Oncogene (2005) 24, 7656–7672

www.nature.com/onc

Retroviral insertional mutagenesis: past, present and future

AG Uren[1,2], J Kool[1,2], A Berns*[,1] and M van Lohuizen*[,1]

[1]*Division of Molecular Genetics, Netherlands Cancer Institute, Plesmanlaan 121, 1066CX Amsterdam, The Netherlands*

Retroviral insertion mutagenesis screens in mice are powerful tools for efficient identification of oncogenic mutations in an *in vivo* setting. Many oncogenes identified in these screens have also been shown to play a causal role in the development of human cancers. Sequencing and annotation of the mouse genome, along with recent improvements in insertion site cloning has greatly facilitated identification of oncogenic events in retrovirus-induced tumours. In this review, we discuss the features of retroviral insertion mutagenesis screens, covering the mechanisms by which retroviral insertions mutate cellular genes, the practical aspects of insertion site cloning, the identification and analysis of common insertion sites, and finally we address the potential for use of somatic insertional mutagens in the study of non-haematopoietic and nonmammary tumour types.
Oncogene (2005) **24,** 7656–7672. doi:10.1038/sj.onc.1209043

Keywords: retroviral insertional mutagenesis; murine leukemia virus; genetic screens; cancer; common insertion site; lymphoma

Forward genetic screens in somatic tissues

Development of cancer requires the accumulation of genetic and epigenetic changes that each confer a selective advantage to the cell at a specific stage of tumorigenesis. Identification of genes involved in tumorigenesis can be greatly facilitated by genetic screens for mutations contributing to tumour formation in model organisms such as the mouse. Success of such screens is dependent upon (a) a high rate of mutagenesis in order to achieve sufficient hits for oncogenic transformation and (b) easy identification of the mutated gene(s). Furthermore, the identified mutations should not be restricted to tumorigenesis in the organism used for the screen, but should ideally also be relevant to human cancers.

Slow transforming retroviruses can efficiently induce oncogenic mutations via insertion of the provirus into the genome, and these oncogenic mutations can be identified relatively easily by determining the insertion site of the provirus. Consequently, these viruses have been widely used in genetic screens for mutations involved in the onset of tumorigenesis in various model organisms.

*Correspondence: A Berns or M van Lohuizen;
E-mails: a.berns@nki.nl or m.v.lohuizen@nki.nl
[2]These authors contributed equally to this work

Transforming retroviruses as tools for genetic screening

Oncogenic retroviruses can be divided into two classes: acute and slow transforming viruses. Acute transforming retroviruses induce polyclonal tumours within 2 to 3 weeks after infection of the host. Transformation by these viruses is mediated by the expression of viral oncogenes such as *v-Abl* in the Abelson Murine leukaemia Virus (reviewed in Shore *et al*., 2002), and *v-Myb* from the avian myoblastosis virus (reviewed in Lipsick and Wang, 1999), which are virally encoded oncogenic versions of normal cellular genes. In contrast, slow transforming retroviruses induce mono- or oligoclonal tumours with a longer latency of 3–12 months. Tumorigenesis caused by slow transforming retroviruses is not induced by the presence of a virally modified proto-oncogene in the virus but results from mutations caused by insertion of the proviral retrovirus into the host genome. Elements in the proviral genome that regulate the viral transcript also act in *cis* on cellular gene transcripts. Depending on whether the provirus integrates into or in the vicinity of genes, these elements can enhance or disrupt normal transcription and thus induce oncogenic mutations.

Retroviral infection of cells is dependent on the binding of viral envelope proteins to cell surface receptors. After incorporation of the provirus into the genome, cells will start producing viral envelope proteins and cell surface receptors become occupied with these proteins, thus inhibiting reinfection of the cell. Recombination of the viral sequences with endogenous viral sequences encoding envelope proteins gives rise to mutant viruses encoding envelope protein variants that utilize different receptors, and thus can reinfect an infected cell. Furthermore, many integrated proviruses appear to harbor defective envelope sequences allowing reinfection by the same virus. In this way, cells can undergo multiple rounds of viral infection and mutations can accumulate in one cell (reviewed in Mikkers and Berns, 2003). The ability to induce multiple mutations in the same cell, combined with the fact that the proviral insertion site can be easily identified by amplification of the genomic sequences flanking the

retroviral insertion (see below), makes slow retroviruses very suitable for genetic screens for oncogenic mutations.

