Mobile Genetic Elements 4:6, 1–6; November/December 2014; Published with license by Taylor & Francis Group, LLC

Applications of DNA integrating elements: Facing the bias bully

Johann de Jong[1], Lodewyk F A Wessels[1,2], Maarten van Lohuizen[3], Jeroen de Ridder[2,*], and Waseem Akhtar[3,*]

[1]Computational Cancer Biology Group; Division of Molecular Carcinogenesis; The Netherlands Cancer Institute; Amsterdam, The Netherlands; [2]Delft Bioinformatics Lab; TU Delft; Delft, The Netherlands; [3]Division of Molecular Genetics; The Netherlands Cancer Institute; Amsterdam, The Netherlands

Keywords: chromatin position effect, epigenomics, gene therapy, insertional mutagenesis, integration bias

Retroviruses and DNA transposons are an important part of molecular biologists' toolbox. The applications of these elements range from functional genomics to oncogene discovery and gene therapy. However, these elements do not integrate uniformly across the genome, which is an important limitation to their use. A number of genetic and epigenetic factors have been shown to shape the integration preference of these elements. Insight into integration bias can significantly enhance the analysis and interpretation of results obtained using these elements. For three different applications, we outline how bias can affect results, and can potentially be addressed.

Introduction

DNA integrating elements are parasitic nucleic acids capable of integrating their DNA into the host genome. These elements can be divided into 2 broad categories, viruses and transposons. The integration process is guided by dedicated sequences at the flanks of these elements called terminal repeats. The rest of the sequence is not required for the integration process and can therefore be replaced by genetic material of interest using molecular engineering techniques. In this way, these elements serve as vectors for the delivery of specialized genetic cargo into the cells of interest. This feature of transposons and viruses has made them ideal tools for studying the function of different genetic components such as genes, promoters and enhancers, by integrating these components into the genome and studying their function in the cellular context. DNA integrating elements can also be used to add a new function or restore a defective function in the cell. This provides the basis for their extensive use in gene therapy.[1] The viruses have strong transcriptional enhancers in their long terminal repeats (LTRs) that can activate the expression of endogenous genes in the vicinity of viral integrations. Additionally, viruses and transposons can be engineered to carry specific sequences that can either activate or disrupt the nearby genes, such as enhancers or transcription stop signals. This has made these elements into powerful tools for forward genetic screens, where they are used to identify the function of endogenous genes.[2,3] An important example of this is the identification of putative oncogenes and tumor suppressor genes from insertional mutagenesis (IM) screens. If an integration activates a proto-oncogene or disrupts a tumor suppressor gene, this can lead to the development of tumors. Mapping the integration loci in resulting tumors and subsequent identification of integration hot spots allows the discovery of the cancer-related genes.[4-6] In addition to forward genetic screens, engineered transposons are increasingly used in the development of functional methods to study mechanisms of gene regulation at a genome-wide level.[7-9]

Retroviruses and transposons do not integrate uniformly across the genome, which limits their usability in molecular biology applications. In this paper, we briefly review the literature on the integration bias of commonly used retroviruses and DNA transposons. We compare the biases between different vectors with a special emphasis on the genetic and epigenetic features of the host cells, which determine these biases. We provide our perspective on how these biases might affect different applications of these vectors. Furthermore, we discuss how the detailed knowledge of the a priori integration bias of these elements can be harnessed to refine some of their applications. Finally, we provide a brief overview of the efforts to control the integration bias of transposons to expand the potential of these tools in molecular biology research.

Chromatin landscapes of integration bias

A common approach for analyzing target site selection is to characterize integration loci in terms of the local genomic and/or chromatin context, and compare them to randomly chosen control loci, e.g.[10-12] Since integration sites are often retrieved using restriction enzymes, and restriction sites are distributed non-uniformly across the genome, each integration site is typically matched to a number of random control loci, based on the distance toward the nearest restriction site. Using this approach, it has been shown that integration preferences differ across different species of retrovirus. Initially, the analyses focused mostly on

*Correspondence to: Jeroen de Ridder; Email: j.deridder@tudelft.nl; Waseem Akhtar; Email: w.akhtar@nki.nl

Submitted: 10/01/2014; Revised: 11/19/2014; Accepted: 11/25/2014
http://dx.doi.org/10.4161/2159256X.2014.992694

genomic marks and gene expression. For example, the Murine Leukemia Virus (MuLV) appeared to strongly favor transcription start sites,[13-15] and the Human and Simian Immunodeficiency Viruses (HIV and SIV) and the Avian Sarcoma-Leukosis Virus (ASLV) favored actively transcribed genes.[13,14,16]

Later studies provided more elaborate analyses of integration profiles, such as scale-based analyses, as well as detailed analyses of sequence specificity and epigenetic marks. By analyzing the profiles of a wide range of retroviruses (among which MuLV and HIV) and 2 transposons (among which the Sleeping Beauty (SB) transposon), it was found that sequence specificity could explain integration bias to a substantial degree, and that conclusions drawn are dependent on the genomic scale at which the insertions and features are analyzed.[10] Furthermore, contrary to HIV, MuLV demonstrated a strong preference for DNase I hypersensitive sites and regions rich in transcription factor binding site motifs.[17,18] Moreover, its bias appeared to be mostly determined by the MuLV-specific integrase [17,18] and the enhancer in the LTR.[17] HIV integrations associated with epigenetic marks such as H3K36me3, consistent with its reported bias for transcriptionally active genes.[19] These and other studies on retroviral integration biases were reviewed extensively in.[20-22]

