Cancer Gene Discovery: Exploiting Insertional Mutagenesis

Marco Ranzani1, Stefano Annunziato1, David J. Adams2, and Eugenio Montini1

Abstract

Insertional mutagenesis has been used 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 are integrating viruses with the capacity to induce the deregulation of genes in the neighborhood of the insertion site. Retroviruses have been used 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 retrovirus-based approaches. Most recently, lentiviral vectors have appeared on the scene for use in cancer gene screens. Lentiviral vectors are replication-defective integrating vectors that have the advantage of being able to infect nondividing 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.

Visual Overview: http://mcr.aacrjournals.org/content/11/10/1141/F1.large.jpg.

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 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.

To conduct cancer gene discovery by insertional mutagenesis, the requirements are: (i) the proper oncogenic agent that efficiently integrates in a random or semirandom 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 conduct 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 toward 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 upregulation of endogenous gene expression (Fig. 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 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 (Fig. 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 (Fig. 2C). Many insertional mutagens contain polyadenylation (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' untranslated region (UTR) of a gene may remove mRNA-destabilizing motifs such as AUUUA hairpins or microRNA (miRNA) target sequences, as has been described to occur for \textit{Pim1} and \textit{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 (Fig. 2C). The abnormal biologic 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 (Fig. 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), because 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 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 (ssRNA) of 7 to 11 kilobases (Fig. 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 (LTR), 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 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 ref. 9.

Oncogenic retroviruses have been classified in two groups, acute- and slow-transforming retroviruses. Acute-transforming retroviruses induce, with short latency (2–3 weeks), the
of polyclonal cancers by highly expressing virus-encoded oncogenes, whereas 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 (MMLV, a γ-retrovirus) and mouse mammary tumor virus (MMTV, a β-retrovirus). These viruses cause lymphomas and mammary tumors, respectively, in mice. The rationale of these screens is infecting the newborn mice (postnatal day 1 or 2, when they haven’t yet developed an immune system) with a viral supernatant, consequently causing a life-long viremia: 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 insertion 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 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 integration 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 multistep progression of human cancers.

Murine leukemia viruses in hematopoietic malignancies

MMLV 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 MMLV recruits the basal transcriptional machinery as it drives the replication of the virus and integrates into the genome (2). Usually, a single genomic integration is not enough to promote tumorigenesis, and multiple rounds of proviral insertions within the same genome are required 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 integration 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).

MMLV mainly induces T- and B-cell leukemia/lymphoma in mice. By analyzing MMLV integrations in murine tumors, several oncogenes have been identified, including c-Myc, Pim1, Pim2, and Pst1 (10). Many studies have created recombinant MLVs by substituting LTRs from other viruses and have tested them in mice, for example, Starkey and colleagues (11) replaced the MMLV 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 MMLV, and indeed led to the identification of new cancer genes, such as Mf8t, Jundm2, Abi1, and Ruas2 (12).

Different spontaneous mutants as well as engineered versions of MMLV displayed the capability to induce other types of hematopoietic malignancies. The SL3-3 MLV 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, whereas the Graffi-1.4 virus induces mainly acute myeloid leukemia (AML; ref. 14). A different approach is to use a specific mouse model to skew the tumor spectrum. For instance, using MMLV 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 viremia. 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

MMTV is a retrovirus with a similar tumorigenic behavior as MMLV but with a specific tropism for mammary cells. Because 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 viremic mother. After a few days, the virus infects B cells in lymphoid tissue of the gut, for example, Peyer’s patches. Here, by stimulating a T-cell 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 neoplastic lesions, which eventually evolve into hormone-independent mammary adenomas and carcinomas.

