophosmin and cooperating pathways drive leukemia initiation and progression in mice. *Nat Genet*. 2011; 43:470–5. [PubMed: 21441929]
54. van der Weyden L, Giotopoulos G, Rust AG, Matheson LS, van Delft FW, Kong J, et al. Modeling the evolution of ETV6-RUNX1-induced B-cell precursor acute lymphoblastic leukemia in mice. *Blood*. 2011; 118:1041–51. [PubMed: 21628403]
55. Mann KM, Ward JM, Yew CC, Kovoichich A, Dawson DW, Black MA, et al. Sleeping Beauty mutagenesis reveals cooperating mutations and pathways in pancreatic adenocarcinoma. *Proc Natl Acad Sci U S A*. 2012; 109:5934–41. [PubMed: 22421440]
56. Perez-Mancera PA, Rust AG, van der Weyden L, Kristiansen G, Li A, Sarver AL, et al. The deubiquitinase USP9X suppresses pancreatic ductal adenocarcinoma. *Nature*. 2012; 486:266–70. [PubMed: 22699621]
57. Wu X, Northcott PA, Dubuc A, Dupuy AJ, Shih DJ, Witt H, et al. Clonal selection drives genetic divergence of metastatic medulloblastoma. *Nature*. 2012; 482:529–33. [PubMed: 22343890]
58. Koso H, Takeda H, Yew CC, Ward JM, Nariai N, Ueno K, et al. Transposon mutagenesis identifies genes that transform neural stem cells into glioma-initiating cells. *Proc Natl Acad Sci U S A*. 2012; 109:E2998–3007. [PubMed: 23045694]
59. Mumert M, Dubuc A, Wu X, Northcott PA, Chin SS, Pedone CA, et al. Functional genomics identifies drivers of medulloblastoma dissemination. *Cancer Res*. 2012; 72:4944–53. [PubMed: 22875024]
60. Rahrman EP, Watson AL, Keng VW, Choi K, Moriarity BS, Beckmann DA, et al. Forward genetic screen for malignant peripheral nerve sheath tumor formation identifies new genes and pathways driving tumorigenesis. *Nat Genet*. 2013
61. Ding S, Wu X, Li G, Han M, Zhuang Y, Xu T. Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice. *Cell*. 2005; 122:473–83. [PubMed: 16096065]
62. Li MA, Pettitt SJ, Eckert S, Ning Z, Rice S, Cadinanos J, et al. The piggyBac Transposon Displays Local and Distant Reintegration Preferences and Can Cause Mutations at Noncanonical Integration Sites. *Mol Cell Biol*. 2013; 33:1317–30. [PubMed: 23358416]
63. Rad R, Rad L, Wang W, Cadinanos J, Vassiliou G, Rice S, et al. PiggyBac transposon mutagenesis: a tool for cancer gene discovery in mice. *Science*. 2010; 330:1104–7. [PubMed: 20947725]
64. Landrette SF, Cornett JC, Ni TK, Bosenberg MW, Xu T. piggyBac transposon somatic mutagenesis with an activated reporter and tracker (PB-SMART) for genetic screens in mice. *PLoS One*. 2011; 6:e26650. [PubMed: 22039523]
65. Cadinanos J, Bradley A. Generation of an inducible and optimized piggyBac transposon system. *Nucleic Acids Res*. 2007; 35:e87. [PubMed: 17576687]
66. Kong J, Wang F, Brenton JD, Adams DJ. Slingshot: a PiggyBac based transposon system for tamoxifen-inducible 'self-inactivating' insertional mutagenesis. *Nucleic Acids Res*. 2010; 38:e173. [PubMed: 20688953]
67. Trobridge GD, Russell DW. Helper-free foamy virus vectors. *Hum Gene Ther*. 1998; 9:2517–25. [PubMed: 9853518]

