Review

Predicting preferential DNA vector insertion sites: implications for functional genomics and gene therapy

Christopher S Hackett*, Aron M Geurts[†] and Perry B Hackett[‡]

*Biomedical Sciences Graduate Program and Department of Neurology, University of California San Francisco, Room U441K, Parnassus Ave, San Francisco, California 94143-0663, USA. [†]Human and Molecular Genetics Center, Medical College of Wisconsin, Room HRC 5200, Watertown Plank Rd, Milwaukee, Wisconsin 53226, USA. [‡]Department of Genetics, Cell Biology, and Development, The Arnold and Mabel Beckman Center for Transposon Research, Gene Therapy Program, and Cancer Center, University of Minnesota, Room 6-160 Jackson Hall, Church St, Minneapolis, Minnesota 55455, USA.

Correspondence: Perry B Hackett. E-mail: perryh@umn.edu

Abstract

Viral and transposon vectors have been employed in gene therapy as well as functional genomics studies. However, the goals of gene therapy and functional genomics are entirely different; gene therapists hope to avoid altering endogenous gene expression (especially the activation of oncogenes), whereas geneticists do want to alter expression of chromosomal genes. The odds of either outcome depend on a vector's preference to integrate into genes or control regions, and these preferences vary between vectors. Here we discuss the relative strengths of DNA vectors over viral vectors, and review methods to overcome barriers to delivery inherent to DNA vectors. We also review the tendencies of several classes of retroviral and transposon vectors to target DNA sequences, genes, and genetic elements with respect to the balance between insertion preferences and oncogenic selection. Theoretically, knowing the variables that affect integration for various vectors will allow researchers to choose the vector with the most utility for their specific purposes. The three principle benefits from elucidating factors that affect preferences in integration are as follows: in gene therapy, it allows assessment of the overall risks for activating an oncogene or inactivating a tumor suppressor gene that could lead to severe adverse effects years after treatment; in genomic studies, it allows one to discern random from selected integration events; and in gene therapy as well as functional genomics, it facilitates design of vectors that are better targeted to specific sequences, which would be a significant advance in the art of transgenesis.

Introduction

Elements such as viruses and transposons, through evolution with their host organisms, have acquired the ability to integrate into host genomes and ultimately shuffle genetic material between organisms. These elements have an established history in molecular biology and genetics research because of their ability to deliver specific genetic cargo, randomly disrupt host genomes for genetic screens, and serve as vectors for delivery of therapeutic expression cassettes to treat human disease. Viral vectors have been the predominant tools for these applications for three reasons: the ease and efficiency with which specific viral genetic cassettes can be introduced into cells; the vast accumulated knowledge of viruses and their mechanisms of gene transfer into chromosomes; and the large number of sites in genomes into which they can integrate. Retroviruses in particular have been used for random insertion into chromatin to interrupt host genes (insertional mutagenesis) and thereby identify their function [1-3] as well as for delivery of therapeutic genes [4-6]. Moreover, viral activation of oncogenes and, more
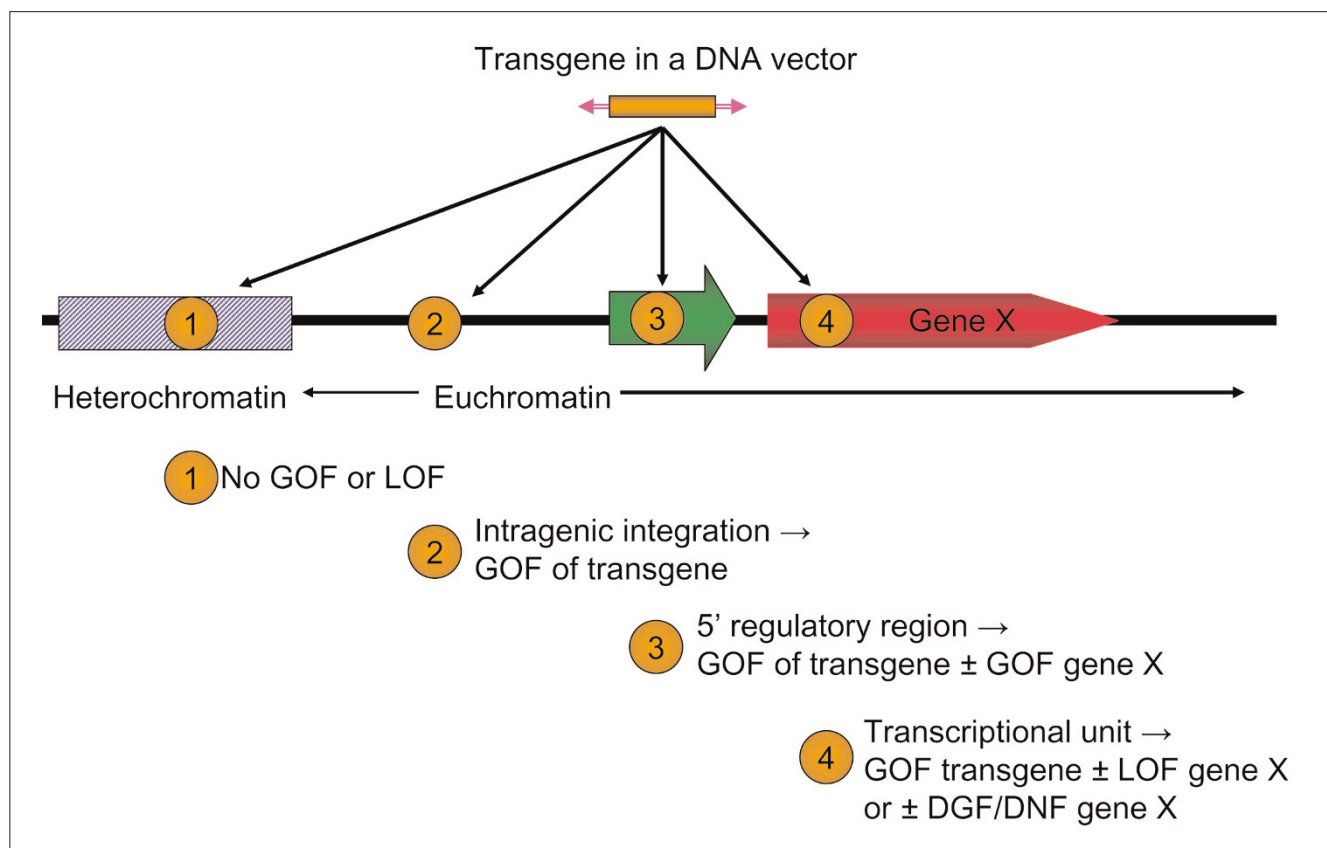
Figure 1
Potential genetic consequences of integration of transgenic cassettes into chromatin. An expression cassette (orange box) in a viral or nonviral vector (represented by purple inverted arrowheads, which indicate either inverted or direct terminal repeats) can integrate into four classes of chromatin. (1) Integration into heterochromatin will most likely result in the suppression of expression of the transgene and essentially no genetic consequences for the host. (2) Integration into intergenic regions of euchromatin is the most desirable outcome; the transgenic cassette is expressed, leading to a gain of function (GOF) in the host cell. (3) Integration into a transcriptional regulatory region can have several outcomes including expression (GOF) of the transgenic cassette, potentially modified by neighboring enhancer and silencer elements in the region. Regulatory elements in the transgenic cassette may either enhance expression of the neighboring gene (GOF for gene X) or, in rare cases, block expression of an active gene. (4) Integration of the vector into a transcriptional unit may allow expression of the transgene but block expression of the host gene leading to a phenotypic loss of function (LOF). Integration within some genes can also lead to a dominant gain of function (DGF) or production of a dominant-negative form (DNF) of the original gene X. A further discussion of effects of insertional mutagenesis can be found in the reports by Carlson and Largaespada [61] and Collier and Largaespada [154].

recently, inactivation of tumor suppressors have been used to discover several novel genes that are involved in cancer progression [7-12]. The consequence of insertional activation of host cell oncogenes by viral vectors, however, has emerged as a major risk/obstacle in gene therapy, with a few cases of leukemia arising from oncogene activation by therapeutic vectors [13,14]. The potential genetic consequences of insertions of integrating vectors are summarized in Figure 1.

Risk of oncogene activation in gene therapy
Activation of oncogenes in mice by insertionally mutagenic retroviruses suggested that inadvertent oncogene activation resulting from the use of relatively benign therapeutic vectors is a potential risk associated with gene therapy. Gene therapy vectors are extensively minimized to eliminate their replicative potential and reduce their collateral effects on the target genome [15]. However, extensive testing in animals demonstrated that the risk of oncogenic activation was real, although variable and dependent on the viral vector used, the genetic cargo, and the background genetics of the model system [16-22]. Given what was assumed to be acceptable risk, retroviral gene therapy trials have been conducted in human patients. Nearly 1,000 clinical gene therapy trials have been initiated, more than half with retroviral vectors [4], but as yet no vectors have been approved in the USA for clinical gene therapy outside the clinical trial setting [23]. (Gendicine, an adenovirus designed to restore p53 function in cancerous cells, has been approved for commercial human gene therapy in China [24], although this vector is essentially nonintegrating and thus carries decreased risk for oncogene activation via vector insertion.)

The worst fears of the gene therapy field, oncogene activation, were realized when three of more than 20 patients treated for X-linked severe combined immunodeficiency disease (X-SCID) developed leukemia. These adverse findings, including one death, occurred 3 years or more after administration of therapeutic murine leukemia virus (MLV)-derived retrovirus vectors [25,26]. The linkage between treatment and leukemias could be inferred because the expanded transformed cell populations harbored clonal integrations of the therapeutic vector, which suggested a biologic selection for the retrovirus-induced mutation [27-30]. However, these studies also indicated that clonal expansions in some cases appeared to be temporary and did not always lead to adverse effects, features that could actually improve the likelihood of successful gene therapy. The cause of at least two of the leukemias appears to be insertion of the MLV vector close to the LMO2 oncogene, which led to LMO2's activation by enhancers in the long terminal repeat (LTR) sequences of the vector [31-33]. Retrospective examination of the role in LMO2 during development supported this conclusion [34,35]. Subsequent studies in which the cargo gene IL2γc was over-expressed in mice (albeit at levels higher than in the X-SCID leukemia patients) suggested that this gene could itself act as an oncogene in T cells [36]. Also, simultaneous activation of IL2γc and LMO2 by oncogenic retroviruses had been observed in one mouse, suggesting a possible genetic interaction between the cargo IL2γc gene and LMO2 [33]. The relevance of these observations to clinical cases, however, is highly debatable [37,38].

In contrast, other gene therapy trials that employed retroviral vectors to treat adenosine deaminase deficiency [39-41] and chronic granulomatosis disease (CGD) [42] have not yet reported any equivalent adverse events. In the CGD study, there appeared to be powerful selection for integration events of the spleen focus-forming virus vector, which also was used as a vector for X-SCID [43], into the neighborhoods of three previously identified genes, namely *MDS-EVI1*, *PRDM16*, and *SETBP1*, which have been associated with enhanced proliferation following integration of retroviruses with activating LTRs [44-46]. As noted previously, findings of preferential integration around certain genes is not necessarily due to a preference for these genes, but may rather be a consequence of clonal expansion that can be transient and thereby beneficial in terms of enhancing the number of therapeutic cells. A similar effect has also been observed in nonhuman primate studies, indicating that this result may not be unique [19]. Despite the striking incidence of common integration sites that are often associated with tumor or leukemia formation [8,47,48], there has been no report of adverse events in the CGD patients and no indication that the corrective gene, *gp91phox*, synergizes with any of the three common integration site genes to promote growth. Likewise, a murine stem cell retrovirus has been used to deliver the α and β chains of the antiMART-1 T-cell receptor complex *ex vivo* into peripheral blood lymphocytes to treat melanoma without any apparent adverse effects, although integration sites were not examined and the patient population had low odds for survival, even with the treatment (two out of 15 survived) for more than 1 year [49].

Taken together, the results of the CGD and X-linked plus adenosine deaminase SCID trials demonstrate that oncogenesis is not necessarily an inherent, inevitable side effect of gene therapy. In more than 20 patients, the genetic deficiencies of more than 80% have been fully corrected, allowing them to lead normal lives. However, tumors and leukemias can take years to manifest, and these trials are in their early years. A clearer understanding of the variables that underlie oncogenesis is needed in order to increase the safety of these trials. These variables include insertion site preferences of therapeutic vectors, their abilities to activate nearby genes, and interactions between specific genetic cargos and activated host genes. Although cargo-host interactions will be specific to each gene therapy approach, the vectors themselves govern other parameters of insertion preference and neighboring gene activation. Analyses of insertion preferences, in particular, have received much recent attention, and have sparked interest in the use of transposons as alternatives to viruses as gene therapy vectors.