Slow transforming retroviruses have been found to induce tumours in various organisms like cats (FeLV), birds (ALV and REV) and mice (murine leukaemia virus (MLV) and mouse mammary tumour virus (MMTV).

LTRs regulate viral transcription

Mutation of cellular genes by retroviral insertions is mediated by proviral elements that drive and regulate retroviral transcription. These elements are present in the Long Terminal Repeats (LTRs) located at each end of the provirus (Figure 1). The LTR is composed of thee parts: U3, which can be subdivided into an enhancer and promoter region, R, containing both the start and termination sites for transcription, and U5.

The promoter in the U3 of MLV contains sequences involved in recruitment of the basal transcription machinery such as a TATA box and GC-rich sequences (Graves *et al*., 1985). The structure of the MLV enhancer region is more complex and contains binding sites for transcription factors, which positively affect transcription, such as Ets (Gunther and Graves, 1994), NF1 (Reisman, 1990) and CBF (Sun *et al*., 1995) family members. Efficient viral replication will be restricted to those cellular environments that provide the various transcription factors required for high levels of viral transcription. Therefore, cell-type specificity of retroviruses is governed by the constitution of the viral LTR. This is exemplified by the tropism of Moloney MLV for T- and B-cells and the induction of predominantly myeloid leukaemia by the Graffi-type murine leukaemia viruses (Erkeland *et al*., 2003, 2004). Substitution of MLV U3 enhancer sequences with Feline leukaemia Virus U3 sequences induces a different spectrum of insertion sites in lymphoma than wild-type MLV virus, indicating that enhancer sequences not only regulate cell-type specificity of viral replication but also determine the pattern of insertional activation by the provirus (Starkey *et al*., 1998; Johnson *et al*., 2005; Nielsen *et al*., 2005).

Mutation of cellular genes by proviral insertions can be achieved in different ways and may result in enhanced transcript levels, chimeric transcripts or truncation of cellular gene transcripts. Alteration of cellular gene transcription is mediated by proviral elements that are involved in the transcriptional regulation of the provirus; in particular, elements present in the LTR and the splice donor sites normally used for splicing of subgenomic sized viral transcripts. In the following sections, the various ways in which retroviral insertions mutate cellular genes are discussed, using examples of oncogenes and tumour suppressors that have been identified in retroviral insertion mutagenesis screens.

Proviral insertions enhance expression of the c-Myc proto-oncogene

The activity of enhancers is not restricted to the most proximal promoter but can also augment the activity of promoters over large distances via chromatin loop interactions (reviewed in West and Fraser, 2005). Thus, viral enhancers of integrated proviruses can affect many genes, explaining why the majority of mutations found in retroviral insertion mutagenesis screens are enhancer mutations resulting in elevated expression of cellular genes. Enhancer mutations are predominantly found upstream of genes in the antisense orientation or downstream in the sense orientation. This may be in order to prevent the positioning of the viral promoter between the enhancer in the 5′LTR and the cellular gene possibly, because the enhancer is only capable of acting bi-directionally on the most proximal promoters.

An alternative explanation may be that viral enhancers (5′ or 3′) only function when they are not transcribed. The later model is supported by results showing that in MMTV-induced tumours enhancer mutations were found downstream of genes with the provirus in the antisense orientation. In these cases, the 5′LTR was deleted and transcription of the cellular

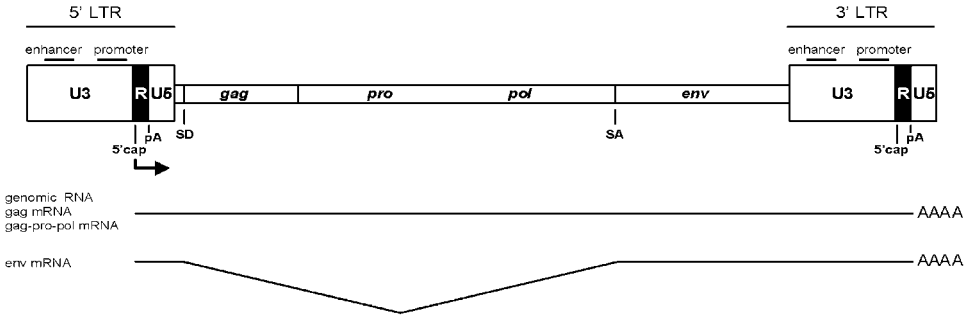

Figure 1 Genomic organization of a retroviral provirus. The provirus contains long terminal repeats (LTRs) at both ends. LTRs can be subdivided into three regions: U3, R and U5. U3 contains the enhancer and promoter sequences that drive viral transcription, the R domains encode the 5′ capping sequences (5′ cap) and the polyA (pA) signal. Gag, pro, pol and env encode the viral components required for the assembly of viral particles. The Splice donor (SD) and splice acceptor (SA) necessary for the formation of the subgenomic mRNAs (e.g. env mRNA) that encode different viral proteins are also indicated