68. Follenzi A, Ailles LE, Bakovic S, Geuna M, Naldini L. Gene transfer by lentiviral vectors is limited by nuclear translocation and rescued by HIV-1 pol sequences. *Nat Genet.* 2000; 25:217–22. [PubMed: 10835641]
69. Vigna E, Naldini L. Lentiviral vectors: excellent tools for experimental gene transfer and promising candidates for gene therapy. *The journal of gene medicine.* 2000; 2:308–16. [PubMed: 11045424]
70. Naldini L, Blomer U, Gally P, Ory D, Mulligan R, Gage FH, et al. In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science.* 1996; 272:263–7. [PubMed: 8602510]
71. Zufferey R, Dull T, Mandel RJ, Bukovsky A, Quiroz D, Naldini L, et al. Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery. *J Virol.* 1998; 72:9873–80. [PubMed: 9811723]
72. Montini E, Cesana D, Schmidt M, Sanvito F, Ponzoni M, Bartholomae C, et al. Hematopoietic stem cell gene transfer in a tumor-prone mouse model uncovers low genotoxicity of lentiviral vector integration. *Nat Biotechnol.* 2006; 24:687–96. [PubMed: 16732270]
73. Montini E, Cesana D, Schmidt M, Sanvito F, Bartholomae CC, Ranzani M, et al. The genotoxic potential of retroviral vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy. *J Clin Invest.* 2009; 119:964–75. [PubMed: 19307726]
74. Biffi A, Bartholomae CC, Cesana D, Cartier N, Aubourg P, Ranzani M, et al. Lentiviral vector common integration sites in preclinical models and a clinical trial reflect a benign integration bias and not oncogenic selection. *Blood.* 2011; 117:5332–9. [PubMed: 21403130]
75. Scaramuzza S, Biasco L, Ripamonti A, Castiello MC, Loperfido M, Draghici E, et al. Preclinical safety and efficacy of human CD34(+) cells transduced with lentiviral vector for the treatment of Wiskott-Aldrich syndrome. *Mol Ther.* 2013; 21:175–84. [PubMed: 22371846]
76. Galy A, Roncarolo MG, Thrasher AJ. Development of lentiviral gene therapy for Wiskott Aldrich syndrome. *Expert Opin Biol Ther.* 2008; 8:181–90. [PubMed: 18194074]
77. Naldini L. Ex vivo gene transfer and correction for cell-based therapies. *Nat Rev Genet.* 2011; 12:301–15. [PubMed: 21445084]
78. Modlich U, Navarro S, Zychlinski D, Maetzig T, Knoess S, Brugman MH, et al. Insertional transformation of hematopoietic cells by self-inactivating lentiviral and gammaretroviral vectors. *Mol Ther.* 2009; 17:1919–28. [PubMed: 19672245]
79. Schambach A, Zychlinski D, Ehrnstroem B, Baum C. Biosafety features of lentiviral vectors. *Hum Gene Ther.* 2013; 24:132–42. [PubMed: 23311447]
80. Biffi A, Montini E, Lorioli L, Cesani M, Fumagalli F, Plati T, et al. Lentiviral Hematopoietic Stem Cell Gene Therapy Benefits Metachromatic Leukodystrophy. *Science.* 2013
81. Aiuti A, Biasco L, Scaramuzza S, Ferrua F, Cicalese MP, Baricordi C, et al. Lentiviral Hematopoietic Stem Cell Gene Therapy in Patients with Wiskott-Aldrich Syndrome. *Science.* 2013
82. Heckl D, Schwarzer A, Haemmerle R, Steinemann D, Rudolph C, Skawran B, et al. Lentiviral Vector Induced Insertional Haploinsufficiency of Ebf1 Causes Murine Leukemia. *Mol Ther.* 2012
83. Mitchell RS, Beitzel BF, Schroder AR, Shinn P, Chen H, Berry CC, et al. Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol.* 2004; 2:E234. [PubMed: 15314653]
84. Wu X, Li Y, Crise B, Burgess SM. Transcription start regions in the human genome are favored targets for MLV integration. *Science.* 2003; 300:1749–51. [PubMed: 12805549]
85. Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, Bushman F. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell.* 2002; 110:521–9. [PubMed: 12202041]
86. Cesana D, Sgualdino J, Rudilosso L, Merella S, Naldini L, Montini E. Whole transcriptome characterization of aberrant splicing events induced by lentiviral vector integrations. *J Clin Invest.* 2012; 122:1667–76. [PubMed: 22523064]
87. Moiani A, Paleari Y, Sartori D, Mezzadra R, Miccio A, Cattoglio C, et al. Lentiviral vector integration in the human genome induces alternative splicing and generates aberrant transcripts. *J Clin Invest.* 2012; 122:1653–66. [PubMed: 22523069]