Nonviral vectors for introduction of genetic cassettes into mammalian genomes

Transposable elements also have been used for insertional mutagenesis and genetic studies in model organisms, and are being developed as gene therapy agents in humans [50-53]. The most well characterized DNA transposon vector used in mammals is the synthetic *Sleeping Beauty* (SB) transposon system [54], which over the past decade has become a powerful tool in functional genomics to identify genes in vertebrates, including fish and mammals [55-61]. Application of transposon-mediated gene transfer to gene therapy has been explored because it avoids several disadvantages of viral delivery systems. These disadvantages of viruses include the following: (1) their preference for integrating into genes [62-65]; (2) the difficulty with purification to eliminate toxic or infectious agents [66]; (3) their potential to elicit unwanted immune or inflammatory responses [67,68]; (4) the constraint on therapeutic cargo size; and (5) the difficulty and expense associated with their production in large quantities [69,70]. In contrast to viral vectors, preparations of nonviral plasmid-based transposon vectors are relatively inexpensive to purify, are largely nonimmunogenic, and have no hard constraints on genetic sequences that can be delivered.

A negative tradeoff with DNA vectors is increased difficulty in delivery. Delivery of nonviral DNA into mammalian genomes involves avoiding or traversing numerous barriers, including enzymes in the blood and cellular environments, the endothelial lining of vessel walls, cellular plasma

membranes, endosomal membranes, nuclear membranes, and chromosomal integrity [71].

There are three delivery approaches that work across the nanoscale, microscale, and macroscale [72]. Nanoscale delivery involves particles or complexes that are most often designed to be about 100 nm or less in diameter, although sizes up to 1 μm fit into this category. The nanoscale approach comprises delivery of single or small numbers of DNA molecules, which most often are collapsed by polycationic polymers (for example, polylysine and other modified amino acids, and various linear and branched forms of polyethylenimine, among others) or lipids, with or without various ligands (for review, see the report by Wagner and coworkers [71]). Some polycationic complexes are cytotoxic or unstable in the blood, which can be circumvented by encasing the complexes in polyethylene glycol [73]. Alternative delivery routes are those at the microscale and macroscale, in which DNA in packages up to 10 μm are phagocytized (microscale) or enter cells via fusions with other cells or entities larger than 10 μm (macroscale).

In mice, the most effective method for *in vivo* gene transfer and expression has been demonstrated in hepatocytes using simple infusion of naked plasmid DNA under increased pressure. This can be accomplished by hydrodynamic delivery of DNA using high pressure/high volume injection [74,75]. In mouse, this procedure involves injection of a large volume (10% volume/weight) of DNA/saline solution through the tail vein in less than 10 seconds. This procedure results in uptake of infused DNA into as many as 10% of hepatocytes in test animals [74,75] by expanding and rupturing liver endothelium, which in mice heals within 24 to 48 hours [76]. Achieving a clinically feasible method of local delivery to liver in large animals, including humans, is a challenge that is being addressed by more localized hydrodynamic delivery using specialized catheters or pressure cuffs [77,78]. On the microscale, condensing DNA with polyamines such as polyethylenimine to a complex small enough to be taken up by cells into endosomes has been studied intensively [79,80]. Our findings (Hackett PB, Podetz-Pedersen K, Bell JB, McIvor RS, unpublished data) suggest that gene expression following hydrodynamic delivery is about 100-fold more effective than delivery using polyethylenimine [81,82] and only about 10-fold to 100-fold less effective than viral delivery to liver [72]. Alternative delivery *ex vivo* using electroporation is under development and has been achieved in hematopoietic stem cells [83].

Since the development of the SB system, nonviral, integrating DNAs have established themselves as potential vectors for gene therapy. Following hydrodynamic delivery, transposons have been used in mice to cure hemophilias A and B [84-87] and tyrosinemia type I [88,89]. Other somatic delivery methods were used to ameliorate blistering skin disease (junctional epidermolysis bullosa) [90], retard

glioma xenographs [91,92], produce Huntingtin protein in a model of Huntington disease [93], and as a preventive treatment for lung allograft fibrosis [94]. Based on the findings summarized above, we estimate that only about one in 10,000 SB transposons that are delivered to liver or lung actually transpose into chromatin (Hackett PB, unpublished data). Although this is a small fraction, it is possible to deliver more than 10^8 therapeutic cassettes to an animal in order to treat as many as 10% to 20% of liver cells with a single injection of plasmids [84,88,95]. This procedure is sufficient to cure diseases such as hemophilia and tyrosinemia type 1, and to ameliorate other diseases such as mucopoly-saccharidoses types I and VII. Although quantifying the number of transposon insertions per cell has not been done because of the difficulty of cloning insertion sites in mostly nondividing cells in most organs of animals, the expression data are consistent with a single integration in most if not all transgene-expressing cells.

In addition to SB, several other transposon vectors and phage integrase-based vectors have been tested for their potential to deliver therapeutic genes, including *Frog Prince* [96], *Tol2* [89], and *piggyBac* [97], as well as other well characterized transposons such as the *Drosophila* P-elements, which are not mobilized very efficiently in mammalian cells [98]. These vectors differ in their efficiency of gene insertion, genetic cargo capacity, integration site preferences, and effects on chromosomal stability. Among other advantages these systems have over retroviruses as gene therapy vectors, transposons present a wide variety of insertion site preferences that differ from those of retroviruses, with possible consequences for oncogene activation. The charac-teristics of these vectors are summarized in Table 1. The remainder of this review discusses these differences as they relate to gene therapy and functional genomics.

Factors governing insertion site preferences and their variation among vectors

Although most vectors will integrate into a vast number of sites scattered throughout the genome, numerous studies have shown that these integrations are not random with respect to several variables. Global preferences for vector integration can be governed by large-scale genomic context such as coding and regulatory regions of genes, and their transcriptional status, as compared with intragenic regions [99]. The fine tuning that determines specific sites of integration is governed by smaller scale, physical features, such as the specific sequences of nucleotides surrounding insertion sites and DNA structural characteristics derived from these sequences. Figure 2 illustrates some of the physical features of DNA that are influenced by local sequence.

Viruses and transposons exhibit a wide range of variability with respect to preference for genes and transcriptional units. Several studies have mapped hundreds to thousands of

Table 1

Properties of nonviral integrating vectors proposed for gene therapy

Vector system	Properties		
	Activity[a]	Target preferences[b]	Positive/negative attributes[c]
Sleeping Beauty	Standard	TA sites, random	Highly tested/cargo capacity decreases efficiency
ΦC31	Lower	Pseudo-*att* sites	Highly tested/induces chromosomal mutations and rearrangements
piggyBac	Same	TTAA sites (genes)	Too new to evaluate/targets transcription units
Tol2	Higher	Unknown	Cured tyrosinemia type 1 in mice/may target genes, too new to evaluate
ΦBT1	Lower	Pseudo-*att* sites	Cured PKU in mice/too new to evaluate
Frog Prince	Same	TA sites	Too new to evaluate

[a]Activity is given relative to *Sleeping Beauty* (SB) in HeLa cells or other cells in which SB has been tested. [b]Target sites for phage integrases φC31 and φBT1 are not found in mammalian genomes; sequences with similarities to the phage attachment sites (*att* sites) are targets, but they vary with cell type. [c]Evaluation with respect to gene therapy: only SB and φC31 have been extensively tested; the others are too new to know positive and negative attributes. PKU, phenylketonuria.

insertions into human or mouse genomes, and correlated insertion positions with known genes. Many retroviruses exhibit a nonrandom preference for genes [65]. This could be due to greater accessibility of the DNA in 'open' chromatin or interaction of integrase enzymes with cellular factors bound to transcriptional regulatory elements. In the case of HIV, the LEDGF/p75 transcriptional factor may act as a tether between the integrase and transcriptionally activated chromatin [100-102], which is similar to an idea that was proposed previously for designer targeting of integrating vectors [103-105]. In a similar approach using the SB transposon, Yant and coworkers [106] found that SB exhibited a much lower (although nonrandom) preference for genes. Although a preference for transcriptional units might seem beneficial for functional genomics studies, the myriad of recently identified noncoding RNA genes [107] (as well as other RNA product genes such as those encoding rRNA and tRNAs) involved in gene regulation may not be targeted by viral vectors that preferentially integrate into or near protein encoding genes. Targeting of various vectors to these non-coding RNAs in gene therapy, and any resulting deleterious effects, has not been extensively examined.

Many vectors appear to exhibit a preference for specific genes. In insertional mutagenesis studies, the identification of recurrent viral insertions into a specific group of genes was taken to mean that viral activation of these putative oncogenes in individual cells led to clonal expansion among a pool of cells in which every host gene was an equal target for integration (as discussed above for LMO2). However, when MLV insertions were mapped in normal HeLa cells that did not undergo any type of selection, oncogenic or otherwise, many of these same genes harbored recurrent integrations, suggesting that vectors may inherently target specific genes [48]. The basis of this selection is not understood, but it may be similar to that discussed above for HIV.

In addition to general preferences for genes, many viral vectors, including retroviruses, lentiviruses, and adeno-associated virus, preferentially target transcriptional units or their promoters. MLV retroviruses have a preference for integration proximal to transcriptional initiation sites [64,65,108-111], which is a problematic trait, considering that MLV-based vectors are the most commonly used vectors in human gene therapy [4]. HIV and adeno-associated viruses have preferences for entire transcriptional units [100,108,111-113] (see Note added in proof, below); this is in contrast to MLV, which targets only the region proximal to promoters. Additionally, expression array studies have shown that HIV has a preference for transcriptionally active genes [65] as well as an avoidance of chromatin regions in which transcription is repressed [114].

In contrast to these viral vectors, SB transposons and avian leukosis virus (a retrovirus) apparently have only a slight preference for either transcriptional units or their regulatory elements [106,115], with little or no preference for transcriptionally active genes [65]. In one survey, SB exhibited an overall preference for microsatellite repeats, found primarily in noncoding regions [106], possibly due to the preferred target sites found in TA repeats [116]. A study that correlated insertions sites with hundreds of genome annotations [99] illustrated the degree to which genomic features and primary sequence influenced vector integration preferences for several vectors (for example, the L1 and SB transposon insertions were much more influenced by primary sequence than were retroviral vectors). This study also found variable preferences between vectors for elements such as CpG islands, DNase I sensitive sites, and transcription factor binding sites. The recent identification of a periodic sequence encoding nucleosome positioning [117] may also correlate with vector integration patterns, because nucleosomes have been shown to affect patterns of retroviral

integration [118]. Similar studies to identify trends for *piggyBac* and *Tol2* with respect to genome-wide integration preferences will be valuable in assessing the relative safety of these vectors for gene therapy.

Local insertional preferences: DNA sequence and structure

Although many vectors exhibit a preference for genes, and even specific genes, few vectors repeatedly integrate into the same precise position with any significant frequency. Rather, most genes harboring frequent insertions show a distribution of insertions into several positions within the same gene. Some vector integrases, such as those for phages φC31 [119-121], φBT1 [122], as well as the *Escherichia coli Tn7* transposon [123], recognize specific DNA sequences or degenerate sequences that exist in mammalian genomes. SB integrates specifically at a TA dinucleotide, and the *piggyBac* transposon integrates into the sequence TTAA. Because the oncogenic potential of a vector is related to its propensity to integrate in or near a select few genes, understanding local parameters that affect integration may contribute to our ability to assess the risk associated with these vectors in gene therapy.

For retroviruses and the SB transposon, consensuses sequences have been described surrounding the sites of integration [111,124-127]. Although retroviruses do not exhibit a strong consensus sequence, the nonrandom pattern of integrations and the observation that frequently hit sites did not match the consensus sequences led investigators to examine other properties of DNA sequences surrounding target sites, including structural characteristics of the DNA itself. DNA structural characteristics are based on non-Watson and Crick interactions between nucleotides and encompass deformations to the regular double helix structure caused by interactions between adjacent, planar bases (Figure 2). Originally characterized from analysis of crystal structures of DNA bound to histones and other proteins, these characteristics include 'protein-induced DNA deformability', 'A-philicity', and trinucleotide 'bendability'. These properties underlie local variations in DNA structure that are probably relevant to recognition of DNA by transposases and integrases. Early investigations into insertion preferences showed that viruses preferred 'bent' DNA [118,128,129], and several groups have investigated secondary DNA structural patterns in sequences that flank mapped insertion sites for both transposons [115,124,130,131] and retroviruses [111,126] to determine general characteristics of the flanking sequence of 'preferred' integration sites. Similarly, the RAG1/2 protein complex, which has properties akin to the cut-and-paste transposases, recognizes a specific sequence/structure for recombination of antigen receptor genes [132].