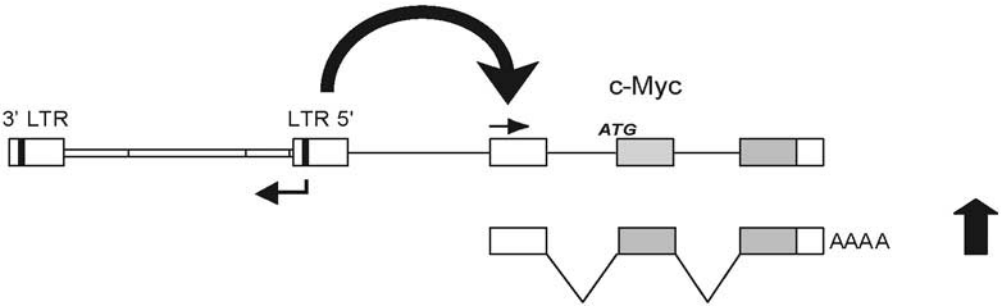

Figure 2 Induction of *c-Myc* expression by proviral enhancer. The enhancer in the 5′ LTR of the provirus, which is integrated upstream of the *c-Myc* gene enhances c-Myc expression. The provirus is inserted such that the transcriptional direction of the provirus is opposite to transcription of the cellular gene. Exons are drawn as boxes (coding sequences depicted in grey). The transcriptional and translational start site of the *c-Myc* gene is indicated (horizontal arrow and ATG, respectively)

gene was augmented by the enhancer in the 3′LTR (Clausse *et al*., 1993).

Overexpression of the *c-Myc* proto-oncogene is a frequently found oncogenic mutation in retrovirally induced tumours, and mainly induces haematopoietic tumours like erythroleukaemias and T-cell lymphomas (reviewed in Dudley *et al*., 2002). Depending on the retrovirus, different strategies are employed to mutate this gene, for example, *c-Myc* is mutated by proviral insertions that lead to fusion transcripts in 80% of avian leucosis virus (ALV)-induced tumours (Rosenberg and Jolicoeur, 1997, see below), whereas *c-Myc* is primarily mutated by enhancer mutations in MLV-induced tumours (Figure 2) (Corcoran *et al*., 1984; Cuypers *et al*., 1984; Selten *et al*., 1984). Enhancer mutations as far as 270 kb 3′ of the promoter region of *cMyc* can affect Myc transcription levels, illustrating that enhancer mutations can have long-range effects on cellular gene transcription (Lazo *et al*., 1990).

The *c-Myc* gene is often found mutated or transcriptionally deregulated in a wide range of human tumours, and overexpression of c-Myc induces tumorigenesis in mice (reviewed by Pelengaris *et al*., 2002). The onset of c-Myc-induced lymhomagenesis is greatly accelerated by inactivation of the $p19^{Arf}$–Mdm2–p53 pathway (Eischen *et al*., 1999; Jacobs *et al*., 1999b), which induces apoptosis in response to high Myc levels (Zindy *et al*., 1998). Indeed, Bmi-1, a repressor of the *Cdkn2a* locus encoding p19ARF and p16 INK4a (Jacobs *et al*., 1999a), is often overexpressed as a result of proviral insertions in MLV-infected *EμMyc*-transgenic mice (van Lohuizen *et al*., 1991). Also, members of the Pim serine/threonine kinase family and Gfi-1 have been identified in retroviral mutagenesis screens as strong collaborators with overexpressed c-Myc in lymphomagenesis (van Lohuizen *et al*., 1989, 1991: van der Lugt *et al*., 1995; Mikkers *et al*., 2002b). Like Bmi-1, both the Pim kinases and the Gfi-1 transcription factor may have anti-apoptotic activities (Grimes *et al*., 1996; Fox *et al*., 2003; Aho *et al*., 2004), which indicates that reduction of apoptosis is an important requirement for lymphomagenesis of Myc-overexpressing cells.