88. Ranzani M, Cesana D, Bartholomae CC, Sanvito F, Pala M, Benedicenti F, et al. Lentiviral vector-based insertional mutagenesis identifies genes associated with liver cancer. *Nat Methods*. 2013; 10:155–61. [PubMed: 23314173]
89. Vigna E, Amendola M, Benedicenti F, Simmons AD, Follenzi A, Naldini L. Efficient Tet-dependent expression of human factor IX in vivo by a new self-regulating lentiviral vector. *Mol Ther*. 2005; 11:763–75. [PubMed: 15851015]
90. Devon RS, Porteous DJ, Brookes AJ. Splinkerettes--improved vectorettes for greater efficiency in PCR walking. *Nucleic Acids Res*. 1995; 23:1644–5. [PubMed: 7784225]
91. Schmidt M, Schwarzwaelder K, Bartholomae C, Zaoui K, Ball C, Pilz I, et al. High-resolution insertion-site analysis by linear amplification-mediated PCR (LAM-PCR). *Nature methods*. 2007; 4:1051–7. [PubMed: 18049469]
92. Uren AG, Kool J, Matentzoglou K, de Ridder J, Mattison J, van Uiter M, et al. Large-scale mutagenesis in p19(ARF)- and p53-deficient mice identifies cancer genes and their collaborative networks. *Cell*. 2008; 133:727–41. [PubMed: 18485879]
93. Uren AG, Mikkers H, Kool J, van der Weyden L, Lund AH, Wilson CH, et al. A high-throughput splinkerette-PCR method for the isolation and sequencing of retroviral insertion sites. *Nat Protoc*. 2009; 4:789–98. [PubMed: 19528954]
94. Paruzynski A, Arens A, Gabriel R, Bartholomae CC, Scholz S, Wang W, et al. Genome-wide high-throughput integrome analyses by nrLAM-PCR and next-generation sequencing. *Nat Protoc*. 2010; 5:1379–95. [PubMed: 20671722]
95. Berry CC, Gillet NA, Melamed A, Gormley N, Bangham CR, Bushman FD. Estimating abundances of retroviral insertion sites from DNA fragment length data. *Bioinformatics*. 2012; 28:755–62. [PubMed: 22238265]
96. Brett BT, Berquam-Vrieze KE, Nannapaneni K, Huang J, Scheetz TE, Dupuy AJ. Novel molecular and computational methods improve the accuracy of insertion site analysis in Sleeping Beauty-induced tumors. *PLoS One*. 2011; 6:e24668. [PubMed: 21931803]
97. de Ridder J, Uren A, Kool J, Reinders M, Wessels L. Detecting statistically significant common insertion sites in retroviral insertional mutagenesis screens. *PLoS Comput Biol*. 2006; 2:e166. [PubMed: 17154714]
98. Sarver AL, Erdman J, Starr T, Largaespada DA, Silverstein KA. TAPDANCE: an automated tool to identify and annotate transposon insertion CISs and associations between CISs from next generation sequence data. *BMC Bioinformatics*. 2012; 13:154. [PubMed: 22748055]
99. Garraway LA, Lander ES. Lessons from the cancer genome. *Cell*. 2013; 153:17–37. [PubMed: 23540688]
100. Mwenifumbo JC, Marra MA. Cancer genome-sequencing study design. *Nat Rev Genet*. 2013; 14:321–32. [PubMed: 23594910]
101. Bernards R, Brummelkamp TR, Beijersbergen RL. shRNA libraries and their use in cancer genetics. *Nature methods*. 2006; 3:701–6. [PubMed: 16929315]
102. Mullenders J, Bernards R. Loss-of-function genetic screens as a tool to improve the diagnosis and treatment of cancer. *Oncogene*. 2009; 28:4409–20. [PubMed: 19767776]
103. Liang Q, Kong J, Stalker J, Bradley A. Chromosomal mobilization and reintegration of Sleeping Beauty and PiggyBac transposons. *Genesis*. 2009; 47:404–8. [PubMed: 19391106]
104. Faschinger A, Rouault F, Sollner J, Lukas A, Salmons B, Gunzburg WH, et al. Mouse mammary tumor virus integration site selection in human and mouse genomes. *J Virol*. 2008; 82:1360–7. [PubMed: 18032509]
105. Sauvageau M, Miller M, Lemieux S, Lessard J, Hebert J, Sauvageau G. Quantitative expression profiling guided by common retroviral insertion sites reveals novel and cell type specific cancer genes in leukemia. *Blood*. 2008; 111:790–9. [PubMed: 17906077]
106. Suzuki T, Minehata K, Akagi K, Jenkins NA, Copeland NG. Tumor suppressor gene identification using retroviral insertional mutagenesis in Blm-deficient mice. *The EMBO journal*. 2006; 25:3422–31. [PubMed: 16858412]
107. Suerth JD, Maetzig T, Brugman MH, Heinz N, Appelt JU, Kaufmann KB, et al. Alpharetroviral self-inactivating vectors: long-term transgene expression in murine hematopoietic cells and low genotoxicity. *Mol Ther*. 2012; 20:1022–32. [PubMed: 22334016]

108. Suerth JD, Maetzig T, Galla M, Baum C, Schambach A. Self-inactivating alpharetroviral vectors with a split-packaging design. *J Virol*. 2010; 84:6626–35. [PubMed: 20410274]
109. Lee E, Iskow R, Yang L, Gokcumen O, Haseley P, Luquette LJ 3rd, et al. Landscape of somatic retrotransposition in human cancers. *Science*. 2012; 337:967–71. [PubMed: 22745252]
110. Shukla R, Upton KR, Munoz-Lopez M, Gerhardt DJ, Fisher ME, Nguyen T, et al. Endogenous retrotransposition activates oncogenic pathways in hepatocellular carcinoma. *Cell*. 2013; 153:101–11. [PubMed: 23540693]
111. Solyom S, Ewing AD, Rahrman EP, Doucet T, Nelson HH, Burns MB, et al. Extensive somatic L1 retrotransposition in colorectal tumors. *Genome Res*. 2012; 22:2328–38. [PubMed: 22968929]
112. Hacin-Bey-Abina S, Garrigue A, Wang GP, Soulier J, Lim A, Morillon E, et al. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *The Journal of clinical investigation*. 2008; 118:3132–42. [PubMed: 18688285]
113. Baum C, Schambach A, Bohne J, Galla M. Retrovirus vectors: toward the plentivirus? *Mol Ther*. 2006; 13:1050–63. [PubMed: 16632409]