Different DNA sequences may produce highly similar patterns of DNA secondary structure, and thus common structural patterns that are preferred for integration may be obscured by approaches that analyze sequence alone. Analysis of secondary structure for a DNA sequence is based on translation of a sliding window of two or three bases into structural values for each 'step'. For example, the tendency of a B-form helix to adopt the A-form (A-philicity; Figure 2) can be predicted by translating each consecutive (over-lapping) dinucleotide into one of 10 A-philicity values for the 16 combinations of base pair transitions [133-135]. Similarly, protein-induced deformability encompasses several changes in base pair orientation from a 'perfect B-form double helix' in a transition between two consecutive base pairs (Figure 2c). All of these changes can be expressed as a single composite parameter of protein-induced DNA deformability known as V_{step} [136-138]. V_{step} represents the physical relationships of any two planar base pairs in terms of their relative shifts and angular orientation. In contrast to A-philicity and protein-induced deformability, DNA bendability is best modeled using a sliding window of three bases, with 64 possible trinucleotide bendability values [139].

An example of DNA structural analysis for the *Tol2* transposon is shown in Figure 3, in which average structural values for each position flanking an insertion site are plotted and compared with a plot of random sequences. In the case of *Tol2*, weak preferences in V_{step} and A-philicity values at specific coordinates are apparent by the peaks in the heavy black lines in Figure 3a,b (left sides), in contrast to the same averages derived from random sequences (right sides). Overall, the bendability around *Tol2* insertion sites exhibits little deviation from a random sequence (Figure 3c), unlike those preferred by SB transposase (Figure 3d). Analysis of hundreds of integration sites for potential gene therapy vectors, including viruses as well as transposons, shows that many have subtle preferences for these variables (Figure 4). For example, the *piggyBac* transposon may favor sites with slightly higher A-philicity, lower bendability, and lower V_{step} values than random sequences. In contrast, 'preferred' SB insertion sites (see below) clearly display a jagged V_{step} pattern and higher bendability. Interestingly, although retroviruses (avian sarcoma virus [ASV], HIV, MLV, and simian immunodeficiency virus) integrate into bent DNA [128], such as that bound to nucleosomes, our analyses of sequences around viral insertion sites do not indicate a particular preference for bendable DNA (Figure 4). A similar, more rigorous approach has been utilized to characterize *Drosophila* P-elements [130] and non-LTR retrotransposons in *Entamoeba histolytica* [140], demonstrating that DNA structural characteristics at insertion sites for both elements are significantly different from collections of random sequences.

For SB, the observation of general structural trends surrounding insertion sites eventually led to the identification of a specific DNA structural pattern governing insertion preference. Vigdal and coworkers [124] observed that
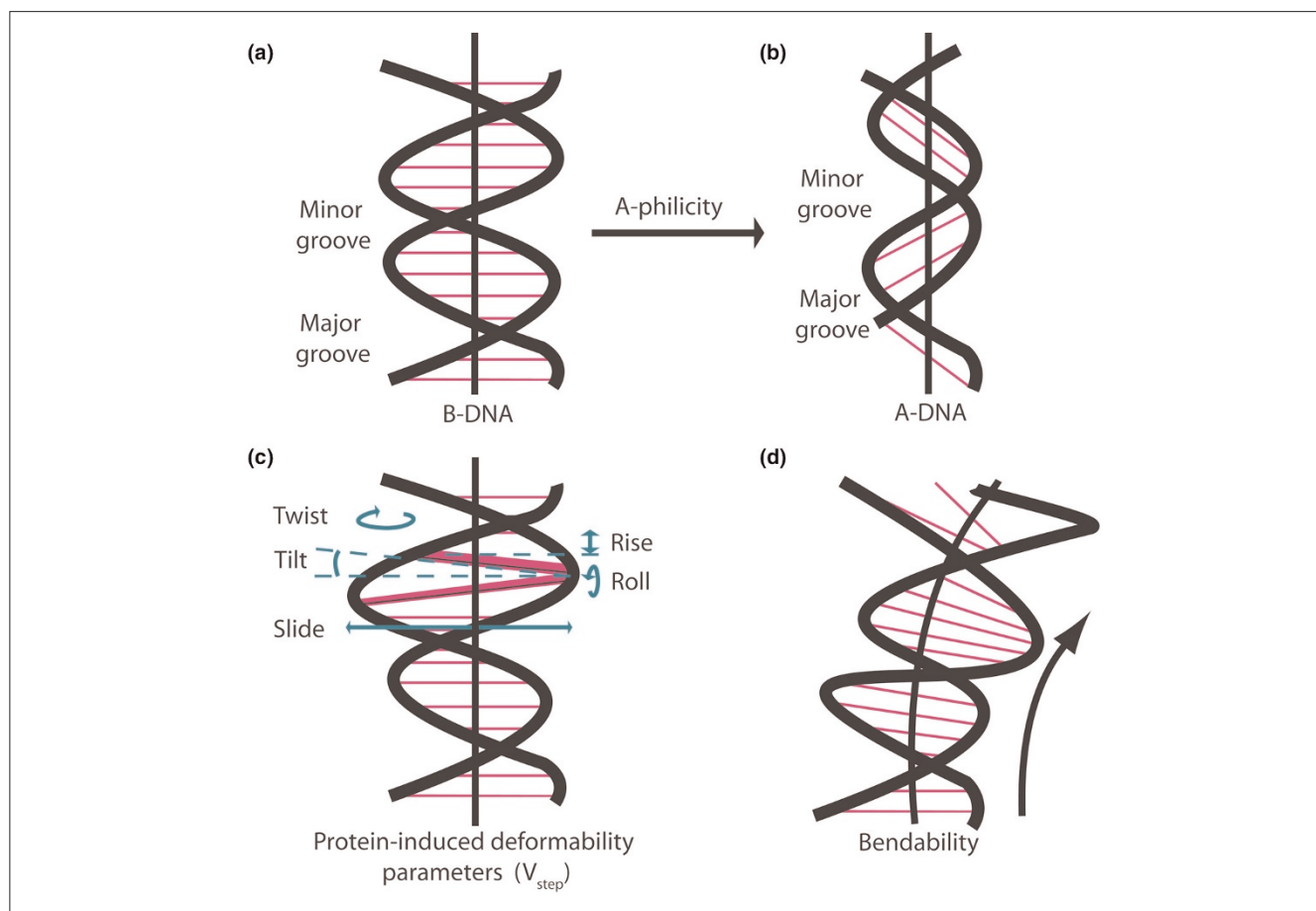
Figure 2
Deviations of DNA structure from the average B-form DNA that play a role modeling three-dimensional structures of specific DNA sequences. The figure illustrates physical parameters of B-form DNA structure that are altered in preferred sites for integration of insertional vectors. **(a)** B-form DNA. **(b)** A-DNA. Interactions between neighboring nucleotides govern the variable energy needed to convert from B-DNA to A-DNA. The propensity of a sequence of B-form DNA to adopt the A-form is referred to as A-philicity [134]. **(c)** Parameters of base pair orientation affected by protein-DNA binding. 'Twist' (horizontal looping arrow) refers to the rotation of base pairs around a central axis (heavy vertical black line); the average rotation between two base pairs is 36°. 'Tilt' (dotted lines) refers to the inclination of the base pairs with respect to the central axis; the average tilt is 0° between base pairs, which are normally parallel in B-form DNA. 'Rise' (vertical double arrowhead) is the distance between adjacent base pairs; the normal spacing is slightly more than 3.3 Å, but it can be more than 3.4 Å at preferred target sites. 'Slide' (horizontal double arrowhead) refers to the shifting of the axis of a base pair out of alignment with the central axis. 'Roll' (vertical looping arrow) refers to rotation of the nucleotide plane around a horizontal axis. A given base pair may be distorted in more than one of these parameters. V_{step} analysis is a method of examining these, and other physical parameters such as 'shift', in terms of a single number that derives from the transition from one base pair to another [131,137]. **(d)** DNA bendability.

increased DNA deformability and A-philicity were features of a consensus sequence that flanked SB TA insertion sites. Subsequently, Liu and colleagues [131] mapped about 200 integrations into a relatively small 7 kilobase plasmid sequence and observed that some common integration sites did not share the consensus sequence. These results identified several 'preferred' TA dinucleotides that harbored recurrent integrations. These preferred integration sites exhibited a striking, specific pattern of alternating high and low deformability (V_{step}) values that were absent in TA sites and that were rarely, if ever, used. This led to the conclusion that SB transposase prefers a 'zigzag' V_{step} pattern of DNA deformability [131], which was later confirmed on a larger, genomic scale [115]. It remains unknown whether these

patterns influence the recognition and binding of the SB transposase, catalysis of the transposon integration, or some other mechanistic factor.