MMTV has been used for forward genetic studies since 1982, when Nusse and Varmus identified Wnt1 as a frequent insertion target (20). Different studies conducted 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 used MMTV than MMLV, but recently Theodorou and colleagues conducted the first large-scale MMTV insertional mutagenesis screen in the mouse mammary gland (22). One hundred and sixty 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 used for insertional mutagenesis screens, the myeloblastosis-associated virus type 2 (MAV-2). MAV-2 is an avian replication-competent nonacute 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 MMLV to induce tumors in nonhematopoietic cell types, by combining tissue-specific viral delivery and oncogene transfer. Autocrine activation of platelet-derived growth factor B-chain (PDGFβ) receptor is a key step in human glioblastoma formation. By intracerebral injection of a replication-defective MMLV-encoding PDGFβ in newborn mice, Uhrbom and colleagues induced the development of gliomas. Viral propagation was enabled by coinfection of a replication-competent MMLV (28). Exploiting this model of glioma, Johansson and colleagues were able to identify 66 candidate cancer genes. Some of these genes probably synergize with PDGFβ 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 MMLV in leukemia. The spectrum of MMLV-induced brain tumors mimics the kinetic and the molecular lesions of human glioblastoma multiforme and oligodendroglia. 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 intermediate, 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 conducted (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 reintegrated 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**

Because DNA transposons are actively mobile only in plants and invertebrates, the Sleeping Beauty transposon system was generated by reverse engineering in 1997, based on sequence comparison between multiple salmonid species of nonfunctional transposase genes of the Tc1/mariner family that had accumulated inactivating mutations (39). The Sleeping Beauty system is composed of two elements: a transposase, that is, the enzyme responsible for mobilization, and the transposon, that is, the actual mobilized sequence of DNA, which is flanked by the binding sites for the Sleeping Beauty transposase, the so-called inverted repeats/direct repeats (Fig. 4A). Transposition occurs when the transposase binds two sites in each inverted repeat/direct repeat, and the closer the inverted repeats/direct repeats, 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 (Fig. 4B). The cargo of the transposon (the segment contained between the two inverted repeats/direct repeats) 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 nonautonomous bipartite 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 nonautonomous 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 Sleeping Beauty cargos, which have been used, are mutagenic elements designed to mimic and potentiate those present in retroviruses. The "oncogene trap" of Sleeping Beauty transposons can induce loss-of-function mutations in TSGs, as well as gain-of-function mutations in proto-oncogenes. Because Sleeping Beauty is a cut and paste transposon, and there is only a 40% to 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. 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 concatemerized 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 use of Sleeping Beauty for oncogenic insertional mutagenesis was first shown 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 Sleeping Beauty transposon, which integrates in a genome locus in the form of a multicopy concatemer. 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 and colleagues 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 (44). The study by
Transposons for insertional mutagenesis. A, Sleeping Beauty (SB) T2/Onc2 can deregulate the expression of oncogenes or inactivate the expression of TSGs. T2/Onc2 contains a MSCV 5'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 (top). T2/Onc2 also contains two splice acceptors and a bidirectional polyA (pA) and can thus prematurely terminate transcription of a TSG when integrated within the TSG 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 β-actin gene (SA; bottom). B, the PB-based insertional mutagenesis system described in ref. (63). Mouse lines carrying the genetic components of the transposon systems. Top, RosaPB and RosaSB knockin mice express PB or Sleeping Beauty transposase under control of the constitutively active Rosa26 promoter. Bottom, 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-inverted repeats/direct repeats and can therefore be mobilized with both transposases. CBASA, Carp β-actin splice acceptor; En2-SA, engrailed-2 exon-2 splice acceptor; SD, Foxf2 exon-1 splice donor; pA, bidirectional SV40 polyadenylation signal. Modified from ref. 8.
Dupuy and colleagues used a knockin 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 (43). Both T2/Onc and T2/Onc2 transposons are “oncogene traps” engineered with engrailed-2 (En-2) splice acceptor sites followed by the simian virus 40 (SV40) pA sequence in both orientations, to intercept upstream splice donor 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 splice donor site, to create fusion transcripts after splicing with downstream endogenous exons (Fig. 5A). The only difference between T2/Onc and T2/Onc2 is that T2/Onc2 contains a larger fragment of a splice acceptor sequence and is flanked by optimized Sleeping Beauty transposase-binding sites that increase Sleeping Beauty 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, Sleeping Beauty 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 (ref. 8; Fig. 5A).