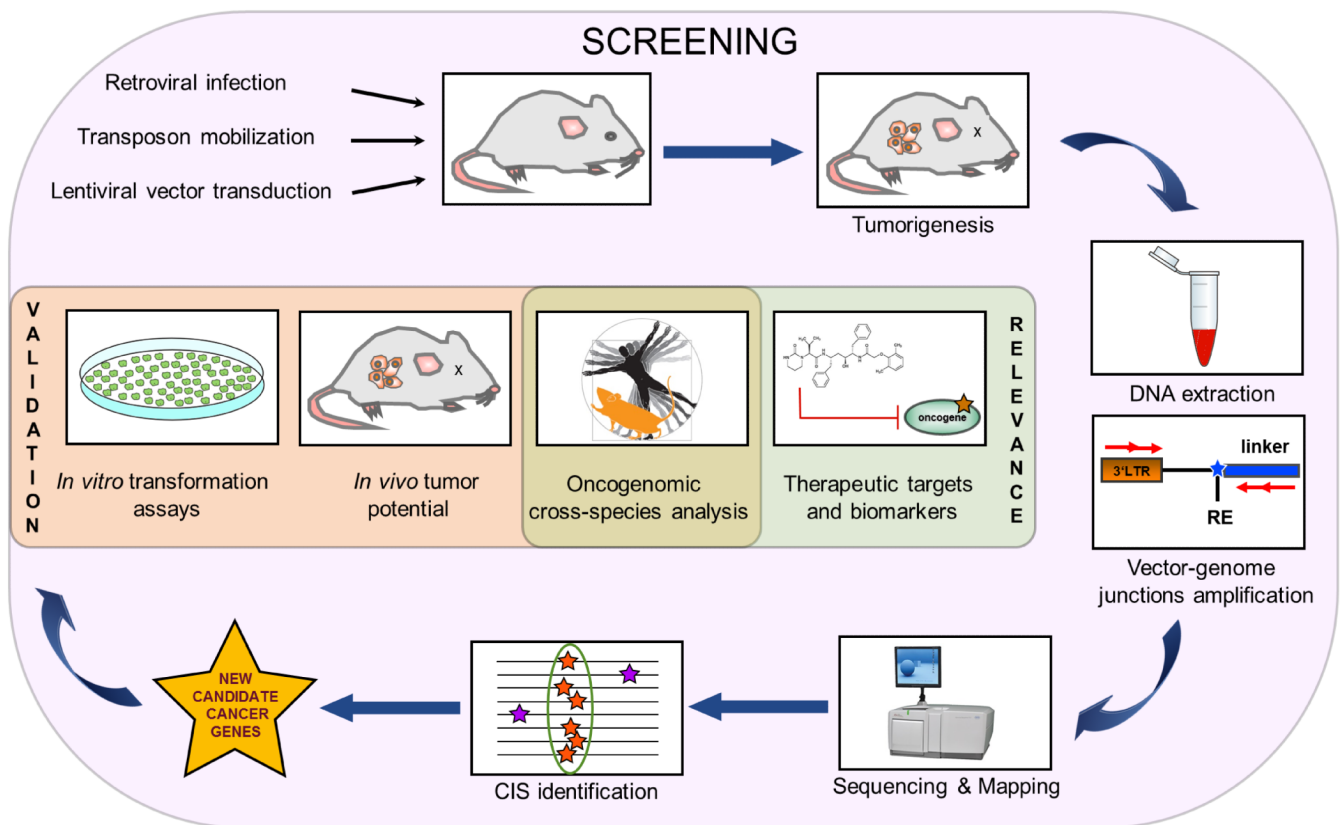


Figure 1. General outline of the work-flow for cancer gene discovery using insertional mutagens
 From top-left, clockwise: the insertional mutagens are administered or activated in the proper mouse model to induce tumorigenesis. Integrations are retrieved from tumor DNA by a PCR-based technique and sequenced. Bioinformatic and statistical analysis leads to map vector integrations events. Clusters of these events are called common insertion sites (CIS). The candidate cancer genes hosted within CIS are then validated by *in vitro* and/or *in vivo* experiments for their transformation potential. The orthologs are then interrogated for their expression level and mutation status in human tumors, allowing the identification of candidate therapeutic targets for the human disease. RE: restriction enzyme.