This analysis was repeated for other vectors, including *piggyBac*, P-elements, and several retroviruses [115]. However, only weak structural signatures were detected, which were no more informative than the weak consensus sequences previously identified. A key difference in the SB screen was the level of saturation of a small target, which allowed for the identification of highly preferred sites over nonpreferred TA dinucleotides. In contrast, the datasets for the other vectors were derived from a relatively small number of insertions into mammalian genomes, which were
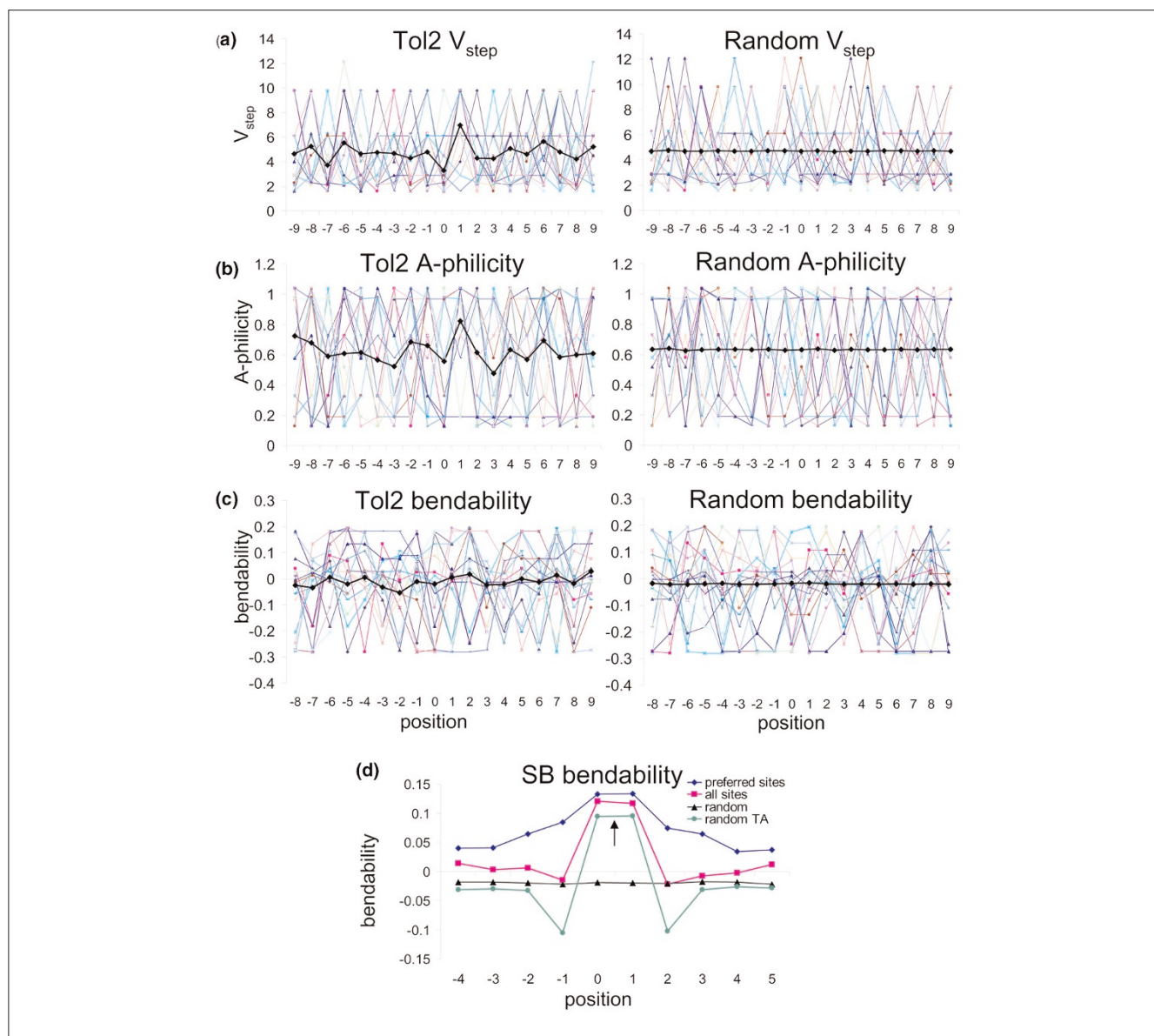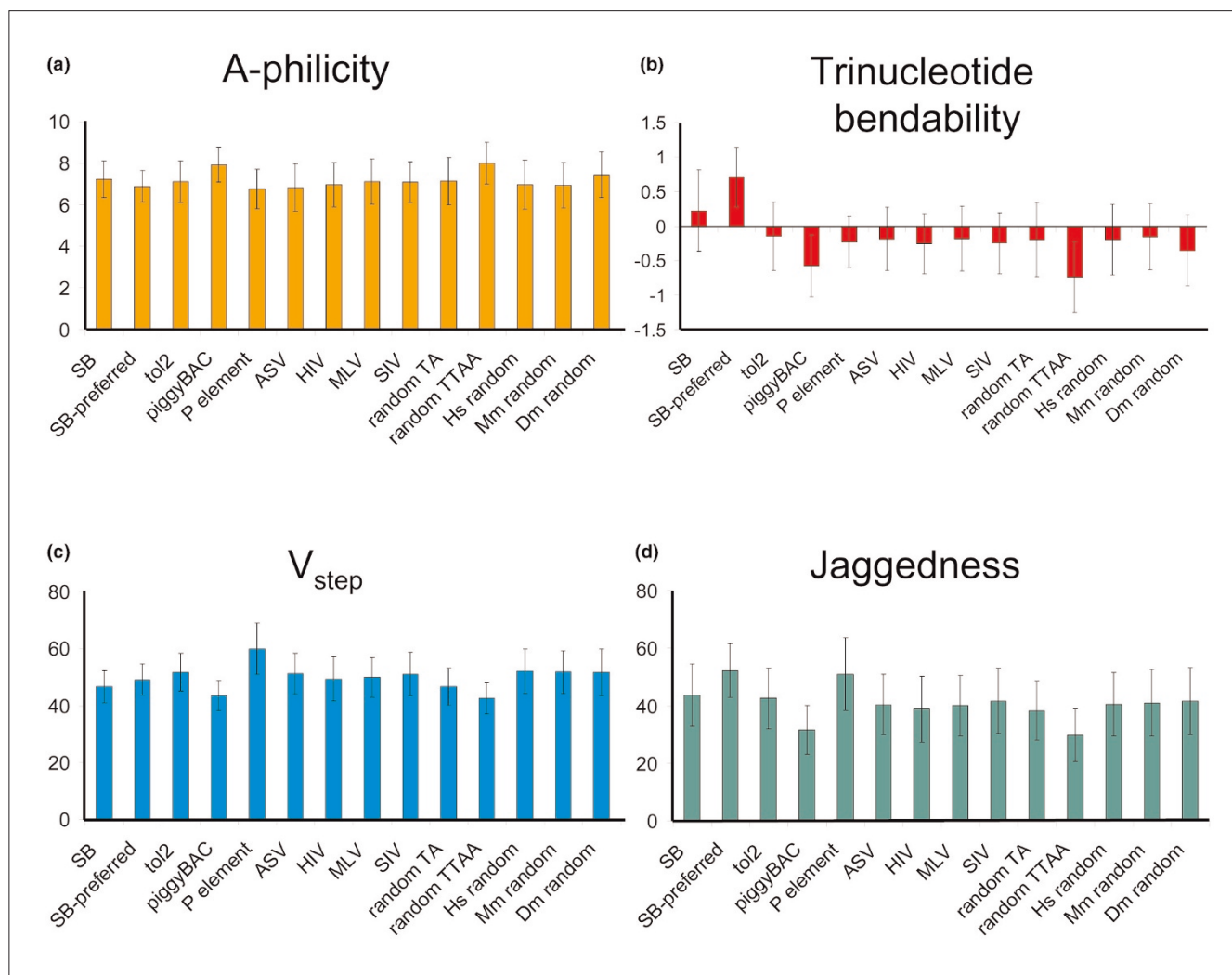
Figure 3

Approaches to identification of DNA structural characteristics governing insertion site preferences for *Tol2* and SB transposons. **(a)** Averaging of all available insertion sites smoothes trends observed in individual plots. Plot of V_{step} profiles of 18 20-base-pair *Tol2* insertions (left, from Balciunas and coworkers [89]) compared with 18 randomly generated sequences (right). Averages are shown by thick black lines. Although individual *Tol2* profiles appear jagged, peaks are not position specific, and so the plot of the average of 36 sites reveals only one small, distinct peak. Individual random sequences also appear jagged, but an average of over 9,000 random sequences is a flat line. **(b)** Analyses of *Tol2* insertion site A-philicity profiles, compared with 18 random sequences. Trends are similar to V_{step} patterns. **(c)** Plot of trinucleotide bendability for *Tol2* and random sites, indicating only small common trends compared with random sequence. The random sequences in panels a to c were acquired from a 10 megabase portion of human chromosome 1p. **(d)** Bendability plots for *Sleeping Beauty* (SB) insertion sites (from Yant and coworkers [106]). The average trinucleotide bendability at each position of 12-base insertion sites is shown for 574 insertions ('all sites'), as well as a subset of 189 insertions classified as 'preferred' based on V_{step} profiles ('preferred sites'). Random TA sites are shown in green, and random sites in black. This plot shows how identification of 'preferred' sites can be useful in distinguishing structural patterns for common insertion sites; preferred sites (based on common patterns of protein-induced deformability in recurrently hit sites) exhibit an overall increase in a separate parameter, DNA bendability, when 'basal' sites are removed.

insufficient to obtain an initial set of preferred sequences. Because nonpreferred sites are likely to vastly outnumber preferred sites in the genome for most vectors, any genome-wide screen will produce a mix of indistinguishable preferred and nonpreferred sites. For example, we have estimated that of the approximately 200,000,000 TA sites in a human genome, only about 10% fall into the preferred category [115], although in the screen conducted by Yant and

Figure 4
Variability in DNA structural characteristics between insertion sites for various vectors. All **(a)** A-philicity, **(b)** trinucleotide bendability, and **(c)** V_{step} values were summed across 12 nucleotides and averaged for all sites of each vector class. **(d)** 'Jaggedness' was measured by taking the absolute value of differences between adjacent V_{step} values, which were then summed and averaged, as in panels a to c. Error bars represent standard deviations. 'SB' indicates 574 *Sleeping Beauty* integrations into human cells identified by Yant and coworkers [106]. 'SB preferred' indicates a subset of 189 sites from the Yant dataset classified as 'preferred' by ProTIS [116]. 'tol2' indicates 63 *Tol2* integrations [89]. 'piggyBac' indicates 297 *piggyBac* insertions deposited into Genbank by Exelexis containing a single TTAA sequence flanked by 10 bases on each side. 'P-element' indicates 920 P-element insertion sites mapped by Liao and coworkers [130]. 'ASV' indicates 357 avian sarcoma leukosis virus (ASLV) insertions into 293T-TVA cells. 'HIV' indicates 334 HIV integrations into SubT1 cells. 'MLV' indicates 695 murine leukemia virus integrations into HeLa cells. 'SIV' indicates 148 simian immunodeficiency virus integrations into CEMx164 cells. All P-element, ASV, HIV, MLV, and SIV sequences were kindly provided by Dr Xioalin Wu. All sites were compared with three sets of over 9,000 randomly selected 12-mers from 10 megabase sections of human chromosome 1 (Hs), mouse chromosome 4 (Mm), and *Drosophila* chromosome 3L (Dm), and 10,000 randomly selected TA and TTAA sites from human chromosome 1.

coworkers [106] 189 out of 573 (33%) genomic SB insertions were classified as preferred sites. Analysis of the bendability of all SB sites mapped in the screen reported by Yant and coworkers shows a peak at the center of the insertion site that is defined by the central TA dinucleotide. However, when only the preferred sites are analyzed, the surrounding nucleotides exhibit a much greater level of bendability (Figure 3d). This effect is in spite of the fact that the preferred sites were identified based on protein-induced

deformability, V_{step}, which is distinct from DNA bendability. The lesson from these studies is that most genome-wide datasets (particularly from experiments involving some form of genetic selection) will probably show a similar dilution effect of preferred sites by greater numbers of nonpreferred sites.

There is a caveat to the analyses discussed up to this point; they all assume that the structures around integration sites

have an absolute center of reference, defined by the site into which the vector integrated. Such analyses could miss structural patterns that are not strictly position specific. For instance, an integrase may have preference for a local region that is highly bendable or deformable, but it may not have a requirement for a particular pattern (or sequence). To account for this, we have examined a parameter called 'jaggedness', which we define as the degree to which V_{step} values alternate from high to low, as in the preferred 'zigzag' sites for SB. We calculated jaggedness by taking the sums of the absolute values of the differences between adjacent V_{step} values across a sequence, so that a jagged/zigzag site would have a higher total value than a flat, basal site, which should have a jaggedness value close to 0. Jaggedness values for several vectors are shown in Figure 4. Although jaggedness values at insertion sites are similar to V_{step} values for most vectors (with the possible exception of *Tol2*), the jaggedness patterns show a high degree of variability across genomic sequences and are somewhat independent of V_{step} patterns (for instance, the c-*myc* gene; Figure 5).

Integration preference versus oncogenic selection

We see two uses for profiling the insertion site preferences for integrating vectors. First, in functional genomics screens, insertion profiles that emerge can be compared with expected profiles that are only structure based rather than genetics based. A striking example of this is evident in the oncogene screens conducted with the SB transposon [58,59], which is illustrated in Figure 6 with respect to the *Braf* gene. Integration sites that emerged from the screen are shown across the entire locus (Figure 6b) and in a selected region comprising exons 10-13/introns 10-12 (Figure 6d), where most of the integrations were selected because of induced expression of a truncated gain-of-function kinase polypeptide. Panels a and c show insertion site preference scores across the region obtained using an automated script (ProTIS) that counts and scores preferred TA dinucleotide insertion sites based on V_{step} values [115]. The results shown in Figure 6 make two strong points. The first is that the frequency of oncogenic insertions in a select region correspond to that predicted on the basis of preference profiling (Figure 6c,d; specifically, microscale structure can be a good predictor of integration site preference). The second is that many predicted hotspots (Figure 6a,b) were not sites that lead to oncogenesis. The combination of these two observations enhances the biologic importance of the integrations into introns 11 and 12.

The second application of predicting profiles of vector insertions may be as part of a risk assessment program. Although current understanding of integration site preferences for most vectors is still inadequate to allow prediction of the probability of integration into specific genes, genome-wide integration datasets may suggest the likelihood that a vector will integrate within the general vicinity of a specific

gene. Similarly, analysis of DNA structural characteristics may be used to assess the likelihood that each vector will integrate within specific regions of genes. For example, although *Braf* can act as a potent oncogene, the pattern of SB integrations into *Braf* suggest that integrations into a relatively small region of the gene (introns 11 and 12) are the most highly selected for oncogenesis, in spite of the presence of hotspots across the entire gene. Thus, the range of possible insertions that are capable of generating an oncogenic transcript, combined with the relative 'attractiveness' of the sequence across these regions, will dictate the chances of insertional activation.

An analysis of several structural characteristics is presented for the mouse c-*myc* gene (Figure 5), the human ortholog of which is activated in many cancers [141]. The figure highlights the 3 kilobase region encompassing the promoter that harbors the bulk of oncogenic retroviral integrations at this locus that have been deposited in the Retroviral-Tagged Cancer Gene Database (RTCGD [142]). The sequence was divided into 50 base pair (bp) bins, and the total values for V_{step}, A-philicity, jaggedness, and bendability were summed across each bin. Measured in 50 bp bins, these structural parameters are highly variable across the sequence, and vary independently from each other. Actual oncogenic retroviral insertions observed in insertional mutagenesis screens and deposited into the RTGCD are shown for comparison in Figure 5a. The profiles indicate two features of transposons under consideration for gene therapy. First, the most likely sites for SB transposons to integrate (Figure 5g) are shifted away from the most commonly found activation sites, as revealed by retroviral integrations (Figure 5a). Second, the profile of TTAA sites, required by the *piggyBac* transposon (Figure 5f), is similar to the preferred SB sites, and further shows that some regions harboring retroviral integrations contain no TTAA sequences, making *piggyBac* insertions into these sites impossible. Thus, at first approximation, it would appear that the transposons are less likely to insert close to the c-*myc* promoter than are retroviral vectors. In support of this, c-*myc* is infrequently hit in SB-based insertional mutagenesis screens; to date, only one c-*myc* integration has been deposited into the RTCGD. In contrast, many retroviral insertions into c-*myc* have been mapped, although the number of deposited retroviral insertions is much higher than the number of transposons.