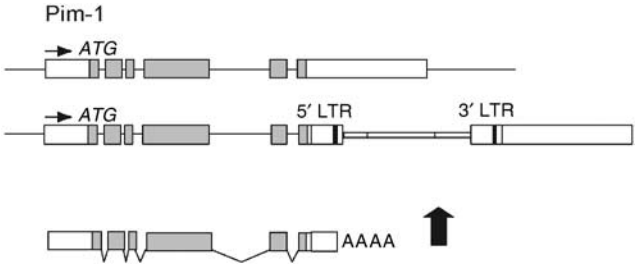

Figure 3 Pim-1 induction by removal of 3′ noncoding sequences. Proviral insertions in the last noncoding exon of the cellular *Pim-1* gene terminate *Pim-1* gene transcription by the proviral polyA signal. Hereby, 3′ UTR sequences that negatively affect *Pim-1* mRNA stability are removed, which results in higher *Pim-1* mRNA and protein levels. Exons are depicted as boxes (coding sequences in grey). The transcriptional and translational start site of the cellular gene is indicated (horizontal arrow and ATG, respectively)

Enhancement of N-myc and Pim-1 expression by removal of 3′ sequences

Proviral insertions in the 3′UTR-encoding region of the gene may remove regulatory or destabilizing motifs in mRNAs by termination of gene transcription by either the polyadenylation signal present or the antisense cryptic polyadenylation signal in the viral enhancer. This kind of mutations will increase the levels of truncated, but wild-type protein-encoding mRNA and subsequently lead to enhanced levels of wild-type protein.

Sense proviral insertions in the 3′-terminal exon of the *Pim-1* gene remove a portion of the 3′UTR, which is thought to contain motifs that reduce mRNA stability (Figure 3). Thus, loss of this region increases *Pim-1* expression levels (Cuypers *et al*., 1984; Selten *et al*. 1985; Domen *et al*., 1987). Overexpression of the Pim-1 serine/threonine kinase strongly collaborates with c-Myc activation in lymphomagenesis (van Lohuizen *et al*., 1991). In fact, combination of an *EμMyc* and an *EμPim* transgene resulted in the perinatal death of doubly transgenic mice caused by pre-B cell leukaemias (Verbeek *et al*., 1991). The genes encoding two other Pim serine/threonine kinases, Pim-2 and Pim-3, were identified as frequent MLV proviral insertion sites in

EμMyc/Pim1$^{-/-}$ and *EμMyc/Pim-1*$^{-/-}$*/Pim-2*$^{-/-}$ mice, respectively, indicating that members of this family can functionally complement each other in tumorigenesis (van der Lugt *et al*., 1995; Mikkers *et al*., 2002b). Pim proteins are most likely involved in growth factor and cytokine signaling. Gene disruption of Pim family members in mice results in mild haematopoietic proliferation defects (Mikkers *et al*., 2004 and references therein). Pim proteins may also suppress apoptosis in haematopoietic cells via inactivation of the proapoptotic Bad protein (Fox *et al*., 2003; Aho *et al*., 2004) and cause genomic instability when overexpressed (Roh *et al*., 2003). Interestingly, high levels of Pim-1 expression have also been found in prostate cancers that were induced by overexpression of human c-Myc in the murine prostate, indicating that Pim-1 can also collaborate with c-Myc in the induction of nonhaematopoietic malignancies (Ellwood-Yen *et al*., 2003). Moreover, increased Pim-1 expression is also found in human prostate cancers and correlates significantly with measures of clinical outcome (Dhanasekaran *et al*., 2001).

Activation of N-myc in MLV-induced leukaemias provides another example of oncogenic mutation of a cellular gene through removal of 3′UTR sequences (van Lohuizen *et al*., 1989). In a panel of 40 tumours, 15 tumours contained insertions in the *N-myc* gene, indicating that *N-myc* is a common target of MLV-insertions. Remarkably, all the insertions were found within 100 bp of the last exon of N-myc, which results in the removal of about 600 bp of the 3′UTR due to termination of gene transcription by the polyA signal in the 5′ LTR. Tumours showed elevated levels of N-myc expression, suggesting that removal of the 3′UTR stabilizes the *N-myc* mRNA (van Lohuizen *et al*., 1989). *N-myc* has been found mutated or amplified in human malignancies like neuroblastoma, medulloblastoma and small cell lung cancers (Table 1) (reviewed by Nesbit *et al*., 1999) indicating that *N-myc*, like its family member *c-Myc*, is a human oncogene.