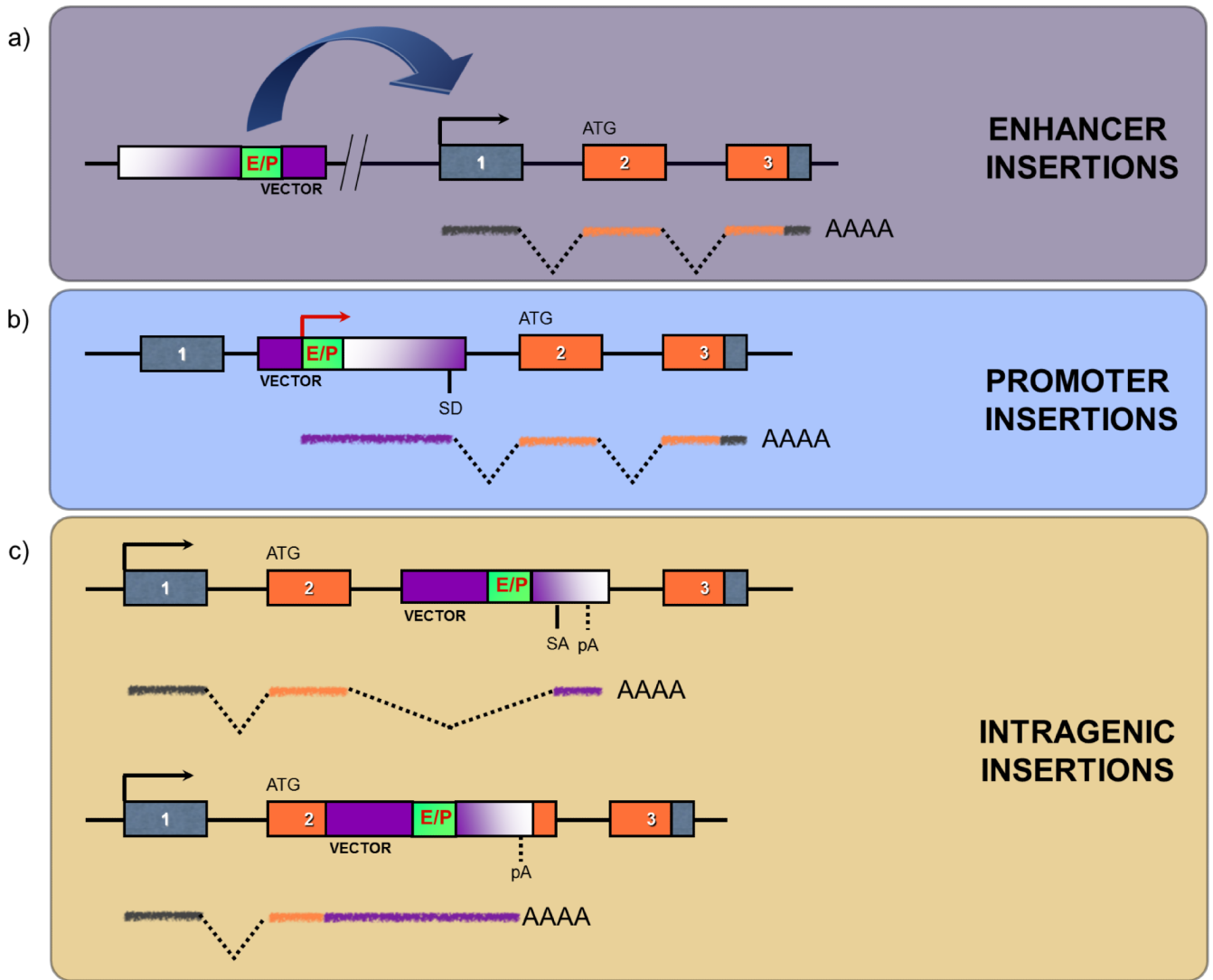


Figure 2. Different mechanisms of gene deregulation by insertional mutagenesis
 Vector integration can induce insertional mutagenesis by different mechanisms: **A)** enhancer insertions occur when the enhancer elements of the integrating vector enhance the transcription of a cellular gene under its own promoter. The integrated vector might be some distance from the cellular gene, and usually is found upstream of the gene in antisense orientation. **B)** Promoter insertions occur when the mutagen integrates in the sense orientation of a gene, and uncouples the gene transcription from its own promoter by placing it under the control of the vector one. A chimeric transcript encoding the full-length protein can be thus generated. **C)** Landing of a vector within the transcriptional unit of a gene (intragenic insertions) can induce the generation of a truncated transcript via the internal polyA signal (upper panel) or disrupt an exon (lower panel). If a chimeric transcript encoding a truncated version of the cellular protein is generated, the new protein might have an abnormal biological activity (e.g. gain-of-function of a proto-oncogene). Alternatively, the intragenic vector might disrupt the normal transcription and impede the correct translation of the cellular protein (e.g. loss-of-function of a TSG). In the three boxes, the vector is depicted in purple, with enhancer/promoter (E/P) elements in green. Splice donors (SD), splice acceptors (SA) and polyadenylation sites (pA) are indicated. A prototypical

cellular gene is shown as a sequence of three boxes corresponding to three exons, with orange boxes indicating protein-coding exons and green segments indicating untranslated regions.

