Retroviral insertion sites and cancer: Fountain of all knowledge?

James C. Neil[1] and Ewan R. Cameron

Molecular Oncology Laboratory, Institute of Comparative Medicine, University of Glasgow Veterinary School, Bearsden, Glasgow G61 1QH, United Kingdom
[1]Correspondence: j.c.neil@vet.gla.ac.uk

Retroviral gene tagging is enjoying a renaissance as a gene discovery method since the completion of the draft mouse genome sequence. The potential of this approach to elucidate the genetic basis of cancer is reviewed in the light of a series of recent papers that report the results of high-throughput screens.

More than 20 years have elapsed since Hayward and colleagues first demonstrated that the c-*myc* proto-oncogene could be activated by retroviral promoter insertion and that this was a common event in the genesis of avian bursal lymphomas (Hayward et al., 1981). Since that time, the search for common retroviral insertion sites in cancer cells has been a highly productive (if sometimes laborious) means to identify host cell genes relevant to cancer. Retroviral replication entails the stable integration of a DNA copy of the viral RNA genome into host cell DNA, a process which is inherently mutagenic. The assumptions underlying the gene tagging strategy are that insertion is essentially a random process and that the occurrence of multiple insertions within a narrow genomic domain in independently derived tumors provides evidence of a selective advantage to cells bearing these insertions. With these defining criteria, many common insertion sites (CIS) have been identified, leading to the discovery of many host genes that are subject to mutation in the carcinogenic process. The most commonly observed outcome of insertional mutagenesis of CIS target genes is transcriptional activation due to the insertion of viral promoters of enhancer elements. In some cases viral integration disrupts the translation unit, leading either to gain of function or inactivation. Murine leukemia virus (MLV) models have been the most widely employed, as investigators have been attracted by the range of retroviral isolates with tissue-specific oncogenic activity and the availability of transgenesis as a means to establish the oncogenic potential of the newly discovered CIS genes. Moreover, the transgenic approach allows the combined effects of candidate oncogenes to be tested in vivo, while the gene-tagging process can be combined with transgenesis in a systematic and directed approach for the identification of specific collaborating gene sets (Jonkers and Berns, 1996).

The detection of the gene(s) affected by insertions at a CIS has until recently been a slow process, often involving genome walking from the clustered site and a painstaking search for the affected transcription unit(s) over many kilobases of host DNA. In some cases these difficulties have led to the search being abandoned without discovery of a relevant target gene. This situation has changed radically following the advent of rapid PCR cloning methods and the recent completion of the draft mouse genome sequence. Together, these advances have reduced the process of isolating and locating proviral insertion sites to a comparatively straightforward exercise. The resulting resurgence of interest in retroviral gene tagging as a gene discovery strategy is documented in a series of recently published papers (Mikkers et al., 2002; Lund et al., 2002; Suzuki et al., 2002).

Collectively, these papers catalog over a hundred new loci with potential relevance to cancer. Has this advance brought within our grasp a comprehensive list of genes with the potential to contribute to hematopoietic malignancy, along with an understanding of the genetic and biochemical pathways that link them? It will certainly aid progress toward this goal, but there are some gaps that must be filled before we can consider it attainable. The significance of some of the new gene targets is strongly reinforced by a degree of overlap between the three studies and the close relatedness of some to established oncogenes, including many that are involved in human cancer (Suzuki et al., 2002). However, much of the newly published data must be regarded as preliminary, as oncogenic significance has yet to be confirmed for most of the novel CISs. The standard of proof previously required for a single locus cannot be waived merely because of the impressive size of the new data sets.

A notable feature of all three studies is the very high frequency of integrations at a small number of previously identified targets, which overlap to some degree but are distinctive and characteristic for each virus/host system. In contrast, the majority of the new common insertion sites are defined by the minimal criterion of only two hits within a specified "window" of genomic DNA (30 kb). Moreover, even single insertions have been scored where these have occurred at known oncogenes or their homologs. The statistical argument that clustering is selected by the oncogenic process depends on the assumption that the process of retroviral insertion is close to random. This is an issue that was debated and tested extensively prior to the availability of complete genome sequences (Weidhaas et al., 2000). It must now be reexamined. Concerns that a nonrandom pattern may apply are highlighted by a recent analysis of over 500 HIV integrations in a human T cell line that exploits the draft human genome sequence to reveal hot-spots of integration as well as a preference for active genes, in the absence of any selection for cellular growth advantage (Schroder et al., 2002). Indeed, a number of the sites of HIV insertion might have been flagged for potential oncogenic significance had they occurred in the MLV studies. Hot-spots of the magnitude described for HIV, which account for up to 1% of integrations, could not account for the radical bias seen at some CISs where up to 90% of a tumor cohort may display integrations at the same locus. They could, however, generate a degree of noise in the system, particularly at the low level of detection of many of the newly reported CISs.