The relative lack of SB insertions into c-*myc* may be due to either a paucity of favorable SB insertion sites in regions of the gene competent for oncogenic activation, or an overall lack of oncogenic selection for insertions into this gene. In support of the former, transposon-free amplification of c-*myc* was one of the few genomic aberrations observed in tumors harboring mobile transposons (Largaespada DA, Collier LC, Hackett CS, unpublished observations), suggesting that activation of c-*myc* plays a role in the biology of these tumors (there was probably oncogenic selection for the
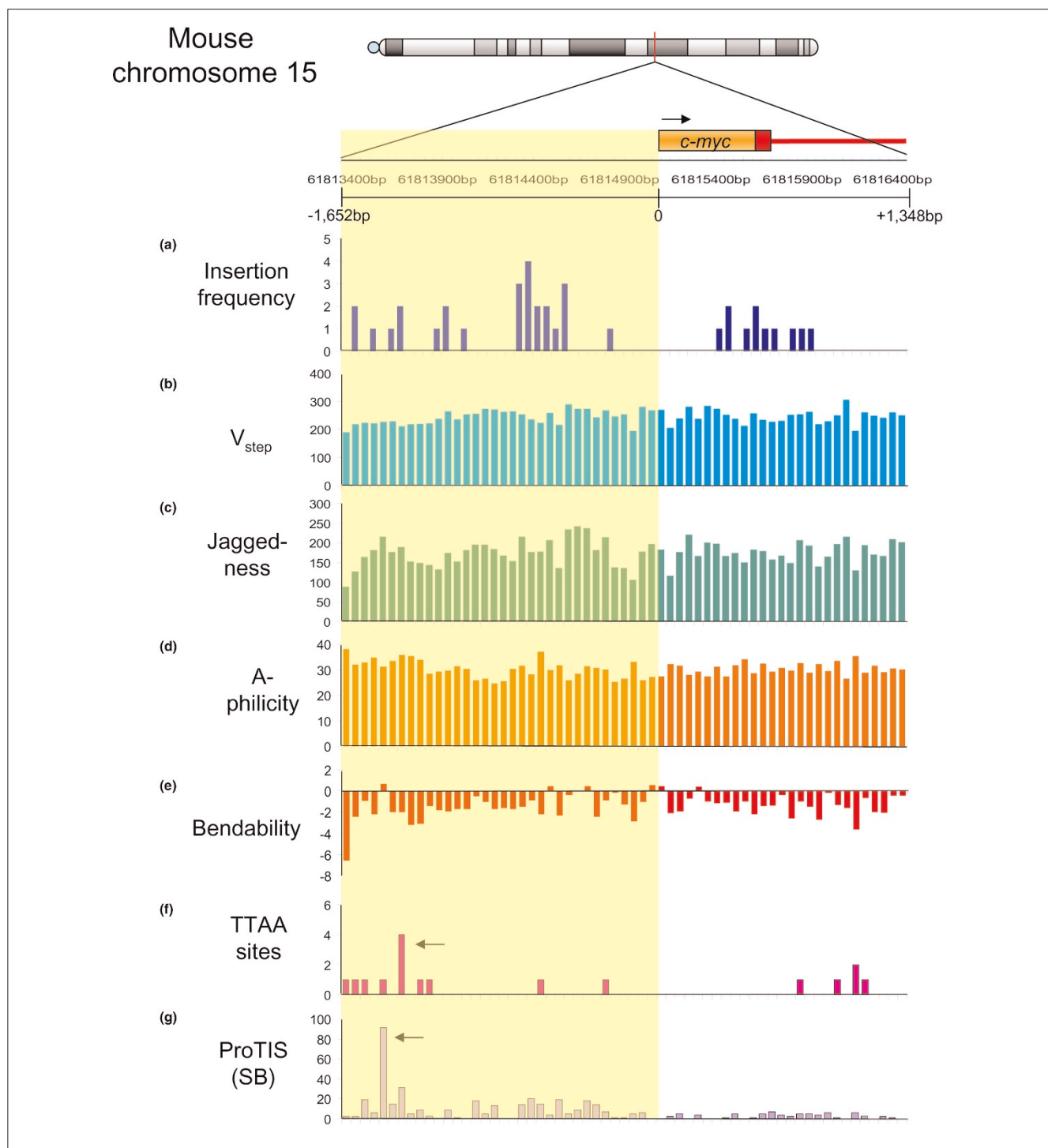
Figure 5
Insertion prediction for transposon vectors surrounding the c-*myc* locus on mouse chromosome 15. A 3 kilobase sequence from the mouse c-*myc* locus (from 61,813,400 to 61,816,400 base pairs) harboring 37 retroviral insertions submitted to the Mouse Retrovirus Tagged Cancer Gene Database [155] is shown. The first exon and intron of c-*myc* are shown in orange; the upstream promoter sequence is shaded in yellow. **(a)** Retrovirus insertion frequency per 50 base pair (bp) segment. Panels (b) to (g) show DNA structural characteristics at 50 bp resolution. **(b)** Total V_{step} for each bin across the region. **(c)** Total V_{step} jaggedness. **(d)** Total A-philicity values. **(e)** Total trinucleotide bendability. **(f)** Number of TTAA sequences per 50 bp bin, representing the total number of possible *piggyBac* insertion sites. Notably, many regions harboring oncogene-selected retroviral insertions have few or no TTAA sequences, suggesting that the likelihood of a *piggyBac* insertion causing an oncogenic event may be lower than that for retroviruses. Arrow represents a potential 'hotspot' for integration, over 1 kilobase upstream of exon 1. **(g)** ProTIS prediction shows a similar, low incidence of preferred SB integration sites. Arrow indicates predicted hotspot for integration over 1 kilobase upstream of exon 1, and slightly upstream of the TTAA hotspot. SB, *Sleeping Beauty*.
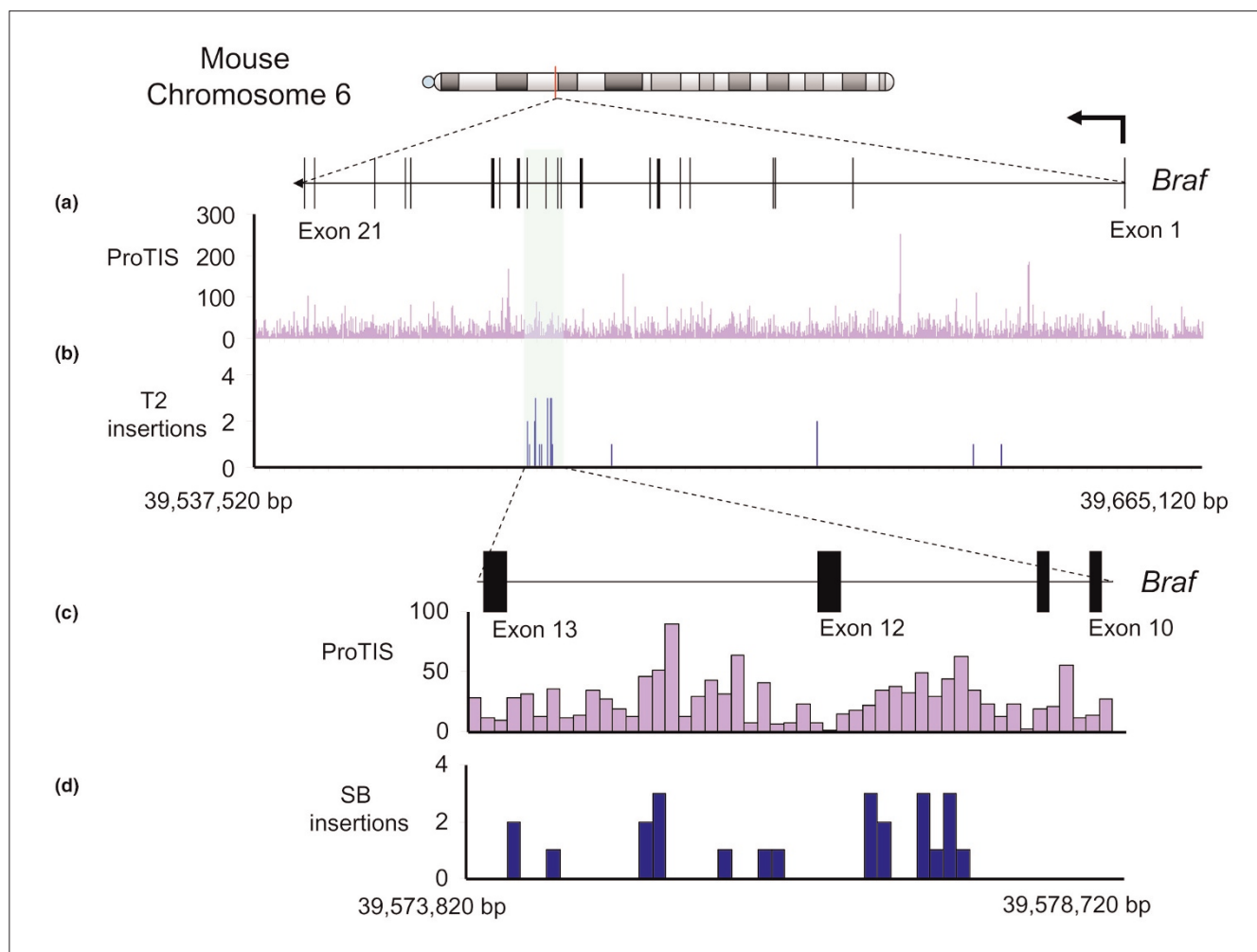
Figure 6
SB insertions across the mouse *Braf* gene. Thirty *Sleeping Beauty* (SB) insertions deposited in the Retroviral-Tagged Cancer Gene Database were mapped across the entire *Braf* transcript and 10 kilobases upstream (NCBI 36 build; note that *Braf* is transcribed right-to-left). Most oncogenic insertions occurred in introns 11 and 12 (formerly annotated as intron 9). **(a)** ProTIS profiling across the entire gene reveals predicted hotspots for SB integration, but **(b)** most actual integrations were found in a relatively low scoring region corresponding to introns 11 and 12. A blowup of this local 4.9 kilobase region demonstrates that **(c)** ProTIS scores closely match **(d)** patterns of actual transposon integration. bp, base pairs.

genomic amplicon). Similar ProTIS analysis of the LMO2 locus revealed the most preferential integration sites for SB transposons that were considerably farther away from the LMO2 promoter than mapped integrations by activating retroviruses [115]. That said, it is evident that prediction of vector integration is not precise and even rare integrations into unfavorable sites have a potential to promote oncogenic expansion, as indicated in Figure 6.

Vector behavior in risk/outcome assessment: lessons from intentional oncogenic insertional mutagenesis

In spite of the inherent behavior of each integrating vector, existing evidence suggests that the oncogenic potential of any given vector can be attenuated depending on how it is used. As with retroviruses, the SB transposon has been used for functional genomics as well as for delivery of therapeutic genes in mouse models of inherited disease. These studies were motivated by two limitations of retroviruses for insertional mutagenesis: the limitation of viruses to infect specific cell types and the tendency of many viral vectors to insert near and activate a possibly limited number of genes [48]. In two recent SB mutagenesis screens, a transgenic concatemer of T2/Onc transposons carried in the germlines of mice was remobilized in somatic cells by a *trans*-acting, transgenic SB transposase. The two screens differed in expression level, domains of expression, and activity of the SB transposase, as well as the copy number of the transposon concatemers [58,59]. An important finding from the two studies was that the oncogenic potential of the same T2/Onc transposon vector, which was engineered specifically

to activate oncogenes and cause cancers in mice, varied between no observable phenotype on one end and rapid development of severe cancer at birth on the other. The oncogenic effect was directly related to the number and types of cells at risk for transposon-induced mutations and perhaps the remobilization rates. The same properties may be relevant for a wide range of other gene therapy vectors.