Recently, with the explosive growth of available epigenomics datasets, attention has shifted more and more to the epigenetic determinants of target site selection. In a comparison of 3700000 MuLV integration sites in K562 cells with the corresponding ENCODE data, the previously reported bias of MuLV [23] for regulatory elements such as enhancers and promoters was confirmed.[24] An especially broad study characterized integration bias of a wide range of retroviruses, MuLV, HIV, ASLV, Porcine Endogenous Retrovirus (PERV), Xenotropic Murine leukemia virus-related Virus (XMRV), Human T-lymphotropic Virus (HTLV), and Foamy Virus (FV) with respect to histone modifications and transcription factor binding as determined by ChIP-seq.[12] Strong association was observed of MuLV, PERV, and XMRV with STAT1, H3/H4 acetylation, and H2AZ/H3K4/K9 methylation. For MuLV specifically, by combining different ChIP-seq data sets, a supermarker was constructed that was present within 2 kb of 75% of the insertion sites. Compared to MuLV, the integration bias of the Mouse Mammary Tumor Virus (MMTV), another retrovirus that is commonly used in IM, is far less extensively studied. Its integration profile was suggested to be the most random across retroviruses, as no preferences could be demonstrated with respect to genes and CpG islands.[25] Based on a large dataset of ~180000 MMTV integrations, we recently demonstrated that biases with respect to genes and CpG islands in fact do exist, but are very weak.[11] As an interesting exception to its generally weak bias, we showed that MMTV did have a strong preference for integrating near the interface between topological domains and their boundary regions.[11,26] Thus, MMTV integration target selection cannot be considered as uniformly random across the genome.

Compared to many retroviruses, transposon integration biases have been less well characterized. Two main systems used in molecular biology are the Sleeping Beauty and the piggyBac (PB) transposons. SB integrates almost exclusively in TA dinucleotides.

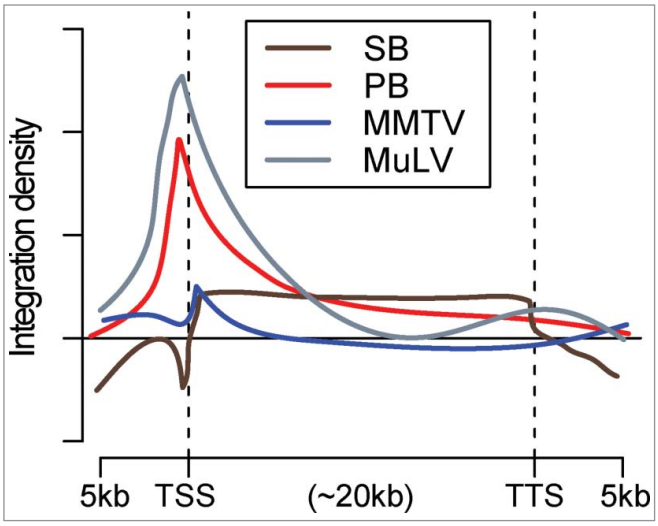

Figure 1. Schematic overview of integration bias with respect to genes for 4 different DNA integrating elements. Adapted from.[11]

Apart from this highly specific recognition sequence, SB did not seem strongly biased.[10,27-29] However, in our recent study,[11] based on a much larger number of integrations (~120000), we found that SB has a strong preference for genes, and preferentially integrates almost uniformly across gene bodies (**Fig. 1**).

PB integrates almost exclusively in TTAA sites,[30,31] and biases were demonstrated with respect to CpG islands, transcription start sites and actively transcribed loci.[11,28,32,33] Interestingly, PB integration profiles are highly similar to those of MuLV [11] (**Fig. 1**).

Across SB, PB and MMTV, we recently identified topological domain boundary interfaces [26] as integration hotspots across different systems.[11] Furthermore, based on a comparison with ~80 publicly available (epi)genomics data sets in the same cell type, we demonstrated that target site selection is directed at multiple genomic scales. At a large scale, it is directed by macrofeatures, i.e. domain-oriented features that are shared between systems, such as expression of proximal genes, proximity to CpG islands and genic features, chromatin compaction and replication timing. At smaller scales, target site selection is directed by microfeatures, i.e., a diverse range of (epi)genomic features, which are generally less domain-oriented and can differ across systems.[11]

The impact of integration bias on applications

As was briefly outlined in the introduction, the integration biases of retroviruses and transposons can pose problems for the applicability of these elements in many areas of molecular biology. In this section we will go into more detail regarding 3 areas of application, 1) cancer gene discovery through insertional mutagenesis (IM) screens, 2) studying the chromatin position effect, i.e. the influence of the genomic location of a genetic unit on its activity, and 3) gene therapy.