Double-transgenic mice (for transposon and transposase) generated by Dupuy and colleagues 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 and colleagues, 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<sup>−/−</sup> mice developed soft tissue sarcomas (similar to what had been observed previously in Arf<sup>−/−</sup> mice; refs. 44, 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 MSCV derived from a cloned MMLV and therefore promote oncogene activation preferably in the hematopoietic compartment. Dupuy and colleagues (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 conducted 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 the tissue-specific pathways of carcinogenesis.

**Tissue-specific Cre-inducible Sleeping Beauty mutagenesis**

To direct mutagenesis to specific tissues, a strategy based on Cre-mediated activation of the Sleeping Beauty transposase has been developed. In this setting, the Sleeping Beauty 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 Sleeping Beauty allele and a tissue-specific Cre-recombinase (47, 48). One of these studies (47) was conducted 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 Sleeping Beauty system in mice heterozygous for loss-of-function mutations in Apc in the germline or somatically induced by Cre recombination. Another study (48) used triple-transgenic mice with the same three transgenes 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 hepatocellular carcinoma, quadruple-transgenic mice were also bred to carry a conditional dominant-negative Tp53 allele. Mice from this model developed different types of liver lesions, ranging from preneoplastic nodules to hepatic adenomas and hepatocellular carcinomas with lung metastases.
Collectively, these studies identified several candidate cancer genes of relevance to human hepatocellular carcinoma 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 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, Sleeping Beauty 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; ref. 51). In a similar way, a recent study exploited inducible tissue-specific activation of Sleeping Beauty transposition in a mouse model of melanoma (52). Here, a tissue-specific tyrosinase promoter (active in melanocytes) was used to drive a tamofoxifen-inducible Cre. Topical administration of tamofoxifen induced melanocyte-specific transposition that when combined with a conditional BrafF595E allele (analogous to human BrafV600E) induced melanogenesis. This screen led to the discovery of several competing endogenous RNAs behaving as tumor suppressors by acting as ‘sponges’ for miRNAs 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 AML (53). Because 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 Grafil1.4 MLV that preferentially promotes myeloid leukemia rather than lymphoid leukemia. In this model, the expression of mutant Npm1 and the Sleeping Beauty transposase was induced using the hematopoietic-specific MxiCre allele. Administration of polyinosinic:polycytidylic acid activated the Cre and resulted both in the expression of the mutated Npm1 and the activation of Sleeping Beauty insertional mutagenesis.

Following a different approach, a model of B-cell precursor ALL (BCP-ALL) was created by combining the expression of the oncogenic fusion product Etv6-RUNXI with Sleeping Beauty insertional mutagenesis (54). ETV6-RUNXI 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 because it requires cooperating mutations to cause overt leukemia. Therefore, van der Weyden and colleagues generated a knockin mouse model that expresses Etv6-RUNXI fusion product and a hyperactive version of Sleeping Beauty transposase 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-RUNXI. These mice developed leukemias with an early onset and a phenotype significantly skewed to BCP-ALL compared with mice carrying the Etv6-RUNXI alone.

Recently, Sleeping Beauty insertional mutagenesis has been widely used 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.

Sleeping Beauty insertional mutagenesis to characterize the process of metastasis