Coupled with the lack of a preference to integrate near genes, the chances that an SB insertion of a therapeutic gene (in contrast to a genetic cassette designed to wreak havoc on transcriptional units) will activate a neighboring host gene would seem to be lower than for vectors that have an affinity to integrate into genes [65,97]. This feature may be a disadvantage for SB-based functional genomics studies aimed at mutating genes, but it may be advantageous for gene therapy.

Engineering safer vectors

As an alternative to finding vectors that do not target genes, several groups are attempting to target vector integration to a specific region of the genome by generating integrase and SB transposase molecules that are fused to DNA-binding domains that recognize specific DNA sequences [143,144]. It appears that targeting introduces a reduction in activity, without much increase in specificity of integration into specific sites in a mammalian genome [144,145]. This is not surprising if the ability of SB transposase to integrate promiscuously into TA sites is not abridged. There are about 2×10^8 potential TA-dinucleotide SB integration sites into which SB transposons can integrate, of which it is estimated that 2×10^7 are preferred integration sites [115]. Consequently, the chances of a sequence-specific targeting motif added to SB transposase actually guiding transposition to a specific, low-copy target sequence is expected to be extremely low compared with the chances of integrating into any of the millions of other available TA sites. Similarly, to overcome the risk for activation of neighboring genes following vector integration, self-inactivating vectors are being engineered to have diminished ability to activate genes over long distances [146,147], although it is not clear whether these vectors will be safer [148]. The ϕC31 phage integrase system targets relatively few sites in mammalian genomes [119,149], but it appears to introduce a relatively high level of chromosomal recombination [149-151]. Thus, further development of safer vectors remains an open area of investigation.

Conclusion

Ultimately, functional genomics and gene therapy would like to answer the same question for any given vector (while hoping for opposite outcomes) - what are the chances of activating genes? There are four major factors influencing the answer, with each retroviral and transposon having different characteristics for each factor. First, what is the overall tendency of the vector to integrate into genes or promoters? Second, are there adequate local target sites around genes of interest to attract the vector? Third, over what distance can the vector activate a gene? Fourth, to what end can the integration activity be modulated to control the overall likelihood of hitting specific insertion sites close enough for activation of specific genes? Theoretically, knowing each of these variables for every vector would allow researchers to choose the vector with the most utility and lowest risk for the specific purpose intended. In gene therapy, these parameters translate into the risk for hitting a specific oncogene or tumor suppressor gene that could lead to a severe adverse effect. If, in the future, hotspots for integration of SB and other potential gene therapy vectors can be predicted, then we should be able to assess more accurately and modify the various risks for adverse effects from therapeutic vectors. This goal should be within reach in the coming years.

Note added in proof

Since submission of the manuscript, adeno-associated viral vectors (AAV) have been implicated in the induction of hepatocellular carcinomas in mice [152] and in the death of a patient in a clinical trial for treatment of rheumatoid arthritis [153].

Competing interests

PBH owns stock in Discovery Genomics, which is conducting research on the SB transposon system. The other authors declare that they have no competing interests.

References

1. Zambrowicz BP, Friedrich GA, Buxton EC, Lilleberg SL, Person C, Sands SL: **Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells.** *Nature* 1998, **392:**608-611.
2. Mitchell KJ, Pinson KI, Kelly OG, Brennan J, Zupicich J, Scherz P, Leighton PA, Goodrich LV, Lu X, Avery BJ, *et al*: **Functional analysis of secreted and transmembrane proteins critical to mouse development.** *Nat Genet* 2001, **28:**198-200.
3. Mikkers H, Berns A: **Retroviral insertional mutagenesis: tagging cancer pathways.** *Adv Cancer Res* 2003, **88:**53-99.

4. Edelstein ML, Abedi MR, Wixon J, Edelstein RM: **Gene therapy clinical trials worldwide 1989-2004-an overview.** *Gene Med* 2004, **6:**597-602.
5. Sinn PL, Sauter SL, McCray PB Jr: **Gene therapy progress and prospects: Development of improved lentiviral and retroviral vectors: design, biosafety, and production.** *Gene Ther* 2005, **12:**1089-1098.
6. Connelly JB: **Lentiviruses in gene therapy clinical research.** *Gene Ther* 2002, **9:**1730-1743.
7. Jonkers J, Berns A: **Retroviral insertional mutagenesis as a strategy to identify cancer genes.** *Biochim Biophys Acta* 1996, **1287:**29-57.
8. Largaespada DA: **Genetic heterogeneity in acute myeloid leukemia: maximizing information flow from MuLV mutagenesis studies.** *Leukemia* 2000, **14:**1174-1184.
9. Lund AH, Turner G, Trubetskoy A, Verhoeven E, Wientjens E, Hulsman D, Russell R, DePinho RA, Lenz J, van Lohuizen M: **Genome-wide retroviral insertional tagging of genes involved in cancer in Cdkn2a-deficient mice.** *Nat Genet* 2002, **32:**160-165.
10. Suzuki T, Shen H, Akag K, Morse HC, Malley JD, Naiman DQ, Jenkins NA, Copeland NG: **New genes involved in cancer identified by retroviral tagging.** *Nat Genet* 2002, **32:**166-174.
11. Kim R, Trubetskoy A, Suzuki T, Jenkins NA, Copeland NG, Lenz J: **Genome-based identification of cancer genes by proviral tagging in mouse retrovirus-induced T-cell lymphomas.** *J Virol* 2003, **77:**2056-2062.
12. Suzuki T, Minehata K, Akagi K, Jenkins NA, Copeland NG: **Tumor suppressor gene identification using retroviral insertional mutagenesis in *Blm*-deficient mice.** *EMBO J* 2006, **25:**3422-3431.
13. Yi Y, Hahm SH, Lee KH: **Retroviral gene therapy: safety issues and possible solutions.** *Curr Gene Therap* 2005, **5:**25-35.
14. Berns A: **Good news for gene therapy.** *N Engl J Med* 2004, **350:**1679-1680.
15. Pages JC, Bru T: **Toolbox for retrovectorologists.** *J Gene Med* 2004, **6(Suppl 1):**S67-S82.
16. Kohn D, Sadelain M, Dunbar C, Bodine D, Kiem HP, Candotti F, Tisdale J, Riviere I, Blau CA, Richard RE, *et al.*: **American Society of Gene Therapy (ASGT) ad hoc subcommittee on retroviral-mediated gene transfer to hematopoietic stem cells.** *Mol Ther* 2003, **8:**180-187.
17. Hematti P, Hong BK, Ferguson C, Adler R, Hanawa H, Sellers S, Holt IE, Eckfeldt CE, Sharma Y, Schmidt M, *et al.*: **Distinct genomic integration of MLV and SIV vectors in primate hematopoietic stem and progenitor cells.** *PLoS Biol* 2004, **2:**e423.
18. Kiem HP, Sellers S, Thomasson B, Morris JC, Tisdale JF, Horn PA, Hematti P, Adler R, Kuramoto K, Calmels B, *et al.*: **Long-term clinical and molecular follow-up of large animals receiving retrovirally transduced stem and progenitor cells: no progression to clonal hematopoiesis or leukemia.** *Mol Ther* 2004, **9:**389-395.
19. Calmels B, Ferguson C, Laukkanen MO, Adler R, Faulhaber M, Kim HJ, Sellers S, Hematti P, Schmidt M, von Kalle C, *et al.*: **Recurrent retroviral vector integration at the Mds1/Evi1 locus in non-human primate hematopoietic cells.** *Blood* 2005, **106:**2530-2533.
20. Bell P, Wang L, Lebherz C, Flieder DB, Bove MS, Wu D, Gao GP, Wilson JM, Wivel NA: **No evidence for tumorigenesis of AAV vectors in a large-scale study in mice.** *Mol Ther* 2005, **12:**299-306.
21. Themis M, Waddington SN, Schmidt M, von Kalle C, Wang Y, Al-Allaf A, Gregory LG, Nivsarkar M, Themis M, Holder MV, *et al.*: **Oncogenesis following delivery of a non-primate lentiviral gene therapy vector to fetal and neonatal mice.** *Mol Ther* 2005, **12:**763-771.
22. Du Y, Spence SE, Jenkins NA, Copeland NG: **Cooperating cancer-gene identification through oncogenic-retrovirus-induced insertional mutagenesis.** *Blood* 2005, **106:**2498-2505.
23. **Center for Biologics Evaluation and Research: Cellular & Gene Therapy** [http://www.fda.gov/cber/gene.htm]
24. Peng S: **Current status of gendicine in China: recombinant human Ad-p53 agent for treatment of cancers.** *Hum Gene Ther* 2005, **16:**1016-1027.
25. Hacein-Bey-Abina S, Le Deist F, Carlier F, Bouneaud C, Hue C, De Villartay JP, Thrasher AJ, Wulffraat N, Sorensen R, Dupuis-Girod S, *et al.*: **Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy.** *N Eng J Med* 2002, **346:**1185-1193.

26. Hacein-Bey-Abina S, von Kalle C, Schmidt M, Le Deist F, Wulffraat N, McIntyre E, Radford I, Villeval JL, Fraser CC, Cavazzana-Calvo M, *et al.*: **A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency.** *N Engl J Med* 2003, **348:**255-256.
27. Baum C, Dullmann J, Li Z, Fehse B, Meyer J, Williams DA, von Kalle C: **Side effects of retroviral gene transfer into hematopoietic stem cells.** *Blood* 2003, **101:**2099-2114.
28. Baum C, Fehse B: **Mutagenesis by retroviral transgene insertion: risk assessment and potential alternatives.** *Curr Opin Mol Ther* 2003, **5:**458-462.
29. Baum C, von Kalle C, Staal FJ, Li Z, Fehse B, Schmidt M, Weerkamp F, Karlsson S, Wagemaker G, Williams DA: **Chance or necessity? Insertional mutagenesis in gene therapy and its consequences.** *Mol Ther* 2004, **9:**5-13.
30. Kustikova O, Fehse B, Modlich U, Yang M, Dullmann J, Kamino K, von Neuhoff N, Schlegelberger B, Li Z, Baum C: **Clonal dominance of hematopoietic stem cells triggered by retroviral gene marking.** *Science* 2005, **308:**1171-1174.
31. Hacein-Bey-Abina S, Von Kalle C, Schmidt M, McCormack MP, Wulffraat N, Leboulch P, Lim A, Osborne CS, Pawliuk R, Morillon E, *et al.*: **LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1.** *Science* 2003, **302:**415-419.
32. McCormack MP, Forster A, Drynan L, Pannell R, Rabbitts TH: **The LMO2 T-cell oncogene is activated via chromosomal translocations or retroviral insertion during gene therapy but has no mandatory role in normal T-cell development.** *Mol Cell Biol* 2003, **23:**9003-9013.
33. Dave' UP, Jenkins NA, Copeland NG: **Gene therapy insertional mutagenesis insights.** *Science* 2004, **303:**33.
34. McCormack MP, Rabbitts TH: **Activation of the T-cell oncogene LMO2 after gene therapy for X-linked severe combined immunodeficiency.** *N Engl J Med* 2004, **350:**913-922.
35. Nam CH, Rabbitts TH: **The role of LMO2 in development and in T cell leukemia after chromosomal translocation or retroviral insertion.** *Mol Ther* 2006, **13:**15-25.
36. Woods NB, Bottero V, Schmidt M, von Kalle C, Verma IM: **Gene therapy: therapeutic gene causing lymphoma.** *Nature* 2006, **440:**1123.
37. Pike-Overzet K, de Ridder D, Weerkamp F, Baert MR, Verstegen MM, Brugman MH, Howe SJ, Reinders MJ, Thrasher AJ, Wagemaker G, *et al.*: **Gene therapy: is IL2RG oncogenic in T-cell development?** *Nature* 2006, **443:**E5.
38. Thrasher AJ, Gaspar HB, Baum C, Modlich U, Schambach A, Candotti F, Otsu M, Sorrentino B, Scobie L, Cameron E, *et al.*: **Gene therapy: X-SCID transgene leukaemogenicity.** *Nature* 2006, **443:**E5.
39. Schmidt M, Carbonaro DA, Speckmann C, Wissler M, Bohnsack J, Elder M, Aronow BJ, Nolta JA, Kohn DB, von Kalle C: **Clonality analysis after retroviral-mediated gene transfer to CD34⁺ cells from the cord blood of ADA-deficient SCID neonates.** *Nat Med* 2003, **9:**463-468.
40. Aiuti A, Ficara F, Cattaneo F, Bordignon C, Roncarolo MG: **Gene therapy for adenosine deaminase deficiency.** *Curr Opin Allergy Clin Immunol* 2004, **3:**461-466.
41. Gaspar HB, Bjorkegren E, Parsley K, Gilmour KC, King D, Sinclair J, Zhang F, Giannakopoulos A, Adams S, Fairbanks LD, *et al.*: **Successful reconstitution of immunity in ADA-SCID by stem cell gene therapy following cessation of PEG-ADA and use of mild preconditioning.** *Mol Ther* 2006, **14:**505-513.
42. Ott MG, Schmidt M, Schwarzwaelder K, Stein S, Siler U, Koehl U, Glimm H, Kuhlcke K, Schilz A, Kunkel H, *et al.*: **Correction of X-linked chronic granulomatous disease by gene therapy, augmented by insertional activation of MDS1-EVI1, PRDM16 or SETBP1.** *Nature Med* 2006, **5:**401-409.
43. Gaspar HB, Parsley KL, Howe S, King D, Gilmour KC, Sinclair J, Brouns G, Schmidt M, Von Kalle C, Barington T, *et al.*: **Gene therapy of X-linked severe combined immunodeficiency by use of a pseudotyped gammaretroviral vector.** *Lancet* 2005, **364:**2181-2187.
44. Buonamici S, Chakraborty S, Senyuk V, Nucifora G: **The role of EVI1 in normal and leukemic cells.** *Blood Cells Mol Dis* 2003, **31:**206-212.
45. Buonamici S, Li D, Chi Y, Zhao R, Wang X, Brace L, Ni H, Sauntharararajah Y, Nucifora G: **EVI1 induces myelodysplastic syndrome in mice.** *J Clin Invest* 2004, **114:**713-719.