Promoter insertions in the oncogene Evi-1

Proviruses inserted in the promoter region of cellular genes may cause gene activation provided that they are inserted in the same transcriptional orientation as the gene. This separates the gene from its cellular promoter and places it under control of the viral promoters present in the 5′ or 3′LTR. Read-through of the transcription machinery from the virus promoter, past the 3′LTR polyadenylation signal and into the cellular gene results in formation of chimeric transcripts (Herman and Coffin, 1986; Swain and Coffin, 1993). Viral splice donors are normally used to form subgenomic sized viral transcripts encoding viral protein components. In chimeric transcripts, these viral splice donors and cryptic splice donors are spliced to splice acceptors or cryptic splice acceptors in the first introns or noncoding exons of the cellular gene. These mature, chimeric, transcripts result in increased levels of the wild-type protein (Martin-Hernandez *et al*., 2001; Mikkers *et al*., 2002a, b; Wotton *et al*., 2002). Deletion of the 5′LTR is frequently found when cellular gene transcription is regulated by the 3′ viral LTR, suggesting that removal of the 5′LTR is required for efficient transcription from the 3′LTR (Cullen *et al*., 1984; Westaway *et al*., 1984; Wotton *et al*., 2002). Promoter insertions of the MLV provirus responsible for gene activation have been found in, for example, the *LCK*, *N-Ras* and *E2a* genes (Voronova *et al*., 1987; Martin-Hernandez *et al*., 2001; Mikkers *et al*., 2002a, b). Perhaps more importantly, retroviral promoter insertions have been instrumental in the identification of Evi-1 (for ecotropic viral insertion site 1) as a potential oncogene both in mice and in humans (Figure 4) (Morishita *et al*., 1988; Mucenski *et al*., 1988; reviewed in Hirai *et al*., 2001). Evi-1 can act both as a positive and negative regulator of gene expression, however, so far, its function is not completely understood. Although it is hardly expressed in normal murine or human haematopoietic cells, elevated expression of Evi-1 is found in several haematopoietic malignancies and may even be involved in other tumour types (reviewed by Hirai *et al*., 2001). The oncogenic effects of Evi-1 overexpression in lymphomagenesis have been attributed to its negative effect on terminal differentiation of haematopoietic cells via repression of transcription (Morishita *et al*., 1992; Kreider *et al*., 1993; Palmer *et al*., 2001).

Translocations on chromosome 3 fusing the *Evi-1* gene to the *Runx1* (or *AML-1*) gene have been found in patients suffering from acute myeloid leukaemia. These translocations result in the production of a chimeric protein that can induce malignant transformation of haematopoietic cells (reviewed by Cameron and Neil, 2004; Mitani, 2004). Runx1 is one of the three members of a family of transcription factors involved in regulation of normal development and tissue-specific gene expression, which all have been linked to haematopoietic malignancies in humans. Interestingly, the *Runx1, 2* and *3* genes are also frequently found to be targeted by retroviral insertions resulting in enhanced transcription in retroviral mutagenesis screens (reviewed by Lund and van Lohuizen, 2002; Cameron and Neil, 2004). *Runx2* was originally identified as the gene targeted by insertions found in the til-1 region in *CD-2 Myc* transgenic mice (Stewart *et al*., 1997), and *Runx1* and *Runx3* have also been shown to be mutated in *CD-2 Myc* transgenic mice by promoter and enhancer insertions, respectively (Stewart *et al*., 2002; Wotton *et al*., 2002).

Targeting of tumour suppressors: the case of p53 and NF1

Proviruses that insert into a cellular gene can prematurely terminate the gene's transcript, either by using the polyadenylation signal of the LTR or by using a cryptic polyadenylation signal present in the antisense orientation of the virus. Premature termination may result in an mRNA encoding either a truncated, inactive or unstable mutant protein or in an active variant protein (discussed in the next section). Alternatively, landing of a provirus within a gene may induce aberrant splicing events, thereby abrogating gene function.