Sleeping Beauty insertional mutagenesis may also be exploited to track and characterize the process of metastasis. Indeed, in contrast with retroviruses that can spread widely within the host organism, transposons undergo multiple cycles of mobilization and reintegration 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 Sleeping Beauty-induced hepatocellular carcinomas (48). As transposition continues as tumors disseminate, drivers of disease spread may be identified. In a recent study, insertional mutagenesis was conducted by activating Sleeping Beauty in the cerebellar progenitor compartment of Pch+/+ or Tpr53-mutated mice, which caused a significant increase in the incidence of medulloblastoma formation and spread of these tumors to distal sites. A comparative analysis of Sleeping Beauty 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 Sleeping Beauty is the PiggyBac (PB) system. 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 Sleeping Beauty and PB are: PB system can efficiently mobilize larger cargos (up to several hundred Kb); PB integrates at a TATA 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 bifunctional transposon containing 5’ and 3’ consensus sequences for both the Sleeping Beauty and PB transposases. The design of the integrating cassettes used was similar to T2/Onc transposon because they contained a promoter, splice donor, splice acceptor, 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 (PGK) promoter (Fig. 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 Sleeping Beauty (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 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 more than 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 5,590 nonredundant transposon integration events resulted in the identification of 72 CIS. Analogous to Sleeping Beauty, gain-of-function events were found near oncogenes, whereas TSGs contained intragenic loss-of-function mutations. Remarkably, 42% of the candidate cancer genes identified by PB insertional mutagenesis were not significantly hit in the previous studies conducted using retroviruses or Sleeping Beauty 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 Sleeping Beauty (64). In addition, 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 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 preintegration complex is capable of integrating into cells that are nonreplacating. Exploiting this feature of lentiviral vectors has allowed them to be used as efficient tools for gene transfer for gene therapy applications.

Because lentiviral vectors 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 lentiviral vectors 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 lentiviral vectors include: (i) a promoter and polyadenylation signal for the viral genome, (ii) a viral packaging signal ($\psi$) to direct incorporation of vector RNA into virions, (iii) signals required for reverse transcription, including a transfer RNA-binding site (PBS), a polyuridine tract 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 polyuridine tract from the polymerase open reading frame 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 lentiviral vectors is achieved by providing the structural and regulatory proteins in trans by transfection into a packaging cell line, together with the vector genome (Fig. 6).

In the last years, several improvements have been made to lentiviral vector packaging systems by modifying either the packaging construct to increase their biosafety, or by making the vector genome self-inactivating (SIN), so that upon integration, lentiviral vectors 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 lentiviral vector development is the viral cellular tropism, which is dictated by the interactions of the viral envelope glycoproteins with cellular receptors on the target cells. Lentiviral vector have the ability to incorporate in their envelope proteins from both related and unrelated viruses when they are provided in trans during lentiviral vector 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 used is the G glycoprotein of the vesicular stomatitis virus (VSV-G). It confers pantropism to the vector, because it mediates fusion of the viral envelope with the cellular membrane and increases particle stability, thus allowing vector concentration by ultracentrifugation.

Production of lentiviral vectors 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, and (iv) the vector genome (Fig. 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). Because their RNA genome does not encode for the structural proteins that are essential for the assembly of the
viral particles, lentiviral vectors are devoid of any replication potential. The resulting VSV-G pseudotyped lentiviral vectors are therefore capable to efficiently transduce in vivo replicating and nonreplicating cells from different tissues, thus providing an efficient single round of integration.

**Specifically modified lentiviral vectors can induce insertional mutagenesis**

Because lentiviral vectors represent a powerful tool for gene therapy and recent clinical trials using γ-retroviral vectors have raised several concerns over the potential of these vectors to act as insertional mutagens, the biosafety profile of different γ-retroviral vector and lentiviral vector 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-retroviral vectors to lentiviral vectors 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 

Because lentiviral vectors represent a powerful tool for gene therapy and recent clinical trials using γ-retroviral vectors have raised several concerns over the potential of these vectors to act as insertional mutagens, the biosafety profile of different γ-retroviral vector and lentiviral vector 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-retroviral vectors to lentiviral vectors 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 

Because lentiviral vectors represent a powerful tool for gene therapy and recent clinical trials using γ-retroviral vectors have raised several concerns over the potential of these vectors to act as insertional mutagens, the biosafety profile of different γ-retroviral vector and lentiviral vector 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-retroviral vectors to lentiviral vectors 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 