Cancer discovery through IM screens

The analysis and interpretation of IM data can be confounded by the a priori integration bias of the DNA integrating elements

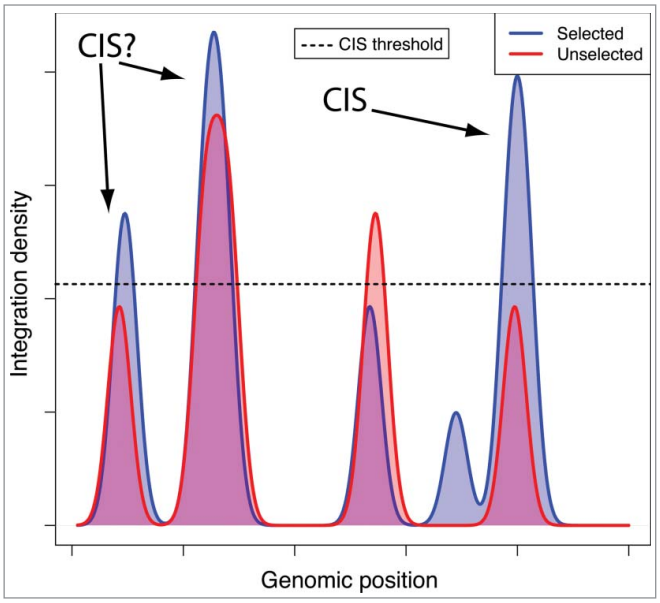

Figure 2. Integration bias can give rise to spurious Common Integration Sites (CISs). Hypothetical integration profile of a tumor screen ("Selected" in blue) and corresponding unselected background integrations ("Unselected" in red). In a typical effort for retrieving cancer genes from an IM screen, a genomic region is called as a CIS if the local integration density exceeds a certain threshold ("CIS threshold;" black dotted line). However, some of these regions may also reflect an a priori integration bias.

used for IM. In IM, putative cancer genes are identified by detecting genomic regions that are recurrently integrated. These genomic regions are called common integration sites (CISs). Identification of CISs is generally done under the assumption that all regions of the genome have an equal probability of hosting an integration event.[34,35] This can lead to spurious CISs, where integration hot spots may be caused merely by passenger integration events rather than tumor-induced selective pressure (**Fig. 2**). In one study, this problem was addressed by assuming that a true CIS gene should harbor significantly more integrations than its flanking genes.[36] In this way, 3 out of 9 CISs in this study were marked as false positive. Other studies have compared control datasets of integrations that were subjected to minimal selective pressure to integration loci retrieved from tumors. One such a study proposed a 47% false positive rate for their MuLV screen.[37] Another SB study found 6 CISs in their control data set, whereas 79 CISs could be found in the tumor screen.[38] We recently compared the integrations from 3 different IM screens utilizing PB, SB and MMTV with the integration profiles of these vectors obtained under unselected conditions.[11] The analysis showed that a substantial fraction of CISs (733%–) overlap with the integration hot spots and are therefore likely not related to the process of tumor development. Especially the integration bias for CIS regions far from endogenous genes was strong. This warrants higher statistical stringency when calling CISs in gene-distant regions of the genome.

Additionally, the use of restriction enzymes for retrieving integration sites could potentially impact CIS calling. To overcome these biases, a method has been developed that can retrieve integration sites by random shearing of DNA.[39] Further, a method for calling CISs based on a Poisson distribution has been used to computationally address restriction site bias.[40]

Depending on the occurrence of integrations relative to endogenous genes and their orientation homogeneity, CISs can have either activating or repressing influence on their target genes, as such allowing to distinguish between putative oncogenes and tumor suppressor genes. In this way, PB was shown to be more efficient at finding oncogenes, whereas SB would be a better tool for mining tumor suppressor genes.[11] These observations highlight the significance of generating large integrations datasets under non-selective conditions in order to refine and prioritize CISs for downstream validation studies.

Studying chromatin position effects

Recently, we presented the TRIP (short for Thousands of Reporters Integrated in Parallel) technology, which depends on mobile genetic elements to study the chromatin position effect in a high-throughput manner.[7,8] A PB construct with a reporter gene was randomly integrated into the genome, and the expression of individual reporter genes was tracked using barcode technology. However, when integrating into the genome, PB shows substantial biases (Section Chromatin landscapes of integration bias). Given these integration biases, one may wonder what the importance of these biases is for computing associations of reporter gene expression with any (epi)genomic features. For example, in the TRIP study [7] we computed the association of reporter gene expression with a number of binarized (epi)genomic features, such as lamina-associated domains (LADs).[41,42] It is known that there is an integration bias against LADs, i.e., there are relatively few integrations within LADs.[11] To demonstrate the influence of this bias on the association of reporter gene expression with LADs we ran a simple simulation. We randomly generated 10^4 integrations in silico, and distributed these in an increasingly uneven fashion across 2 classes, e.g., LADs and inter-LADs (iLADs),[41,42] from completely even (i.e. 5000 in one class and 5000 in the other) to highly uneven (i.e., 9998 in one class and 2 in the other). For each integration, depending on the class of an integration, we simulated expression values by sampling from a certain class-specific expression distribution, i.e. a normal distribution with mean 0.1 and standard deviation 1 for class 1, and a normal distribution with mean 0 and standard deviation 1 for class 2. Then, for each distribution, we performed Welch's *t*-test to distinguish between the 2 classes. The results of the simulation are shown in **Figure 3**. It shows 2 measures, 1) the statistical significance of the *t*-test, expressed as a *z*-normalized *t*-statistic, and 2) the effect size, expressed as the difference in mean reporter gene expression between the 2 classes. As could be expected, it clearly illustrates that with an increasingly uneven distribution, the expected effect size remains the same. However, the variance in the effect size increases, and with it the statistical significance reduces. In other words, given a certain distribution of a number of integrations across 2 classes, a more asymmetric distribution will require a larger total number of integrations to detect a certain effect size as statistically significant. Based on our TRIP data,[7] we computationally estimated that approximately