Table 1 Cancer genes that map to common insertion sites

Symbol	*Name*	*Tumour types*
BCL10	B-cell CLL/lymphoma 10	Mucosa-associated lymphoid tissue lymphoma
BCL11A	B-cell CLL/lymphoma 11A	B-cell chronic lymphatic leukaemia
CBFA2T3	Core-binding factor, runt domain, alpha subunit 2; translocated to, 3 (MTG-16)	Acute myelogenous leukaemia
CCND1	Cyclin D1	Chronic lymphatic leukaemia, B-cell acute lymphocytic leukaemia, breast
CCND2	Cyclin D2	Non-Hodgkin's lymphoma, chronic lymphatic leukaemia
CCND3	Cyclin D3	Multiple myeloma
DDX6	DEAD (Asp-Glu-Ala-Asp) box polypeptide 6	B-cell non-Hodkin's lymphoma
EPS15	Epidermal growth factor receptor pathway substrate 15 (AF1p)	Acute lymphocytic leukaemia
EVI1	Ecotropic viral integration site 1	Acute myelogenous leukaemia, chronic myeloid leukaemia
FANCC	Fanconi anaemia, complementation group C	Acute myelogenous leukaemia, leukaemia
FCGR2B	Fc fragment of IgG, low affinity IIb, receptor for (CD32)	Acute lymphocytic leukaemia
FLI1	Friend leukaemia virus integration 1	Ewings sarcoma
FLT3	fms-related tyrosine kinase 3	Acute myelogenous leukaemia, acute lymphocytic leukaemia
FNBP1	Formin binding protein 1 (FBP17)	Acute myelogenous leukaemia
GPC3	Glypican 3	Wilms tumour,
HMGA1	High mobility group AT-hook 1	Microfollicular thyroid adenoma, various benign mesenchymal tumours
HOXA9	Homeo box A9	Acute myelogenous leukaemia
HRAS	v-Ha-ras Harvey rat sarcoma viral oncogene homolog	Infrequent sarcomas, rare other types
KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog	Gastrointestinal stromal tumour, acute myelogenous leukaemia, testicular germ cell tumour, mastocytosis, epithelioma,
KRAS2	v-Ki-ras2 Kirsten rat sarcoma 2 viral oncogene homolog	Pancreatic, colorectal, lung, thyroid, acute myelogenous leukaemia, others
LASP1	LIM and SH3 protein 1	Acute myelogenous leukaemia
LMO2	LIM domain only 2 (rhombotin-like 1) (RBTN2)	T-cell acute lymphocytic leukaemia
MEN1	Multiple endocrine neoplasia type 1 gene	Parathyroid tumours, parathyroid adenoma, pituitary adenoma, pancreatic islet cell, carcinoid,
MYC	v-myc myelocytomatosis viral oncogene homolog (avian)	Burkitt lymphoma, amplified in other cancers, B-cell Lymphocytic leukaemia
MYCN	v-myc myelocytomatosis viral related oncogene, neuroblastoma derived (avian)	Neuroblastoma
NCOA2	Nuclear receptor coactivator 2 (TIF2)	Acute myelogenous leukaemia
NF1	Neurofibromatosis type 1 gene	Neurofibroma, glioma,
NFKB2	Nuclear factor of kappa light polypeptide gene enhancer in B-cells 2 (p49/p100)	B-cell non-Hodkin's lymphoma
NOTCH1	Notch homolog 1, translocation-associated (Drosophila) (TAN1)	T-cell acute lymphocytic leukaemia
NRAS	Neuroblastoma RAS viral (v-ras) oncogene homolog	Melanoma, multiple myeloma, acute myelogenous leukaemia, thyroid
PAX5	Paired box gene 5 (B-cell lineage specific activator protein)	Non Hodgkin lymphoma
PIM1	Pim-1 oncogene	Non-Hodgkin's lymphoma
PLAG1	Pleiomorphic adenoma gene 1	Salivary adenoma
PRDM16	PR domain containing 16	Myelodysplastic syndrome, acute myelogenous leukaemia
RAD51L1	RAD51-like 1 (*S. cerevisiae*) (RAD51B)	lipoma, uterine leiomyoma
REL	v-rel reticuloendotheliosis viral oncogene homolog (avian)	Hodgkin's lymphoma
RUNX1	Runt-related transcription factor 1 (AML1)	Acute myelogenous leukaemia, pre-B-cell acute lymphocytic leukaemia
SET	SET translocation	Acute myelogenous leukaemia
SFRS3	Splicing factor, arginine/serine-rich 3	Follicular lymphoma
TCF12	Transcription factor 12 (HTF4, helix–loop–helix transcription factors 4)	Extraskeletal myxoid chondrosarcoma
TP53	Tumour protein p53	Breast, colorectal, lung, sarcoma, adrenocortical, glioma, multiple other tumour types,