Because lentiviral vectors represent a powerful tool for gene therapy and recent clinical trials using γ-retroviral vectors have raised several concerns over the potential of these vectors to act as insertional mutagens, the biosafety profile of different γ-retroviral vector and lentiviral vector 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-retroviral vectors to lentiviral vectors 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 

Because lentiviral vectors represent a powerful tool for gene therapy and recent clinical trials using γ-retroviral vectors have raised several concerns over the potential of these vectors to act as insertional mutagens, the biosafety profile of different γ-retroviral vector and lentiviral vector 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-retroviral vectors to lentiviral vectors 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 

Because lentiviral vectors represent a powerful tool for gene therapy and recent clinical trials using γ-retroviral vectors have raised several concerns over the potential of these vectors to act as insertional mutagens, the biosafety profile of different γ-retroviral vector and lentiviral vector 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-retroviral vectors to lentiviral vectors 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 

Because lentiviral vectors represent a powerful tool for gene therapy and recent clinical trials using γ-retroviral vectors have raised several concerns over the potential of these vectors to act as insertional mutagens, the biosafety profile of different γ-retroviral vector and lentiviral vector 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-retroviral vectors to lentiviral vectors 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 

Because lentiviral vectors represent a powerful tool for gene therapy and recent clinical trials using γ-retroviral vectors have raised several concerns over the potential of these vectors to act as insertional mutagens, the biosafety profile of different γ-retroviral vector and lentiviral vector 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-retroviral vectors to lentiviral vectors 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 

Because lentiviral vectors represent a powerful tool for gene therapy and recent clinical trials using γ-retroviral vectors have raised several concerns over the potential of these vectors to act as insertional mutagens, the biosafety profile of different γ-retroviral vector and lentiviral vector 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-retroviral vectors to lentiviral vectors 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 

Because lentiviral vectors represent a powerful tool for gene therapy and recent clinical trials using γ-retroviral vectors have raised several concerns over the potential of these vectors to act as insertional mutagens, the biosafety profile of different γ-retroviral vector and lentiviral vector 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-retroviral vectors to lentiviral vectors 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 

Because lentiviral vectors represent a powerful tool for gene therapy and recent clinical trials using γ-retroviral vectors have raised several concerns over the potential of these vectors to act as insertional mutagens, the biosafety profile of different γ-retroviral vector and lentiviral vector 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-retroviral vectors to lentiviral vectors 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 

Because lentiviral vectors represent a powerful tool for gene therapy and recent clinical trials using γ-retroviral vectors have raised several concerns over the potential of these vectors to act as insertional mutagens, the biosafety profile of different γ-retroviral vector and lentiviral vector 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-retroviral vectors to lentiviral vectors 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 

Because lentiviral vectors represent a powerful tool for gene therapy and recent clinical trials using γ-retroviral vectors have raised several concerns over the potential of these vectors to act as insertional mutagens, the biosafety profile of different γ-retroviral vector and lentiviral vector 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-retroviral vectors to lentiviral vectors 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 

Because lentiviral vectors represent a powerful tool for gene therapy and recent clinical trials using γ-retroviral vectors have raised several concerns over the potential of these vectors to act as insertional mutagens, the biosafety profile of different γ-retroviral vector and lentiviral vector 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-retroviral vectors to lentiviral vectors 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
Lentiviral vector-based insertional mutagenesis for cancer gene discovery

We have recently developed a lentiviral vector-based insertional mutagen and exploited it to conduct a screening for new liver cancer genes (88). We constructed a transgeneless lentiviral vector with highly active hepatospecific enhancer–promoter sequences (enhanced transthyretin; ref. 89) in the LTR to activate genes upon integration in hepatocytes while avoiding unwanted effects in nonparenchymal cells. By a single injection of this lentiviral vector into newborn mice, we efficiently induced hepatocellular carcinomas in 3 different mouse models and generated a collection of 30 hepatocellular carcinomas covering different grades of hepatocellular carcinoma (from G1 to G3) and displaying gene expression signatures reminiscent of those found in human hepatocellular carcinomas.