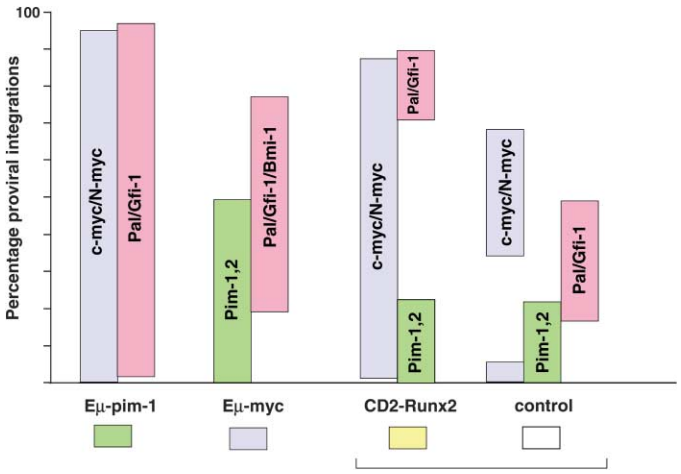

Figure 1. Collaborating genes and complementation groups identified by retroviral gene tagging in oncogene transgenic mice

The diagram, which is collated from two published studies (Scheijen et al., 1997; Blyth et al., 2001), depicts the number of tumors affected by proviral insertion at specific CISs in MoMLV induced tumors of Eμ-*pim-1* (n = 43), Eμ-*myc* (n = 85), CD2-*Runx2* (n = 44), and control mice from the CD2-*Runx2* cohort (n = 18). The colored bars indicate the frequency of hit at each complementation group, as defined by mutually exclusive CISs. The overlap between bars indicates tumors that display insertions at CISs from both complementation groups. The colored boxes at the foot are coded to indicate the complementation group represented by the initiating oncogene (transgene) in each case. A bracket underneath the two cohorts on the right denotes comparison between transgenic and littermate controls in a single study.

Moreover, if levels of gene activity also drive selection for MLV, integration preference will be tissue specific and liable to evolve during the carcinogenic process. An intriguing scenario can be envisaged where activation of an initiating gene primes other genes for targeting and leads to their selective activation as secondary or progression events. While this speculative notion would seem unlikely to undermine the general validity of gene tagging as a means of identifying complementing oncogenes, which is backed up by a large body of data, it illustrates the need to follow up the initial observation of a new CIS with formal evidence of its oncogenic consequences.

How many of the newly identified genes are functionally important targets for activation (or inactivation)? Previous studies have shown that MLV can affect target gene expression over several hundred kilobases (Lazo et al., 1990), suggesting that some of the novel integrations might be targeting known rather than new players in the oncogenic process. A candidate for rationalization is *HBS1*, a gene on mouse chromosome 10 that encodes a putative elongation release factor and was scored as one of the most frequently targeted new CISs with a combined total of 11 hits in two of the recent papers (Mikkers et al., 2002; Lund et al., 2002). This gene is only 120 kb upstream of *Myb* and is the closest gene to a previously identified CIS, *fit-1(fti-1)* (Barr et al., 1999). Could insertions at this site be activating both *Myb* and *HBS1*? A potentially telling observation is that these two genes are coamplified in some cases of human pancreatic cancer, but the fact that *HBS1* is often missing from the amplification unit has led to the argument that it is a passive partner in this process, with *Myb* alone playing the essential oncogenic role (Wallrapp et al., 1997).

How many hits of retroviral integration are required to convert a normal cell to a fully malignant clone? Suzuki et al. report up to six CISs hit in a single tumor (Suzuki et al., 2002), although this is an exceptional case, while Mikkers et al. found an average of 3.53 per tumor (Mikkers et al., 2002). It is unclear whether these numbers have been corrected for clonal complexity of the tumors, which can be assessed for lymphoid tumors using the clonal markers of rearranged TCR and Ig genes. This issue is important as oligoclonal tumors may contain a series of independently transformed cells or clonal derivatives of a single primary tumor that have acquired different secondary hits. Other recent studies shed light on this issue and on possible limitations to the gene tagging approach in defining oncogenic complementation groups, a concept that was first elaborated in detail in studies of Eμ-myc mice by the Berns group, who defined two groups of *Myc* collaborating genes by neonatal MLV infection and screening for target genes in the accelerated lymphomas (van Lohuizen et al., 1991). A complementation group is defined by genes that are subject to mutually exclusive targeting in the tumors (see also Mikkers et al., 2002). This phenomenon is readily understood where it affects partially redundant gene families such as the *Pim* kinases, but it may also include distinct target genes such as *Bmi-1* and *Gfi-1*, two unlinked genes that are each flanked by extended clusters of CIS (*Gfi-1* here is shorthand for four closely linked CISs, *eis-1pal-1/gfi1/evi-5*) (Scheijen et al., 1997).