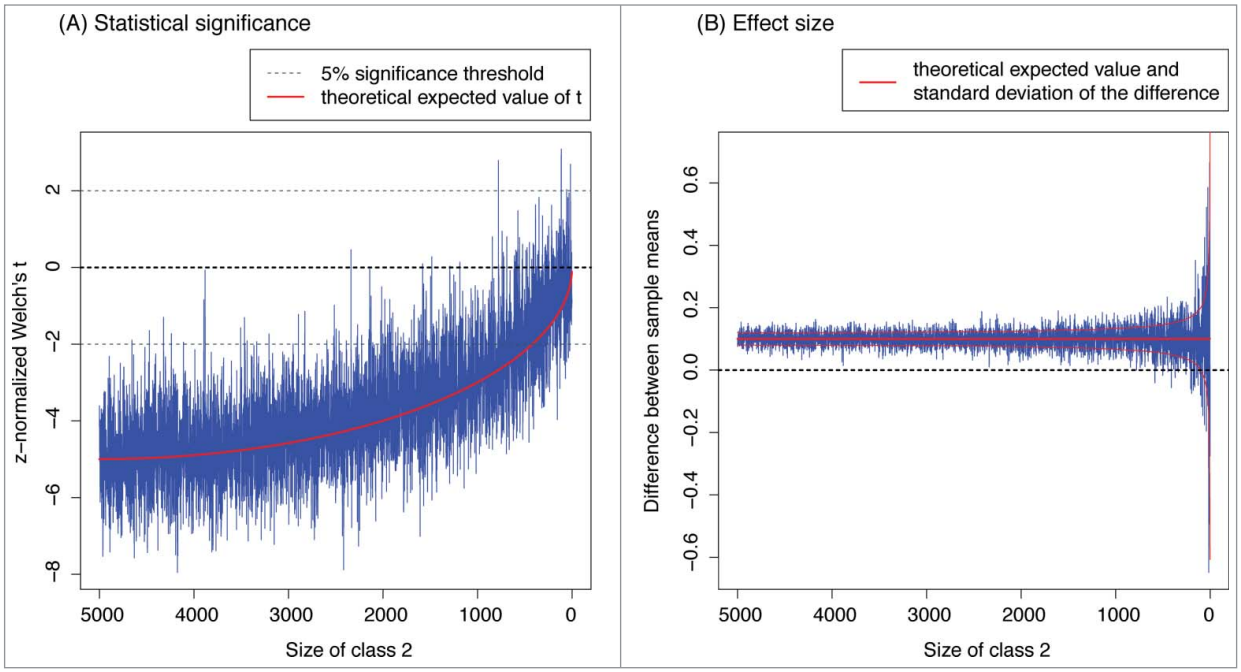

Figure 3. Integration bias reduces statistical power in TRIP applications. 10^4 integrations were generated *in silico*, and distributed across 2 classes (Class 1 and Class 2) in an increasingly uneven fashion. Depending on the assigned class, reporter gene expression was simulated by drawing from a class-specific distribution. Then, (**A**) the significance of the difference between the 2 classes was determined by Welch's t-test as a function of the size of Class 2 (dashed gray line: 2-sided 5% significance threshold; red solid line: theoretical expected value of the *z*-normalized *t*-statistic), and (**B**) the effect size was determined as the difference in means between the 2 classes as a function of the size of Class 2 (red solid lines: theoretical expected value and standard deviation of the sample distribution of the difference). The *x*-axis represents the size of Class 2, which indicates how uneven the distribution across the 2 classes is.

120 PB integrations in total would be sufficient to distinguish between LADs and iLADs in terms of PGK-driven reporter gene expression (in 95% of cases, at a significance level of 5%; data not shown).

Not for all questions it is equally straightforward to determine the (lack of) influence of integration bias. In these cases, it may be needed to regularize the genome-wide integration profile. We provided one such example when inferring PGK domains reflecting genome-wide domains of transcriptional permissiveness, using a hidden Markov model (HMM).[7] Since by inferring an HMM, equidistant spacing of integrations on the genome was assumed, we asked to what extent integration bias affected the eventual domain calling. For this purpose, a non-homogeneous HMM was additionally inferred, with the HMM transition probabilities depending on the distance between integrations.[43] The domains inferred using both approaches were highly similar.

In conclusion, while in the case of interpreting TRIP results it should always be kept in mind that integration is random but biased, the impact of these biases on results seems often limited. However, an important drawback of integration bias is that it reduces statistical power, which can be regained by generating a larger data set of integrations.