Although the retrieval of vector integration resulted in low number of univocally mapped integrations sites (172 on 30 hepatocellular carcinomas) compared with 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 hepatocellular carcinomas. Moreover, when we stratified human patients with hepatocellular carcinoma according to the expression level of the novel cancer genes or to the gene expression signatures induced by their upregulation (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 hepatocellular carcinoma.

In this lentiviral vector-based screening, we observed a strong and unprecedented enrichment for tumors bearing a single CIS integration (83%), when compared with other insertional mutagenesis systems. This high efficiency of CIS retrieval is likely the result of the combination of nonrepllicative nature of lentiviral vectors, the high efficiency of lentiviral vector transduction in hepatocytes, and the efficient coverage of genes by integrations. The lentiviral vector 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 lentiviral vectors when compared with transposons and retroviruses. On the other hand, this exclusive peculiarity of lentiviral vectors may facilitate the identification of earlymutated genes in carcinogenesis by eliminating the bystander and progression-related integrations that are accumulated during the entire lifetime 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 lentiviral vectors 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; ref. 90) and linear amplification-mediated PCR (LAM-PCR; ref. 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 site, with the products of these reactions being sequenced on capillary (92) or next-generation sequencing platforms (91, 93). Because 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 restriction enzyme sites have been developed. These approaches use polymerase-based strategy (94) or DNA shearing (24, 95) to provide a semiquantitative 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.

Because 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). Although 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% to 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, to account for the preference of Sleeping Beauty transposon to integrate in TA dinucleotides, frequency of integration was corrected for the TA content of each chromosome (48). Because lentiviral vector integrations cluster in large chromosomal areas giving rise to false-positive calls in CIS stats, tests to calculate whether 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 of 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, whereas shRNAs are notorious for off-target effects and incomplete silencing. In addition, they rely on the composition of predefined libraries, while the genome wide distribution of insertional mutagens allows to identify even nonannotated 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 conducting screenings for cancer genes by insertional mutagenesis is that each insertional mutagen has its own intrinsic integration biases. Transposons, and especially Sleeping Beauty 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, that is, 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 2 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 Sleeping Beauty 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, MLV's display a strong bias for integration in the neighborhood of the TSS of genes and in CpG islands. Moreover, γ-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). Lentiviral vectors are characterized by a strong bias of integration within genes and specifically within actively transcribed genes (83–85). Although some of the lentiviral vectors, 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 by only characterizing the integration biases of each specific insertional mutagen in the specific tissue of interest in nonselective 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 nonproliferating cells typical of β- and γ-retroviruses have been overcome by transposable elements. However, the development of specific knockin mouse strains for the activation of transposon insertional mutagenesis in several different tissues is still labor intensive and time consuming. Because 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 lentiviral vectors 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 untag transformation events that were relevant for the tumor induction. Moreover, Sleeping Beauty 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 by only increasing the number of tumors and integrations that are analyzed. Retroviruses are devoid of this problem, but they also have intrinsic limitations. Besides
Cancer Gene Discovery by Insertional Mutagenesis

their limited tropism, it was shown that, because 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 lentiviral vectors are characterized by a preference to drive the upregulation 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 conducted 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 biologic validation is required.

To overcome this issue, recently a SIN PB 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 lentiviral vectors 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. Lentiviral vectors represent a promising tool for insertional mutagenesis, but their extensive exploitation in the cancer gene discovery arena has just started.

Nonetheless, lentiviral vectors may suffer from some potential limitations too. Indeed, because lentiviral vectors are replication deficient, an extensive transduction of the targeted organ is required to obtain significant levels of mutagenesis and eventually cell transformation. Therefore, lentiviral vector-mediated insertional mutagenesis may be inefficient in organs that are difficult to access. Lentiviral vector 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 produced, 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 the advantages and disadvantages in their use and make them complementary to each other. Importantly, the recent studies with PB transposons and lentiviral vectors 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, α-retroviral vectors have been recently developed for gene therapy applications (107, 108). Because they have a different integration profile compared with other integrating vectors, their modification to generate efficient insertional mutagens could provide a new tool for cancer gene discovery screens, as we did with lentiviral vectors.