46. Li Z, Dullmann J, Schiendlmeier B, Schmidt M, von Kalle C, Meyer J, Forster M, Stocking C, Wahlers A, Frank O, *et al.*: **Murine leukemia induced by retroviral gene marking.** *Science* 2002, **296:**497.

47. Mikkers H, Allen J, Knipscheer P, Romeijn L, Hart A, Vink E, Berns A: **High-throughput retroviral tagging to identify components of specific signaling pathways in cancer.** *Nature Genet* 2002, **32:**153-159.

48. Wu X, Luke BT, Burgess SM: **Redefining the common insertion site.** *Virol* 2006, **344:**292-295.

49. Morgan RA, Dudley ME, Wunderlich JR, Hughes MS, Yang JC, Sherry RM, Royal RE, Topalian SL, Kammula US, Restifo NP, *et al.*: **Cancer regression in patients after transfer of genetically engineered lymphocytes.** *Science* 2006, **314:**126-129.

50. Ivics Z, Izsvak Z: **Transposable elements for transgenesis and insertional mutagenesis in vertebrates: a contemporary review of experimental strategies.** *Meth Mol Biol* 2004, **260:**255-276.

51. Hackett PB, Ekker SC, Largaespada DA, Mclvor RS: *Sleeping Beauty* **transposon-mediated gene therapy for prolonged expression**. *Adv Genet* 2005, **54:**187-229.

52. Hackett PB, Ekker SE, Essner JJ: **Applications of transposable elements in fish for transgenesis and functional genomics.** In *Fish Development and Genetics*. Edited by Gong Z, Korzh V. Hackensack, NJ, USA: World Scientific, Inc; 2004:532-580.

53. Ivics Z, Izsvak Z: **Transposons for gene therapy!** *Curr Gene Ther* 2006, **6:**593-607.

54. 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:**501-510.

55. Dupuy AJ, Fritz S, Largaespada DA: **Transposition and gene disruption using a mutagenic transposon vector in the male germline of the mouse.** *Genesis* 2001, **30:**82-88.

56. Davidson AE, Balciunas D, Mohn D, Shaffer J, Hermanson S, Sivasubbu S, Hackett PB, Ekker SC: **Efficient gene delivery and expression in zebrafish using** *Sleeping Beauty***.** *Dev Biol* 2003, **263:**191-202.

57. Balciunas D, Davidson AE, Sivasubbu S, Hermanson SB, Welle Z, Ekker SC: **Enhancer trapping in zebrafish using the** *Sleeping Beauty* **transposon.** *BMC Genomics* 2004, **5:**62.

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

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

60. Keng VW, Yae K, Hayakawa T, Mizuno S, Uno Y, Yusa K, Kokubu C, Kinoshita T, Akagi K, Jenkins NA, *et al.*: **Region-specific saturation germline mutagenesis in mice using the** *Sleeping Beauty* **transposon system.** *Nat Methods* 2005, **2:**763-769.

61. Carlson CM, Largaespada DA: **Insertional mutagenesis in mice: new perspectives and tools.** *Nat Rev Genet* 2005, **6:**568-580.

62. Nakai H, Montini E, Fuess S, Storm TA, Grompe M, Kay MA: **AAV serotype 2 vectors preferentially integrate into active genes in mice.** *Nat Genet* 2003, **34:**297-302.

63. Wu X, Burgess SM: **Integration target site selection for retroviruses and transposable elements.** *Cell Mol Life Sci* 2004, **61:**2588-2596.

64. Wu X, Li Y, Crise B, Burgess SM: **Transcription start regions in human genome are favored targets for MLV integration.** *Science* 2003, **300:**1749-1751.

65. Bushman F, Lewinski M, Ciuffi A, Barr S, Leipzig J, Hannenhalli S, Hoffmann C: **Genome-wide analysis of retroviral DNA integration.** *Nat Rev Microbiol* 2005, **3:**848-858.

66. Kay MA, Glorioso JC, Naldini L: **Viral vectors for gene therapy: the art of turning infectious agents into vehicles of therapeutics.** *Nat Med* 2001, **7:**33-40.

67. Muruve DA, Barnes MJ, Stillman IE, Libermann TA: **Adenoviral gene therapy leads to rapid induction of multiple chemokines and acute neutrophil-dependent hepatic injury in vivo.** *Hum Gene Ther* 1999, **10:**965-976.

68. Graham A, Walker R, Baird P, Hahn CN, Fazakerley JK: **CNS gene therapy applications of the semliki forest virus 1 vector are limited by neurotoxicity.** *Mol Ther* 2006, **13:**631-635.

69. Reeves L, Smucker P, Cornetta K: **Packaging cell line characteristics and optimizing retroviral vector titer: The National**

Gene Vector Laboratory experience. *Hum Gene Ther* 2000, **11:**2093-2103.

70. Kumar M, Bradow BP, Zimmerberg J: **Large-scale production of pseudotyped lentiviral vectors using baculovirus GP64.** *Hum Gene Ther* 2003, **14:**67-77.

71. Wagner E, Culmsee C, Boeckle S: **Targeting polyplexes: toward synthetic virus vector systems.** *Adv Genet* 2005, **53:**333-354.

72. Putnam D: **Polymers for gene delivery across length scales.** *Nat Mater* 2006, **5:**439-451.

73. Merdan T, Kunath K, Petersen H, Bakowsky U, Voigt KH, Kopecek J, Kissel T: **PEGylation of poly(ethylene imine) affects stability of complexes with plasmid DNA under in vivo conditions in a dose-dependent manner after intravenous injection into mice.** *Bioconjugate Chem* 2005, **16:**785-792.

74. Liu F, Song Y, Liu D: **Hydrodynamics-based transfection in animals by systemic administration of plasmid DNA.** *Gene Ther* 1999, **6:**1258-1266.

75. Zhang G, Budker V, Wolff JA: **High levels of foreign gene expression in hepatocytes after tail vein injections of naked plasmid DNA.** *Hum Gene Ther* 1999, **10:**1735-1737.

76. Suda T, Gao X, Stolz DB, Liu D: **Structural impact of hydrodynamic injection on mouse liver.** *Gene Ther* 2007, **14:**129-137.

77. Yoshino H, Hashizume K, Kobayashi E: **Naked plasmid DNA transfer to the porcine liver using rapid injection with large volume.** *Gene Ther* 2006, **13:**1696-1702.

78. Herweijer H, Wolff JA: **Gene therapy progress and prospects: Hydrodynamic gene delivery.** *Gene Ther* 2007, **14:**99-107.

79. Kichler A: **Gene transfer with modified polyethylenimines.** *J Gene Med* 2004, **6:**S3-S10.

80. Demeneix B, Behr JP: **Polyethylenimine (PEI).** *Adv Genet* 2005, **53:**217-230.

81. Neu M, Fischer D, Kissel T: **Recent advances in rational gene transfer vector design based on poly(ethylene imine) and its derivatives.** *J Gene Med* 2005, **7:**992-1009.

82. Breunig M, Lungwitz U, Liebl R, Fontanari C, Klar J, Kurtz A, Blunk T, Goepferich A: **Gene delivery with low molecular weight linear polyethylenimines.** *J Gene Med* 2005, **7:**1287-1298.

83. Huang X, Wilber AC, Bao L, Tuong D, Tolar J, Orchard PJ, Levine D, June CH, Mclvor RS, Blazar BL, Zhou X: **Stable gene transfer and expression in human primary T-cells by the** *Sleeping Beauty* **transposon system.** *Blood* 2005, **107:**483-491.

84. Yant SR, Meuse L, Chiu W, Ivics Z, Izsvak Z, Kay MA: **Somatic integration and long-term transgene expression in normal and haemophilic mice using a DNA transposon system.** *Nat Genet* 2000, **25:**35-41.

85. Ohlfest JR, Frandsen JL, Fritz S, Lobitz PD, Perkinson SG, Clark KJ, Nelsestuen G, Key NS, Mclvor RS, Hackett PB, *et al.*: **Phenotypic correction and long-term expression of factor VIII in hemophilic mice by immunotolerization and nonviral gene transfer using the** *Sleeping Beauty* **transposon system.** *Blood* 2005, **105:**2691-2698.

86. Baus J, Liu L, Heggestad AD, Sanz S, Fletcher BS: **Correction of murine hemophilia a by hematopoietic stem cell gene therapy.** *Mol Ther* 2005, **12:**1034-1042.

87. Liu L, Mah C, Fletcher BS: **Sustained FVIII expression and phenotypic correction of hemophilia A in neonatal mice.** *Mol Ther* 2006, **13:**1006-1015.

88. Montini EP, Held PK, Noll M, Morcinek N, Al-Dhalimy M, Finegold M, Yant SR, Kay MA, Grompe M: **In vivo correction of murine tyrosinemia type I by DNA-mediated transposition.** *Mol Ther* 2002, **6:**759-769.

89. Balciunas D, Wagensteen KJ, Wilber AC, Bell JB, Geurts AM, Sivasubbu S, Wang X, Hackett PB, Largaespada DA, Mclvor RS, *et al.*: **Harnessing an efficient large cargo-capacity transposon for vertebrate gene transfer applications.** *PLoS Genet* 2006, **4:**e169.

90. Ortiz S, Lin Q, Yant SR, Keene D, Kay MA, Khavari PA: **Sustainable correction of junctional epidermollysis bullosa via transposon-mediated nonviral gene transfer.** *Gene Ther* 2003, **10:**1099-1104.

91. Ohlfest JR, Lobitz PD, Perkinson SG, Largaespada DA: **Integration and long-term expression in xenografted human glioblastoma cells using a plasmid-based transposon system.** *Mol Ther* 2004, **10:**260-268.

92. Ohlfest JR, Demorest ZL, Motooka Y, Vengco I, Oh S, Chen E, Scappaticci FA, Saplis RJ, Ekker SC, Low WC, *et al.*: **Combinatorial anti-angiogenic gene therapy by nonviral gene transfer using the** *Sleeping Beauty* **transposon causes tumor regression and**

improves survival in mice bearing intracranial human glioblastoma. *Mol Ther* 2005, **12**:778-788.

93. Chen ZT, Kren BT, Wong PYP, Low WC, Steer CJ: *Sleeping Beauty*-mediated down-regulation of huntingtin expression by RNA interference. *Biochem Biophys Res Commun* 2005, **329**:646-652.

94. Liu H, Liu L, Fletcher BS, Visner GA: *Sleeping Beauty*-based gene therapy with indoleamine 2,3-dioxygenase inhibits lung allograft fibrosis. *FASEB J* 2006, **20**:2384-2386.

95. Aronovich EL, Bell JB, Belur LR, Gunther R, Koniar B, Erickson DC, Schachern PA, Matise I, McIvor RS, Whitley CB, *et al.*: *Sleeping Beauty* transposon-mediated gene therapy in the murine models of mucopolysaccharidoses (MPS) Type I and MPS Type VII. *J Gene Med* 2007, **9**:403-415.

96. Miskey C, Izsvak Z, Plasterk RHA, Ivics Z: **The Frog Prince: a reconstructed transposon from *Rana pipiens* with high transpositional activity in vertebrate cells.** *Nucl Acids Res* 2003, **31**:6873-6881.