Gene therapy

Another important area of research where DNA integrating elements are of great use is gene therapy. Retroviruses and transposons are extensively used in *ex vivo* gene therapy as a molecular vehicle for introducing a therapeutic gene into cells with genetic defects.[1] For this purpose, sustained expression of the introduced gene is desirable. However, this comes with many complications depending on the site of integration of the vector carrying the therapeutic gene. For example, initial gene therapy trials using MuLV showed that viruses integrated in the proximity of proto-oncogenes led to the formation of tumors in some of the treated patients.[44,45] The reason for this is that many MuLV integrations occur in the vicinity of endogenous genes, and more specifically near transcription start sites.[13-15] The use of DNA integrating elements that preferably integrate away from endogenous genes can potentially circumvent this problem. Unfortunately, currently there are no such elements with a distinct preference of integrating away from genes (Section Chromatin landscapes of integration bias). Some insight into this type of bias can be gained by studying large datasets of integrations generated under minimal pressure.[11] When considering the bias of 3 currently used integrating elements,[11] one can see that SB has a higher proportion of integrations landing more than 5kb away from the endogenous genes compared to PB (**Fig. 4**). This means that the chance of gene disruption is comparatively smaller when using SB for gene therapy. Note that, when only considering the integration with respect to genes, MMTV would be even less likely to disrupt endogenous genes, as MMTV has a mild bias against integrating near genes.

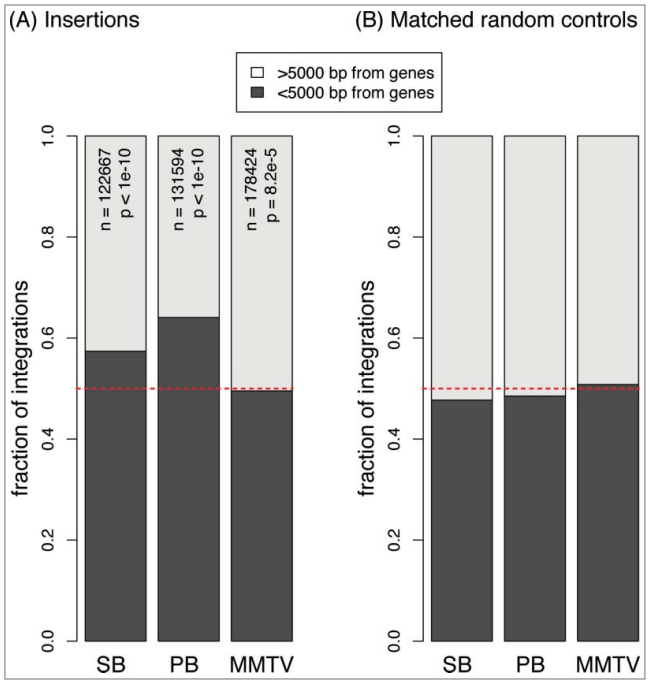

Figure 4. Bias for genic regions of 3 DNA integrating elements. (**A**) For 3 DNA integration elements [11], integrations were counted within and outside 5kb from genes, and *p*-values were determined by 2-sided binomial tests. (**B**) For comparison, the same analysis was done for 3 sets of matched random controls. Refer to [11] for a detailed description of how these controls were generated.

However, limited tropism would restrict its potential use in gene therapy. Hitting cancer-related genes by using any of the integration vectors cannot be completely ruled out. This risk can be substantially reduced by genetically engineering gene therapy vectors capable of integrating the transgene at specific loci away from any endogenous genes (see below).

Future perspectives

As has been outlined above, the bias in the integration profile of DNA integrating elements can be an impediment to realizing the full potential of many applications of these elements such as gene therapy, forward genetic screens and massively parallel chromatin sensor assays such as TRIP. One way to circumvent the bias issue is to genetically modify the integration behavior of these elements, for example by redirecting the integration of these elements to gene-poor regions of the genome. Lentiviral integrations could be directed to heterochromatic regions by fusing the integrase binding domain of host cell encoded LEDGF (involved in the integration of lentiviruses) to CBX1β, which binds to heterochromatin.[46] Blocking the activity of BET proteins, which are cellular binding partners of MuLV integrase, reduces the strong preference of MuLV for endogenous promoters.[47] Along similar lines, the bias of the transposons can be altered by genetically engineering the transposases. Attempts at this have already been made by fusing transposase to the adeno-associated virus Rep protein,[48] zinc finger modules targeting specific sequences[49,50] and custom transcription activator like effector DNA-binding domains.[51] Until now these efforts have yielded only limited success. It is however foreseeable that emerging DNA targeting technologies such as the CRISPR-Cas9 system,[52,53] as well as a deeper understanding of the mechanism of action of transposases, will lead to the engineering of more effective transposition systems. These systems would be capable of precisely targeting the integrations to safe but nonetheless transcriptionally permissive loci of the genome. Such magic transposons will not only make gene therapeutic approaches safer and more controllable, but will also be valuable in studying the chromatin landscape of genomic regions of interest with TRIP-like approaches, at an unprecedented resolution.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