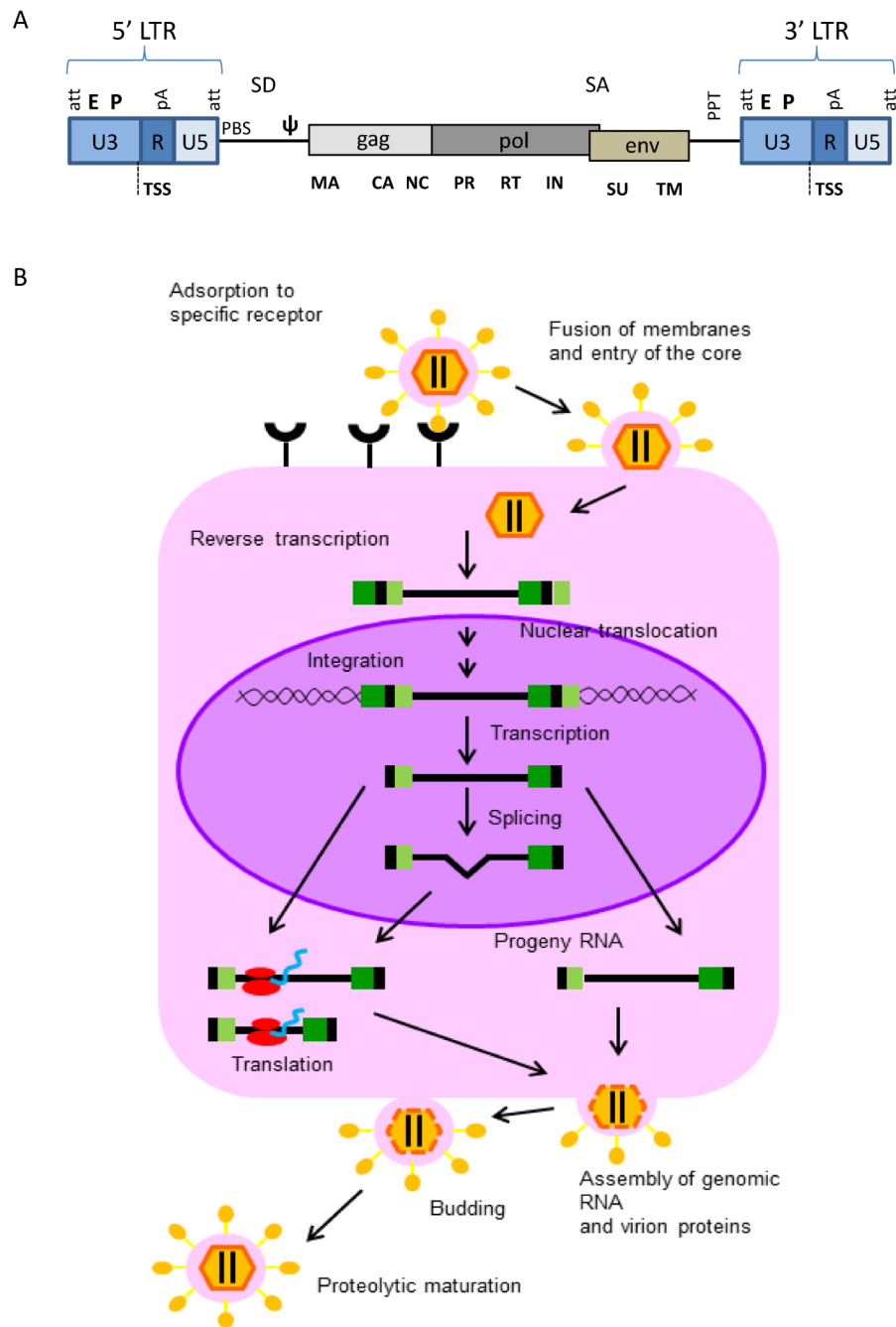


Figure 3. Structure of the genome and replication life cycle of retroviruses

A) Schematic representation of the proviral form of a retrovirus with its simple genomic architecture. The coding sequences of the three essential retroviral genes, gag, pol, and env, and their relative protein subunits are indicated below. EP: enhancer-promoter; att: attachment site; TSS: transcription start site; pA: polyadenylation signal; PBS: primer binding site, necessary for the binding of the primer from which the retrotranscription is triggered; SD: splice donor; Ψ : viral packaging signal; SA: splice acceptor; PPT: polypurine tract; SU: surface; TM: transmembrane; RT: retrotranscriptase; IN: integrase; NC: nucleocapsid; CA: capsid; PR: protease; MA: matrix. Modified from (113). **B)** Retroviral replication cycle. The parental virus attaches to a specific receptor on the surface of a

susceptible cell with the SU portion of the viral Env protein leading to fusion and entry of the core. Reverse transcription then generates a double-stranded DNA copy of the RNA genome. The provirus is transported into the nucleus and integrated into chromosomal DNA. It is then transcribed by cellular RNA polymerase II. Transcription generates RNA copies with the terminal structures organized as in the parental genome. These copies become full-length and spliced messenger RNAs as well as full-length progeny virion RNAs. Viral messages are translated in the cytoplasm. Virion proteins and progeny RNAs assemble at the cell periphery and the plasma membrane, and progeny viruses are released by a process of budding and subsequent maturation into infectious viruses. Modified from (9).