Figure 1 is distilled from two published studies that address the targeting of known CISs in different oncogene transgenic mouse strains (Scheijen et al., 1997; Blyth et al., 2001). Accelerated Eμ-*myc* lymphomas display insertions at *Pim* family genes and at the *Bmi-1/Gfi-1* group. The reciprocal experiment of gene tagging in Eμ-*Pim-1* mice shows a remarkable level of saturation with insertions at c-*Myc* or *N-Myc* and *Pal-1/Gfi-1* as secondary events in the vast majority of the tumors. However, this apparently simple closed system of three collaborating oncogenic pathways breaks down in virus-accelerated T cell lymphomas of CD2-*Runx2* transgenic mice (Blyth et al., 2001). A *Pim*-like preference for activation of the *Myc* pathway was noted in the *Runx2* transgenic mice that display a highly significant increase in the hit rate at *Myc* family genes compared to their littermate controls (Figure 1). The fact that *Pim1,2* and *Pal-1/Gfi-1* loci are also targeted in these mice places the *Runx2* oncogene in a new complementation group that can synergize with either of the previously defined pathways. As the *Pim* and *Gfi-1* insertions did not overlap in these tumors, it seems that targeting both pathways is no longer required when *Runx2* is the initiating gene and *Myc* is activated. However, it is important to note that the lower hit rate at either gene group on this background renders the lack of overlap statistically insignificant. This problem highlights the difficulty we face in functional grouping of the new CIS that are hit only rarely and have no clear structural homologs.

What factors currently limit the spectrum of genes that can be detected by the retroviral tagging approach? The exquisite oncogenic specificity of the MLV family stems mainly from their LTR enhancer elements that consist of unique and complex arrays of binding sites for tissue-specific transcription factors including Runx, Ets, and Myb (Lewis et al., 1999). The strong selection for specific CIS targets must be due in part to the fact that the relevant target gene promoters are among those most amenable to activation by these enhancer elements. Moreover, as many studies have shown that the spectrum and hit rate of CIS is heavily influenced by mouse strain background, other as

yet unidentified genetic factors must be operative. This point must be borne in mind as it generally invalidates statistical comparisons between studies.

Marked tissue-specific differences are also evident in the interaction of MLVs with tumor suppressor pathways that may be direct or indirect. While inactivation of the p53 gene in Friend MLV-induced erythroleukemias has recently been reevaluated and is proposed to be a prerequisite for tumor progression in vivo (Prasher et al., 2001), Moloney MLV favors the development of lymphomas that retain functional p53 (Baxter et al., 2001). The latter phenomenon may be reflected in the low rate of loss of heterozygosity detected in Moloney induced lymphomas (Lander and Fan, 1997), but this could not be explained in terms of MLV insertion at specific loci, as the p53 null genotype had no significant impact on the hit rate at a number of CISs compared to littermate controls (Baxter et al., 1996, 2001). Similar findings are now reported by Lund and colleagues, who studied MLV tumors in Cdkn2A null mice. These mice are defective for expression of the overlapping gene products p16Ink4a and p19Arf, which regulate the Rb and p53 pathways (Lund et al., 2002). The study revealed a remarkable lack of effect of Cdkn2A deficiency on the hit rate at oncogenes that are known to collaborate with this genotype (such as Myc) (Eischen et al., 1999) or that might be expected to fall into a Cdkn2A complementation group (e.g., Bmi-1/Gfi-1) (Scheijen et al., 1997; Jacobs et al., 1999). The strongest skew they detected affected the *Tpl-2* gene, but it is notable that these insertions were observed exclusively in histiocytic sarcomas, a novel tumor type seen only in the Cdkn2A background (Lund et al., 2002). These results support our previous suggestion that Moloney MLV gene tagging may have a "blind spot" with respect to the targeting of certain tumor suppressor pathways in T-lymphoma cells (Baxter et al., 1996). Although we have no definitive leads to the mechanism, Moloney MLV may have evolved the capacity to suppress the Arf-p53 response in T cells. If the effect is mediated in *trans,* this would account for the apparent lack of requirement for specific insertion events targeting the pathway.