References

1. Kaufmann KB, Buening H, Galy A, Schambach A, Grez M. Gene therapy on the move. EMBO Mol Med 2013; 5(11):1642-61; PMID:24106209
2. Bouwman P, Aly A, Escandell JM, Pieterse M, Bartkova J, van der Gulden H, Hiddingh S, Thanasoula M, Kulkarni A, Yang Q, et al. 53BP1 loss rescues BRCA1 deficiency and is associated with triple-negative and BRCA-mutated breast cancers. Nat Struct & Mol Biol 2010; 17:688-95; PMID:20453858; http://dx.doi.org/10.1038/nsmb.1831
3. Miller EH, Obernosterer G, Raaben M, Herbert AS, Deffieu MS, Krishnan A, Ndungo E, Sandesara RG, Carette JE, Kuehne AI, et al. Ebola virus entry requires the host-programmed recognition of an intracellular receptor. EMBO J 2012; 31(8):1947-60; PMID:22395071; doi:10.1038/emboj.2012.53
4. Kool J, Uren AG, Martins CP, Sie D, de Ridder J, Turner G, van Uitert M, Matentzoglu K, Lagcher W, Krimpenfort P, et al. Insertional mutagenesis in mice deficient for p15Ink4b, p16Ink4a, p21Cip1, and p27Kip1 reveals cancer gene interactions and correlations with tumor phenotypes. Cancer Res 2010; 70:520-31; PMID:20068150; http://dx.doi.org/10.1158/0008-5472.CAN-09-2736
5. Mattison J, Kool J, Uren AG, de Ridder J, Wessels L, Jonkers J, Bignell GR, Butler A, Rust AG, Brosch M, et al. Novel candidate cancer genes identified by a large-scale cross-species comparative oncogenomics approach. Cancer Res 2010; 70:883-95; PMID:20103622; http://dx.doi.org/10.1158/0008-5472.CAN-09-1737
6. Uren AG, Kool J, Matentzoglu K, de Ridder J, Mattison J, van Uitert M, Lagcher W, Sie D, Tanger E, Cox T, et al. Large-scale mutagenesis in p19(ARF)- and p53-deficient mice identifies cancer genes and their collaborative networks. Cell 2008; 133:727-41; PMID:18485879; http://dx.doi.org/10.1016/j.cell.2008.03.021
7. Akhtar W, de Jong J, Pindyurin AV, Pagie L, Meuleman W, de Ridder J, Berns A, Wessels LFA, van Lohuizen M, van Steensel B. Chromatin position effects assayed by thousands of reporters integrated in parallel. Cell 2013; 154:914-27; PMID:23953119; http://dx.doi.org/10.1016/j.cell.2013.07.018
8. Akhtar W, Pindyurin AV, de Jong J, Pagie L, Ten Hoeve J, Berns A, Wessels LF, van Steensel B, van Lohuizen M. Using TRIP for genome-wide position effect analysis in cultured cells. Nat Protoc 2014; 9:1255-81; PMID:24810036; http://dx.doi.org/10.1038/nprot.2014.072
9. Ruf S, Symmons O, Uslu VV, Dolle D, Hot C, Ettwiller L, Spitz F. Large-scale analysis of the regulatory architecture of the mouse genome with a transposon-associated sensor. Nat Genet 2011; 43:379-86; PMID:21423180; http://dx.doi.org/10.1038/ng.790
10. Berry C, Hannenhalli S, Leipzig J, Bushman FD. Selection of Target Sites for Mobile DNA Integration in the Human Genome. PLoS Comput Biol 2006; 2:e157; PMID:17166054; http://dx.doi.org/10.1371/journal.pcbi.0020157
11. de Jong J, Akhtar W, Badhai J, Rust AG, Rad R, Hilkens J, Berns A, van Lohuizen M, Wessels LFA, de Ridder J. Chromatin Landscapes of Retroviral and Transposon Integration Profiles. PLoS Genet 2014; 10:e1004250+; PMID:24721906; http://dx.doi.org/10.1371/journal.pgen.1004250
12. Santoni FA, Hartley O, Luban J. Deciphering the Code for Retroviral Integration Target Site Selection. PLoS Comput Biol 2010; 6:e1001008; PMID:21124862; http://dx.doi.org/10.1371/journal.pcbi.1001008
13. Hematti P, Hong B-K, 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; PMID:15550989; http://dx.doi.org/10.1371/journal.pbio.0020423
14. 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 Biol 2004; 2:e234;