97. 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**:473-483.

98. Rio DC, Barnes G, Laski FA, Rine J, Rubin GM: **Evidence for *Drosophila* P element transposase activity in mammalian cells and yeast.** *J Mol Biol* 1988, **200**:411-415.

99. Berry C, Hannenhalli S, Leipzig J, Bushman FD: **Selection of target sites for mobile DNA integration in the human genome.** *PLoS Comp Biol* 2006, **2**:e157.

100. Ciuffi A, Mitchell RS, Hoffmann C, Leipzig J, Shinn P, Ecker JR, Bushman FD: **Integration site selection by HIV-based vectors in dividing and growth-arrested IMR-90 lung fibroblasts.** *Mol Ther* 2006, **13**:366-373.

101. Ciuffi A, Diamond TL, Hwang Y, Marshall HM, Bushman FD: **Modulating target site selection during human immunodeficiency virus DNA integration in vitro with an engineered tethering factor.** *Hum Gene Ther* 2006, **17**:960-967.

102. Lewinski MK, Yamashita M, Emerman M, Ciuffi A, Marshall H, Crawford G, Collins F, Shinn P, Leipzig J, Hannenhalli S, *et al.*: **Retroviral DNA integration: viral and cellular determinants of target-site selection.** *PLoS Pathog* 2006, **2**:e60.

103. Bushman FD: **Tethering human immunodeficiency virus 1 integrase to a DNA site directs integration to nearby sequences.** *Proc Natl Acad Sci USA* 1994, **91**:9233-9237.

104. Bushman FD, Miller MD: **Tethering human immunodeficiency virus type 1 preintegration complexes to target DNA promotes integration at nearby sites.** *J Virol* 1997, **71**:458-464.

105. Zhu Y, Dai J, Fuerst PG, Voytas DF: **Controlling integration specificity of a yeast retrotransposon.** *Proc Natl Acad Sci USA* 2003, **100**:5891-5895.

106. Yant SR, Wu X, Huang Y, Garrison B, Burgess SM, Kay MA: **High-resolution genome-wide mapping of transposon integration in mammals.** *Mol Cell Biol* 2005, **25**:2085-2094.

107. Eddy SR: **Non-coding RNA genes and the modern RNA world.** *Nat Rev Genet* 2006, **2**:919-925.

108. Mitchell RS, Beitzel BF, Schroder AR, Shinn P, Chen H, Berry CC, Ecker JR, Bushman FD: **Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences.** *PLoS* 2004, **2**:1127-1136.

109. Laufs S, Nagy KZ, Giordano F, Hotz-Wagenblatt A, Zeller WJ, Fruehauf S: **Insertion of retroviral vectors in NOD/SCID repopulating human peripheral blood progenitor cells occurs preferentially in the vicinity of transcription start regions and in introns.** *Mol Ther* 2004, **10**:874-881.

110. De Palma M, Montini E, de Sio FR, Benedicenti F, Gentile A, Medico E, Naldini L: **Promoter trapping reveals significant differences in integration site selection between MLV and HIV vectors in primary hematopoietic cells.** *Blood* 2005, **105**:2307-2315.

111. Holman AG, Coffin JM: **Symmetrical base preferences surrounding HIV-1, avian sarcoma/leukosis virus, and murine leukemia virus integration sites.** *Proc Natl Acad Sci USA* 2005, **102**:6103-6107.

112. Schroder ARW, 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**:521-529.

113. Nakai H, Wu X, Fuess S, Storm TA, Munroe D, Montini E, Burgess SM, Grompe M, Kay MA: **Large-scale molecular characterization of adeno-associated virus vector integration in mouse liver.** *J Virol* 2005, **79**:3606-3614.

114. Lewinski MK, Bisgrove D, Shinn P, Chen H, Hoffmann C, Hannenhalli S, Verdin E, Berry CC, Ecker JR, Bushman FD: **Genome-wide analy-** sis of chromosomal features repressing human immunodeficiency virus transcription. *J Virol* 2005, **79**:6610-6619.

115. Geurts AM, Hackett CS, Bell JB, Bergemann TM, Carlson CM, Collier LS, Largaespada DA, Hackett PB: **DNA structural patterns influence integration site preferences for mobile elements.** *Nucl Acids Res* 2006, **34**:2803-2811.

116. Geurts AM, Yang Y, Clark KJ, Cui Z, Dupuy AJ, Largaespada DA, Hackett PB: **Gene transfer into genomes of human cells by the *Sleeping Beauty* transposon system.** *Mol Therap* 2003, **8**:108-117.

117. Segal E, Fondufe-Mittendorf Y, Chen L, Thastrom A, Field Y, Moore IK, Wang JP, Widom J: **A genomic code for nucleosome positioning.** *Nature* 2005, **442**:772-778.

118. Pryciak PM, Muller HP, Varmus HE: **Simian virus 40 minichromosomes as targets for retroviral integration in vivo.** *Proc Natl Acad Sci USA* 1992, **89**:9237-9241.

119. Groth AC, Olivares EC, Thyagarajan B, Calos MP: **A phage integrase directs efficient site-specific integration in human cells.** *Proc Nat Acad Sci USA* 2000, **97**:5995-6000.

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

121. Olivares EC, Hollis RP, Chalberg TW, Meuse L, Kay MA, Calos MP: **Site-specific genomic integration produces therapeutic Factor IX levels in mice.** *Nat Biotechnol* 2002, **20**:1124-1128.

122. Chen L, Woo SLC: **Complete and persistent phenotypic correction of phenylketonuria in mice by site-specific genome integration of murine phenylalanine hydroxylase cDNA.** *Proc Natl Acad Sci USA* 2005, **102**:15581-15586.

123. Kuduvalli PN, Mitra R, Craig NL: **Site-specific Tn7 transposition into the human genome.** *Nucl Acids Res* 2005, **33**:857-863.

124. Vigdal TJ, Kaufman CD, Izsvak Z, Voytas DF, Ivics Z: **Common physical properties of DNA affecting target site selection of *Sleeping Beauty* and other Tc1/mariner transposable elements.** *J Mol Biol* 2002, **323**:411-452.

125. Carlson CM, Dupuy AJ, Fritz S, Roberg-Perez KJ, Fletcher CF, Largaespada DA: **Transposon mutagenesis of the mouse germline.** *Genetics* 2003, **165**:243-256.

126. Wu X, Li Y, Crise B, Burgess SM, Munroe DJ: **Weak palindromic consensus sequences are a common feature found at the integration target sites of many retroviruses.** *J Virol* 2005, **79**:5211-5214.

127. Grandgennet DP: **Symmetrical recognition of cellular DNA target sequences during retroviral integration.** *Proc Nat Acad Sci USA* 2005, **102**:5903-5904.

128. Pryciak PM, Sil A, Varmus HE: **Retroviral integration into minichromosomes in vitro.** *EMBO J* 1992, **11**:291-303.

129. Muller HP, Varmus HE: **DNA-bending creates favored sites for retroviral integration: an explanation for preferred insertion sites in nucleosomes.** *EMBO J* 1994, **13**:4704-4714.

130. Liao GC, Rehm EJ, Rubin GM: **Insertion site preferences of the P transposable element in *Drosophila melanogaster*.** *Proc Natl Acad Sci USA* 2000, **97**:3347-3351.

131. Liu G, Geurts AM, Yae K, Srinivassan AR, Fahrenkrug SC, Largaespada DA, Takeda J, Horie K, Olson WK, Hackett PB: **Target-site preference for *Sleeping Beauty* transposons.** *J Mol Biol* 2005, **346**:161-173.

132. Posey JE, Pytlos MJ, Sinden RR, Roth DB: **Target DNA structure plays a critical role in RAG transposition.** *PLoS Biol* 2006, **4**:e350.

133. Gorin AA, Zhurkin VB, Olson WK: *B*-DNA twisting correlates with base-pair morphology. *J Mol Biol* 1995, **247**:34-48.

134. Ivanov VI, Minchenkova LE, Chernov BK, McPhie P, Ryu S, Garges S, Barber AM, Zhurkin VB, Adhya S: **CRP-DNA complexes: inducing the *A*-like form in the binding sites with an extended central spacer.** *J Mol Biol* 1995, **245**:228-240.

135. Lu XJ, Shakked Z, Olson WK: *A*-form conformational motifs in ligand-bound DNA structures. *J Mol Biol* 2000, **300**:819-840.

136. Olson WK, Bansal M, Burley SK, Dickerson RE, Gerstein M, Harvey SC, Heinemann U, Lu XJ, Neidle S, Shakked Z, *et al.*: **A standard reference frame for the description of nucleic acid base-pair geometry.** *J Mol Biol* 2001, **313**:229-237.

137. Olson WK, Gorin AA, Lu XJ, Hock LM, Zhurkin VB: **DNA sequence-dependent deformability deduced from protein-DNA crystal complexes.** *Proc Natl Acad Sci USA* 1998, **95**:11163-11168.

138. Olson WK, Zhurkin VB: **Modeling DNA deformations.** *Curr Opin Struct Biol* 2000, **10**:286-297.

139. Brukner I, Sanchez R, Suck D, Pongor S: **Sequence-dependent bending propensity of DNA as revealed by DNase I: parameters for trinucleotides.** *EMBO J* 1995, **14:**1812-1818.

140. Mandal PK, Rawal K, Ramaswamy R, Bhattacharya A, Bhattacharya S: **Identification of insertion hot spots for non-LTR retrotransposons: computational and biochemical application to *Entamoeba histolytica*.** *Nucl Acids Res* 2006, **34:**5752-5763.

141. Nesbit CE, Tersak JM, Prochownik EV: **MYC oncogenes and human neoplastic disease.** *Oncogene* 1999, **18:**3004-3006.

142. **Retrovirus Tagged Cancer Gene Database** [http://rtcgd.abcc.ncifcrf.gov/]

143. Maragathavally KJ, Kaminski JM, Coates CJ: **Chimeric *Mos1* and *piggyBac* transposases result in site-directed integration.** *FASEB J* 2006, **20:**1880-1882.

144. Yant SR, Huang Y, Akache B, Kay MA: **Fusion proteins consisting of the *Sleeping Beauty* transposase and the polydactyl zinc finger protein hE2C direct transposon integration into a unique human chromosomal sequence.** *Nucleic Acids Res* 2007:in press.

145. Ivics Z, Katzer A, Stuwe EE, Fiedler D, Knespel S, Izsvak Z**: Targeted *Sleeping Beauty* transposition in human cells.** *Mol Ther* 2007, **15:**1137-1144.

146. CPMP: **Insertional mutagenesis and oncogenesis: update from non-clinical and clinical studies. Gene Therapy Expert Group of the Committee for Proprietary Medical Products (CPMP).** *J Gene Med* 2004, **6:**127-129.

147. Levine BL, Humeau LM, Boyer J, Macgregor RR, Rebello T, Lu, X., Binder GK, Slepushkin V, Lemiale F, Mascola JR, Bushman FD, *et al.*: **Gene transfer in humans using a conditionally replicating lentiviral vector.** *Proc Natl Acad Sci USA* 2006, **103:**17372-17377.

148. Buchholz CJ, Cichutek K: **Is it going to be SIN?** *J Gene Med* 2006, **8:**1274-1276.

149. Chalberg TW, Portlock JL, Olivares EC, Thyagarajan B, Kirby PJ, Hillman RT, Hoelters J, Calos MP: **Integration specificity of phage phiC31 integrase in the human genome.** *J Mol Biol* 2006, **357:**28-48.

150. Liu J, Jeppesen I, Nielsen K, Jensen TG: **phiC31 integrase induces chromosomal aberrations in primary human fibroblasts.** *Gene Ther* 2006, **13:**1188-1190.

151. Ehrhardt A, Engler JA, Xu H, Kay MA: **Molecular analysis of chromosomal rearrangements in mammalian cells after phiC31-mediated integration.** *Hum Gene Ther* 2006, **17:**1077-1094.

152. Donsante A, Miller DG, Li Y, Vogler C, Brunt EM, Russell DW, Sands MS. **AAV vector integration sites in mouse hepatocellular carcinoma.** *Science* 2007, **317:**477.

153. Kaiser J. **Clinical research. Death prompts a review of gene therapy vector.** *Science* 2007, **317:**580.

154. Collier LS, Largaespada DA: **Hopping around the tumor genome: transposons for cancer gene discovery.** *Cancer Res* 2005, **65:**9607-9610.

155. **Mouse Retrovirus Tagged Cancer Gene Database** [http://rtcgd.abcc.ncifcrf.gov/].