The clearest example of the power and specificity of retroviral tagging in dissecting pathways is provided by the study from Mikkers and colleagues who used Eμ-*myc* transgenic mice deficient in both *Pim-1* and *Pim-2* to screen for genes that would act as alternative targets in their absence (Mikkers et al., 2002). Reassuringly, the new targets detected by this process included the related *Pim-3* gene and were able to be validated by examination of the effects of insertion on gene transcription. However, it will be noted that this heroic effort would not be applied easily to other gene families where functional inactivation compromises viability.

These studies on MLV and HIV represent the first fruits of whole genome sequences as applied to phenomena associated with retroviral integration. As similar high-throughput screens are in progress in the public and commercial sectors, publication of the details will be vital to allow rapid comparison of data sets and assist in distinguishing integration preference from selective processes operating on the infected cell. These databases will also inform assessments of the potential hazards of retroviral vector integration in the human genome and hopefully aid the design of safer delivery systems for future clinical application.

Selected reading

Barr, N.I., Stewart, M., Tsatsanis, C., Fulton, R., Hu, M., Tsujimoto, H., and Neil, J.C. (1999). Mamm. Genome *10*, 556–559.

Baxter, E.W., Blyth, K., Donehower, L.A., Cameron, E.R., Onions, D.E., and Neil, J.C. (1996). J. Virol. *70*, 2095–2100.

Baxter, E.W., Blyth, K., Cameron, E.R., and Neil, J.C. (2001). J. Virol. *75*, 9790–9798.

Blyth, K., Terry, A., Mackay, N., Vaillant, F., Bell, M., Cameron, E.R., Neil, J.C., and Stewart, M. (2001). Oncogene *20*, 295–302.

Eischen, C.M., Weber, J.D., Roussel, M.F., Sherr, C.J., and Cleveland, J.L. (1999). Genes Dev. *13*, 2658–2669.

Hayward, W.S., Neel, B.G., and Astrin, S.M. (1981). Nature *290*, 475–480.

Jacobs, J.J., Kieboom, K., Marino, S., DePinho, R.A., and van Lohuizen, M. (1999). Nature *397*, 164–168.

Jonkers, J., and Berns, A. (1996). Biochim. Biophys. Acta *1287*, 29–57.

Lander, J.K., and Fan, H. (1997). J. Virol. *71*, 3940–3952.

Lazo, P.A., Lee, J.S., and Tsichlis, P.N. (1990). Proc. Natl. Acad. Sci. USA *87*, 170–173.

Lewis, A.F., Stacy, T., Green, W.R., Taddesse-Heath, L., Hartley, J.W., and Speck, N.A. (1999). J. Virol. *73*, 5535–5547.

Lund, A.H., Turner, G., Trubetskoy, A., Verhoeven, E., Wientjens, E., Hulsman, D., Russell, R., DePinho, R.A., Lenz, J., and van Lohuizen, M. (2002). Nat. Genet. *32*, 160–165.

Mikkers, H., Allen, J., Knipscheer, P., Romeyn, L., Hart, A., and Berns, A. (2002). Nat. Genet. *32*, 153–159.

Prasher, J.M., Elenitoba-Johnson, K.S., and Kelley, L.L. (2001). Oncogene *20*, 2946–2955.

Scheijen, B., Jonkers, J., Acton, D., and Berns, A. (1997). J. Virol. *71*, 9–16.

Schroder, A.R., Shinn, P., Chen, H., Berry, C., Ecker, J.R., and Bushman, F. (2002). Cell *110*, 521–529.

Suzuki, T., Shen, H., Akagi, K., Morse, H.C., Malley, J.D., Naiman, D.Q., Jenkins, N.A., and Copeland, N.G. (2002). Nat. Genet. *32*, 166–174.

van Lohuizen, M., Verbeek, S., Scheijen, B., Wientjens, E., van der Gulden, H., and Berns, A. (1991). Cell *65*, 737–752.

Wallrapp, C., Muller-Pillasch, F., Solinas-Toldo, S., Lichter, P., Friess, H., Buchler, M., Fink, T., Adler, G., and Gress, T. (1997). Cancer Res. *57*, 3135–3139.

Weidhaas, J.B., Angelichio, E.L., Fenner, S., and Coffin, J.M. (2000). J. Virol. *74*, 8382–8389.