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 futures 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 shown to occur in some gene therapy trials with γ-retroviral vectors (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.

Disclosure of Potential Conflicts of Interest
No potential conflicts of interest were disclosed.

Authors’ Contributions
Conception and design: M. Ranzani, E. Montini

References
4. van Lohuizen M, Breuer M, Berns A. N-myc is frequently activated by proviral insertion in MuLV-induced T cell lymphomas. EMBO J 1989;8:133–6.
Cancer Gene Discovery by Insertional Mutagenesis


www.aacajournals.org Mol Cancer Res; 11(10) October 2013

Published OnlineFirst August 8, 2013; DOI: 10.1158/1541-7786.MCR-13-0244

© 2013 American Association for Cancer Research.
77. Naldini L. *Ex vivo* gene transfer and correction for cell-based tera-

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 gamma retroviral vectors. Mol Ther 2009;

79. Schambach A, Zychlinski D, Ehmstroem B, Baum C. Biosafety

Lentiviral hematopoietic stem cell gene therapy benefits metachro-

81. Aiuti A, Biasco L, Scaramuzza S, Ferrua F, Cicalese MP, Baricordi C,
et al. Lentiviral hematopoietic stem cell gene therapy in patients with

82. Hecker D, Schwarzer A, Haemmerle R, Steinemann D, Rudolph C,
Skawran B, et al. Lentiviral vector induced insertional haploinsuffi-

83. Mitchell RS, Beitzel BF, Schroder AR, Shinn P, Chen H, Berry CC,
et al. High-resolution insertion-site analysis by linear ampli-

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.

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

86. Cesana D, Sguaidino J, Rudolosso L, Merella S, Naldini L, Montini E.
Whole transcriptome characterization of aberrant splicing events indi-
cated by lentiviral vector integrations. J Clin Invest 2012;122:
1667–76.

Lentiviral vector integration in the human genome induces alternative
splicing and generates aberrant transcripts. J Clin Invest 2012;122:
1653–66.

88. Ranzani M, Cesana D, Bartholomea CC, Sarvito F, Pala M, Benedic-
enti F, et al. Lentivirus vector-based insertional mutagenesis iden-
tifies genes associated with liver cancer. Nat Methods 2013;10:
137–9.

89. Vigna E, Amendola M, Benedicenti F, Simmons AD, Pollenz A, Naldini
L. Efficient Tet-dependent expression of human factor IX in vivo by a

90. Devon RS, Porteous DJ, Brookes AJ. Splinkerettes 91.

91. Schmidt M, Schwarzwaelder K, Bartholomea CC, Zoua K, Ball C, Pilz I,
et al. High-resolution insertion-site analysis by linear amplification

92. Uren AG, Kool J, Matentzoglu K, de Ridder J, Maitson J, van Uitert M,
et al. Large-scale mutagenesis in p19ARF- and p53-deficient mice
identifies cancer genes and their collaborative networks. Cell
2008;133:727–41.

W, et al. Genome-wide high-throughput integrase analyses by
rILAM-PCR and next-generation sequencing. Nat Protoc 2010;5:
1379–95.
Cancer Gene Discovery: Exploiting Insertional Mutagenesis

Marco Ranzani, Stefano Annunziato, David J. Adams, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1541-7786.MCR-13-0244

Visual Overview
A diagrammatic summary of the major findings and biological implications:
http://mcr.aacrjournals.org/content/11/10/1141/F1.expansion.html

Cited articles
This article cites 110 articles, 33 of which you can access for free at:
http://mcr.aacrjournals.org/content/11/10/1141.full.html#ref-list-1

Citing articles
This article has been cited by 4 HighWire-hosted articles. Access the articles at:
/content/11/10/1141.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.