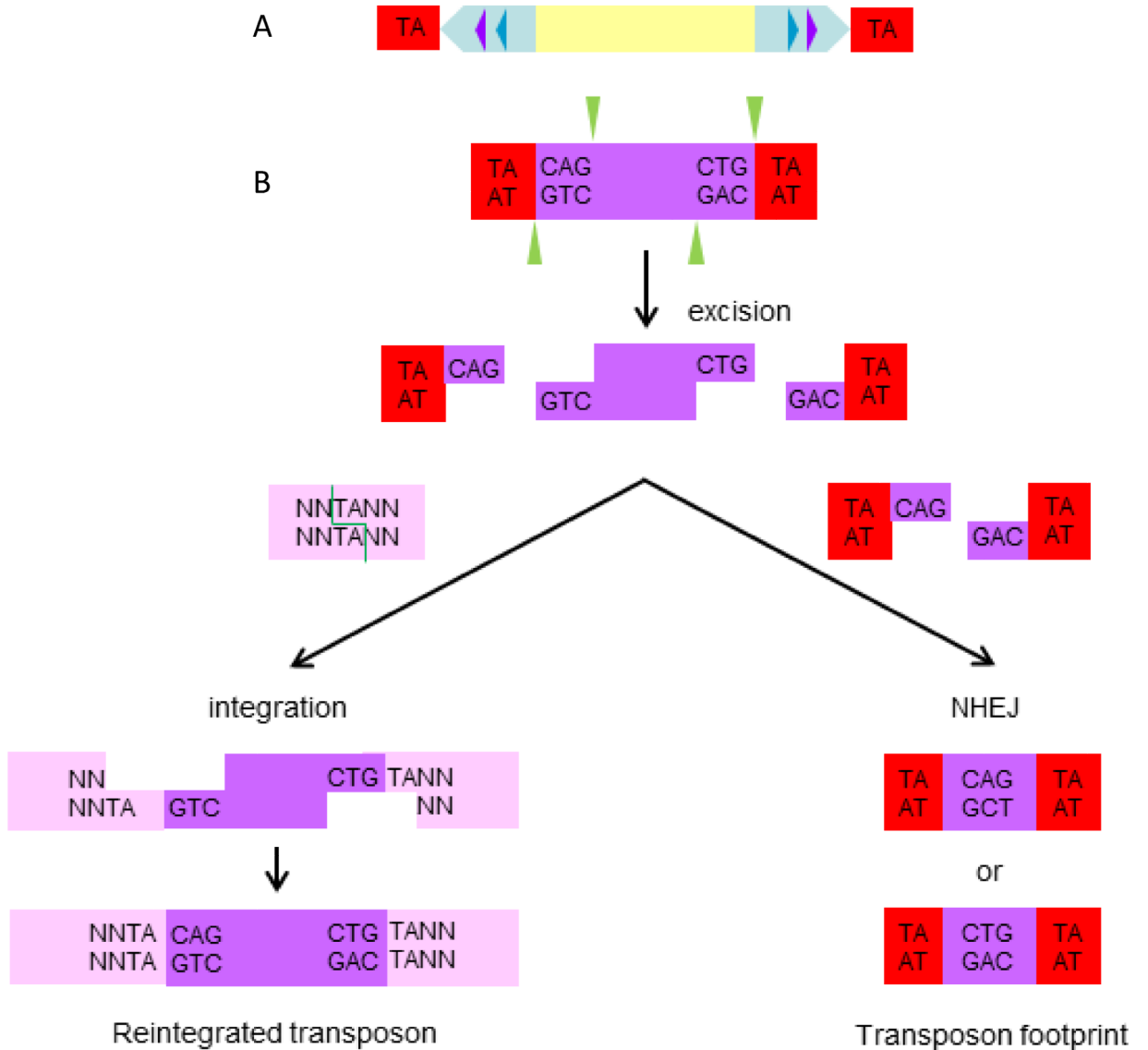


Figure 4. *Sleeping Beauty* structure and mechanism of transposition

(A) The *Sleeping Beauty* (*SB*) inverted terminal repeats (light blue arrows) each contain two direct repeats (DRs) (violet and dark blue triangles) that are the binding sites for the transposase. The genomic TA dinucleotide that is duplicated during integration and flanks the transposon is shown in red. (B) Cut and paste mechanism of *SB* transposition. The transposase (green triangles) initiates transposition by cleaving both ends of the transposon to generate 3 base pair (bp) 5' overhangs (violet) and also cleaves a genomic TA dinucleotide at the integration site (pink) to create a gap with a 3' TA overhang at both ends. The host non-homologous end joining (NHEJ) DNA repair machinery then repairs the single-stranded gaps at the integration site and the double-strand breaks in the donor DNA. A small 5 bp footprint (one TA and 3 bps from the end of the transposon) remains at the excision site.

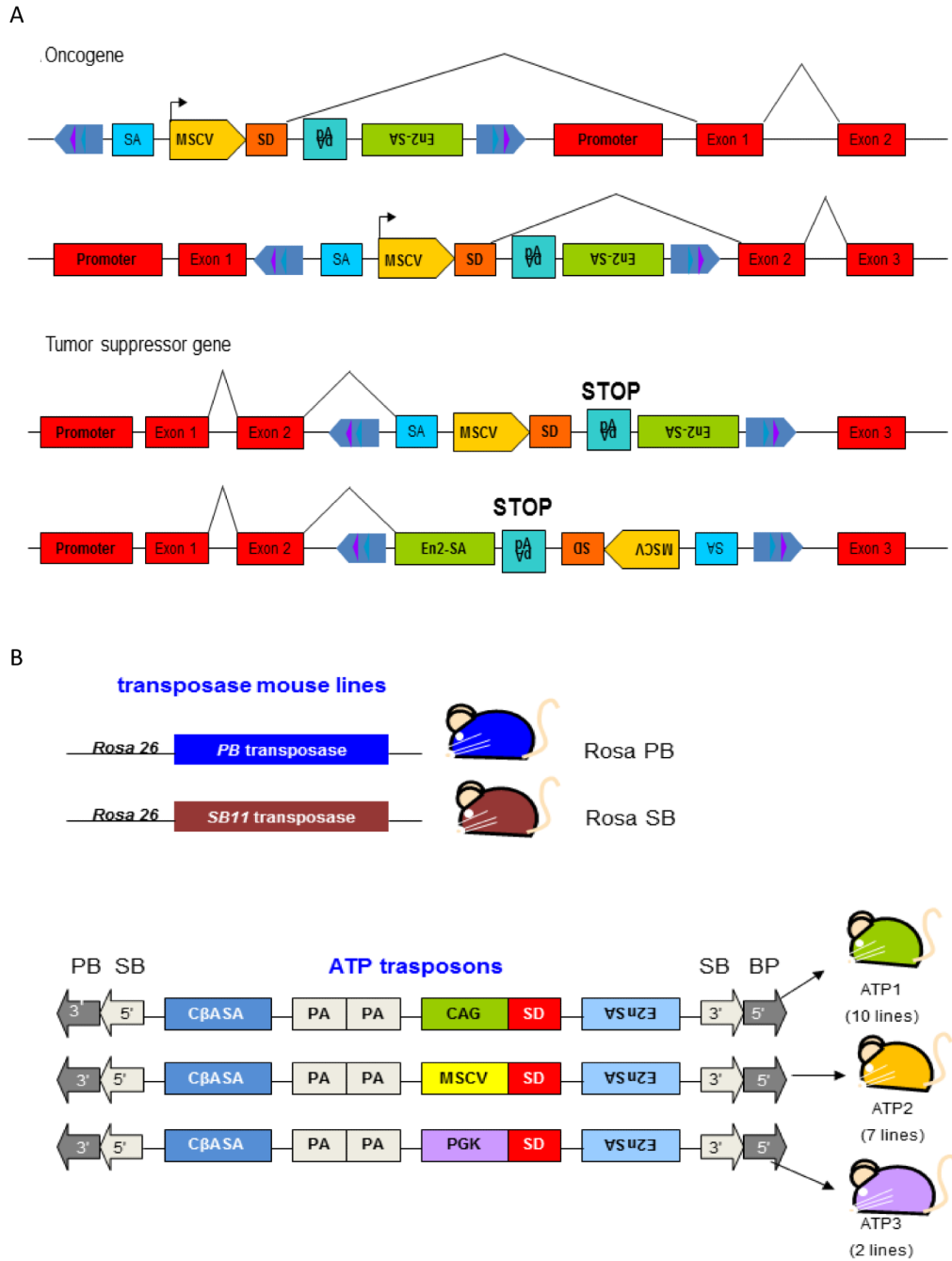
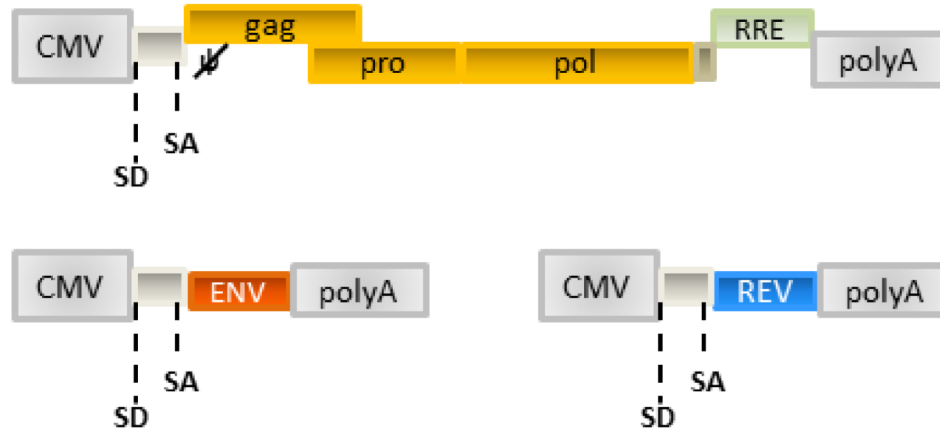