PMID:15314653; http://dx.doi.org/10.1371/journal.pbio.0020234
15. 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; PMID:12805549; http://dx.doi.org/10.1126/science.1083413
16. Schroder AR, Shinn P, Chen H, Berry C, Ecker JR, Bushman F. HIV-1 integration in the human genome favors active genes and local hotspots. Cell 2002; 110:521-9; PMID:12202041; http://dx.doi.org/10.1016/S0092-8674(02)00864-4
17. Felice B, Cattoglio C, Cittaro D, Testa A, Miccio A, Ferrari G, Luzi L, Recchia A, Mavilio F. Transcription factor binding sites are genetic determinants of retroviral integration in the human genome. PLoS One 2009; 4:e4571; PMID:19238208; http://dx.doi.org/10.1371/journal.pone.0004571
18. 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; PMID:16789841; http://dx.doi.org/10.1371/journal.ppat.0020060
19. Wang GP, Ciuffi A, Leipzig J, Berry CC, Bushman FD. HIV integration site selection: Analysis by massively parallel pyrosequencing reveals association with epigenetic modifications. Genome Res 2007; 17:1186-94; PMID:17545577; http://dx.doi.org/10.1101/gr.6286907
20. 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; 3:848-58; PMID:16175173; http://dx.doi.org/10.1038/nrmicro1263
21. Desfarges Se, Ciuffi A. Retroviral Integration Site Selection. Viruses 2010; 2:111-30; PMID:21994603; http://dx.doi.org/10.3390/v2010111
22. Lim K-I. Retroviral integration profiles: their determinants and implications for gene therapy. BMB Rep 2012; 45:207-12; PMID:22531129; http://dx.doi.org/10.5483/BMBRep.2012.45.4.207
23. Cattoglio C, Pellin D, Rizzi E, Maruggi G, Corti G, Miselli F, Sartori D, Guffanti A, Di Serio C, Ambrosi A, et al. High-definition mapping of retroviral integration sites identifies active regulatory elements in human multipotent hematopoietic progenitors. Blood 2010; 116(25):5507-17 blood-2010-05-283523; PMID:20864581; http://dx.doi.org/10.1182/blood-2010-05-283523
24. LaFave MC, Varshney GK, Gildea DE, Wolfsberg TG, Baxevanis AD, Burgess SM. MLV integration site selection is driven by strong enhancers and active promoters. Nucleic Acids Res 2014; 42:4257-69; PMID:24464997; http://dx.doi.org/10.1093/nar/gkt1399
25. Faschinger A, Rouault F, Johannes S, Lukas A, Salmons B, Guenzburg WH, Indik S. Mouse mammary tumor virus integration site selection in human and mouse genomes. J Virol 2007; 82:1360-7; PMID:18032509; http://dx.doi.org/10.1128/JVI.02098-07
26. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, Hu M, Liu JS, Ren B. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature 2012; 485:376-80; PMID:22495300; http://dx.doi.org/10.1038/nature11082
27. Copeland NG, Jenkins NA. Harnessing transposons for cancer gene discovery. Nat Rev Cancer 2010; 10:696-706; PMID:20844553; http://dx.doi.org/10.1038/nrc2916
28. Liang Q, Kong J, Stalker J, Bradley A. Chromosomal mobilization and reintegration of Sleeping Beauty and PiggyBac transposons. Genesis 2009; 47:404-8; PMID:19391106; http://dx.doi.org/10.1002/dvg.20508
29. Vandendriessche T, Ivics Z, Izsvak Z, Chuah MK. Emerging potential of transposons for gene therapy and generation of induced pluripotent stem cells. Blood 2009; 114:1461-8; PMID:19471016; http://dx.doi.org/10.1182/blood-2009-04-210427
30. Balu B, Chauhan C, Maher S, Shoue D, Kissinger J, Fraser M, Adams J. piggyBac is an effective tool for functional analysis of the Plasmodium falciparum genome. BMC Microbiol 2009; 9:83+; PMID:19422698; http://dx.doi.org/10.1186/1471-2180-9-83
31. Li MA, Pettitt SJ, Eckert S, Ning Z, Rice S, Cad1 nanos J, Yusa K, Conte N, Bradley A. The piggybac transposon displays local and distant reintegration preferences and can cause mutations at non-canonical integration sites. Mol Cell Biol 2013; 33(7):1317-30; PMID:23358416; http://dx.doi.org/10.1128/MCB.00670-12
32. Galvan DL, Nakazawa Y, Kaja A, Kettlun C, Cooper LJN, Rooney CM, Wilson MH. Genome-wide mapping of PiggyBac transposon integrations in primary human T cells. J Immunother 2009; 32:837-44; PMID:19752750; http://dx.doi.org/10.1097/CJI.0b013e3181b2914c
33. Rad R, Rad L, Wang W, Cadinanos J, Vassiliou G, Rice S, Campos LS, Yusa K, Banerjee R, Li MA, et al. PiggyBac transposon mutagenesis: a tool for cancer gene discovery in mice. Science 2010; 330:1104-7; PMID:20947725; http://dx.doi.org/10.1126/science.1193004
34. de Jong J, de Ridder J, van der Weyden L, Sun N, van Uitert M, Berns A, van Lohuizen M, Jonkers J, Adams DJ, Wessels LFA. Computational identification of insertional mutagenesis targets for cancer gene discovery. Nucleic Acids Res 2011; 39:e105; PMID:21652642; http://dx.doi.org/10.1093/nar/gkr447
35. de Ridder J, Uren A, Kool J, Reinders M, Wessels L. Detecting statistically significant common insertion sites in retroviral insertional mutagenesis screens. PLoS Comput Biol 2006; 2:e166; PMID:17154714