Figure 5. Transposons for insertional mutagenesis

A) *Sleeping Beauty* T2/Onc2 can deregulate the expression of oncogenes or inactivate the expression of tumor suppressor genes. T2/Onc2 contains a murine stem cell virus (MSCV) 5 long terminal repeat (LTR) and a splice donor (SD) site derived from exon 1 of the mouse *Foxf2* gene. T2/Onc2 can thus promote the expression of an oncogene when integrated upstream of or within the gene in the same transcriptional orientation. T2/Onc2 is flanked by optimized *SB* transposase binding sites (violet and dark blue triangles) that are located within the transposon inverted terminal repeats (light blue arrows), which increase the frequency of *SB* transposition (upper panel). T2/Onc2 also contains two splice acceptors and a bi-directional polyA (pA) and can thus prematurely terminate transcription of a tumor

suppressor gene when integrated within the tumor suppressor gene in either orientation (pale blue). One splice acceptor is derived from exon 2 of mouse engrailed 2 (En2-SA) and the other from the Carp b-actin gene (SA) (lower panel). **B**) The PB-based insertional mutagenesis system described in (63). Mouse lines carrying the genetic components of the transposon systems. Upper Panel: RosaPB and RosaSB knock-in mice express PiggyBac or Sleeping Beauty transposase under control of the constitutively active *Rosa26* promoter. Lower panel: Transposon design and transposon mouse lines. Three transposon constructs were designed, which differ in their promoter and enhancer. All three transposons have PB as well as SB IR/DRs and can therefore be mobilized with both transposases. **C** ASA, Carp -actin splice acceptor; En2SA, Engrailed-2 exon-2 splice acceptor; SD, Foxf2 exon-1 splice donor; pA, bidirectional SV40 polyadenylation signal; CAG, cytomegalovirus enhancer and chicken beta-actin promoter; MSCV, murine stem cell virus long terminal repeat; PGK, phosphoglycerate kinase promoter. Modified from (8)

Packaging constructs



Transfer construct

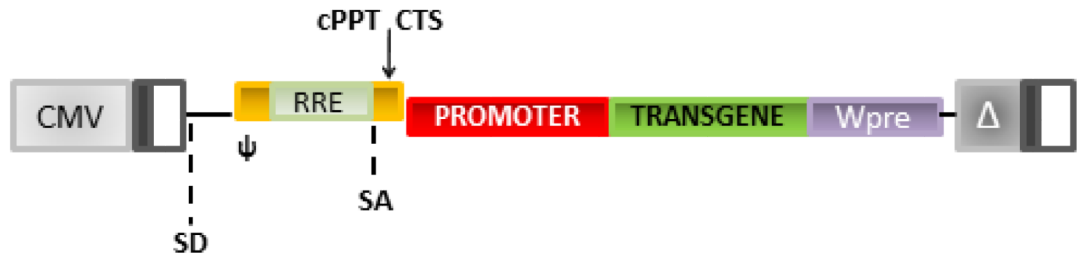


Figure 6. Four plasmids system for the production of late generation LVs

Top panels: LV packaging constructs encoding HIV structural proteins (*gag/pro/pol*), Envelope construct (*Env*) encoding for the envelope glycoprotein and Rev. Lower panel: LV transfer vector; the HIV 5 LTR U3 has been substituted with Cytomegalovirus promoter-enhancer sequence; the 3 LTR U3 sequence containing the HIV promoter-enhancer region has been deleted. Upon retrotranscription the 3 LTR is copied at the 5' obtaining a Self-Inactivating (SIN) vector. Wpre: Woodchuck post-transcriptional regulatory element; CMV: cytomegalovirus enhancer/promoter sequence; SD: splice donor; SA: splice acceptor; Δ : packaging signal.