36. Biffi A, Bartolomae C, Cesana D, Cartier N, Aubourg P, Ranzani M, Cesani M, Benedicenti F, Plati T, Rubagotti E, et al. Lentiviral vector common integration sites in preclinical models and a clinical trial reflect a benign integration bias and not oncogenic selection. Blood 2011; 117:5332-9; PMID:21403130; http://dx.doi.org/10.1182/blood-2010-09-306761
37. Wu X, Luke B, Burgess S. Redefining the common insertion site. Virology 2006; 344:292-5; PMID:16271739; http://dx.doi.org/10.1016/j.virol.2005.08.047
38. Starr TK, Allaei R, Silverstein KAT, Staggs RA, Sarver AL, Bergemann TL, Gupta M, O'Sullivan MG, Matise I, Dupuy AJ, et al. A Transposon-Based Genetic Screen in Mice Identifies Genes Altered in Colorectal Cancer. Science 2009; 323:1747-50; PMID:19251594; http://dx.doi.org/10.1126/science.1163040
39. Koudijs MJ, Klijn C, van der Weyden L, Kool J, ten Hoeve J, Sie D, Prasetyanti PR, Schut E, Kas S, Whipp T, et al. High-throughput semiquantitative analysis of insertional mutations in heterogeneous tumors. Genome Res 2011; 21:2181-9; PMID:21852388; http://dx.doi.org/10.1101/gr.112763.110
40. Bergemann TL, Starr TK, Yu H, Steinbach M, Erdmann J, Chen Y, Cormier RT, Largaespada DA, Silverstein KA. New methods for finding common insertion sites and co-occurring common insertion sites in transposon- and virus-based genetic screens. Nucleic Acids Res 2012; 40:3822-33; PMID:22241771; http://dx.doi.org/10.1093/nar/gkr1295
41. Guelen L, Pagie L, Brasset E, Meuleman W, Faza MB, Talhout W, Eussen BH, de Klein A, Wessels L, de Laat W, et al. Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature 2008; 453:948-51; PMID:18463634; http://dx.doi.org/10.1038/nature06947
42. Peric-Hupkes D, Meuleman W, Pagie L, Bruggeman SWM, Solovei I, Brugman W, Graf S, Flicek P, Kerkhoven RM, van Lohuizen M, et al. Molecular maps of the reorganization of genome-nuclear lamina interactions during differentiation. Mol Cell 2010; 38:603-13; PMID:20513434; http://dx.doi.org/10.1016/j.molcel.2010.03.016
43. Marioni JC, Thorne NP, Tavare S. BioHMM: a heterogeneous hidden Markov model for segmenting array CGH data. Bioinformatics 2006; 22:1144-6; PMID:16533818; http://dx.doi.org/10.1093/bioinformatics/btl089
44. Hacein-Bey-Abina S, Garrigue A, Wang GP, Soulier J, Lim A, Morillon E, Clappier E, Caccavelli L, Delabesse E, Beldjord K, et al. Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. J Clin Invest 2008; 118:3132-42; PMID:18688285; http://dx.doi.org/10.1172/JCI35700
45. 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; 302:415-9; PMID:14564000; http://dx.doi.org/10.1126/science.1088547
46. Gijsbers R, Ronen K, Vets S, Malani N, De Rijck J, McNeely M, Bushman FD, Debyser Z. LEDGF hybrids efficiently retarget lentiviral integration into heterochromatin. Mol Ther 2010; 18:552-60; PMID:20195265; http://dx.doi.org/10.1038/mt.2010.36
47. Sharma A, Larue RC, Plumb MR, Malani N, Male F, Slaughter A, Kessl JJ, Shkriabai N, Coward E, Aiyer SS, et al. BET proteins promote efficient murine leukemia virus integration at transcription start sites. Proc Natl Acad Sci U S A 2013; 110:12036-41; PMID:23818621; http://dx.doi.org/10.1073/pnas.1307157110
48. Ammar I, Gogol-Doering A, Miskey C, Chen W, Cathomen T, Izsvak Z, Ivics Za. Retargeting transposon insertions by the adeno-associated virus Rep protein. Nucleic Acids Research 2012; 40(14):6693-712; PMID:22523082; http://dx.doi.org/10.1093/nar/gks317
49. Kettlun C, Galvan DL, George AL, Kaja A, Wilson MH. Manipulating piggyBac transposon chromosomal integration site selection in human cells. Mol Ther 2011; 19:1636-44; PMID:21730970; http://dx.doi.org/10.1038/mt.2011.129
50. Voigt K, Gogol-Doering A, Miskey C, Chen W, Cathomen T, Izsvak Z, Ivics Z. Retargeting sleeping beauty transposon insertions by engineered zinc finger DNA-binding domains. Mol Ther 2012; 20:1852-62; PMID:22776959; http://dx.doi.org/10.1038/mt.2012.126
51. Owens JB, Mauro D, Stoytchev I, Bhakta MS, Kim M-S, Segal DJ, Moisyadi S. Transcription activator like effector (TALE)-directed piggyBac transposition in human cells. Nucleic Acids Research 2013; 41:9197-207; PMID:23921635; http://dx.doi.org/10.1093/nar/gkt677
52. Cheng AW, Wang H, Yang H, Shi L, Katz Y, Theunissen TW, Rangarajan S, Shivalila CS, Dadon DB, Jaenisch R. Multiplexed activation of endogenous genes by CRISPR-on, an RNA-guided transcriptional activator system. Cell Research 2013; 23(10):1163-71; PMID:23979020; http://dx.doi.org/10.1038/cr.2013.122
53. Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK. CRISPR RNA-guided activation of endogenous human genes. Nature Methods 2013; 10:977-9; PMID:23892898; http://dx.doi.org/10.1038/nmeth.2598