The table is a subset of the set of human genes known to be causally mutated in cancer, as collated and curated by the Cancer Genome Project of the Wellcome Trust Sanger Institute http://www.sanger.ac.uk/genetics/CGP/Census/(Futreal *et al*., 2004). At the time of submission, of the 343 genes present in the Cancer Gene Census, 41 of these loci were also identified within the list of 358 common insertion sites present in the RTCGD http://RTCGD.ncifcrf.gov.

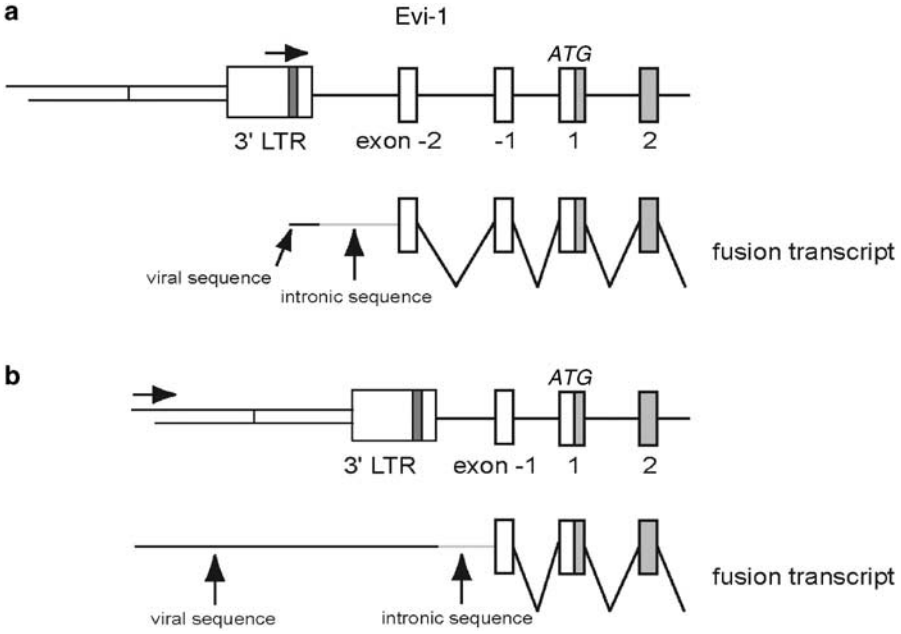

Figure 4 Promoter insertions activate the *Evi-1* proto-oncogene. Proviral insertions in the sense orientation that are either upstream (**a**) or within the first noncoding exons (**b**) of the *Evi-1* gene result in viral-cellular fusion transcripts. Transcriptional bypass of the proviral polyA signal in the 3′LTR and subsequent transcription into the *Evi-1* gene creates mRNAs that contain viral, intronic (in grey) and cellular Evi-1 sequences. The fusion transcript contains all Evi-1 coding sequences and therefore gives rise to a normal, wild-type protein. Exons are drawn as boxes (coding sequences in grey). The transcriptional and translational start site of the *Evi-1* gene is indicated (horizontal arrow and ATG, respectively)

Viral integrations inactivating the p53 tumour suppressor have been found in erythroleukaemic cell lines derived from spleens from mice infected with Friend leukaemia virus (Mowat *et al*., 1985; Ben-David *et al*., 1988). Insertions in the gene resulted in high expression of mutant proteins or in a complete absence of p53 mRNA, indicating that second copy of the p53 gene has also been mutated by proviral insertions or other mechanisms. (Mowat *et al*., 1985; Ben-David *et al*., 1988, 1990). In a similar fashion, the Neurofibromatosis-1 tumour suppressor gene, which encodes the neurofibromin GTP-ase activating protein, is found inactivated by proviral insertions of MLV and a murine AIDS virus-related endogenous virus (Cho *et al*., 1995; Largaespada *et al*., 1995). A high percentage of cell lines derived from these tumours possessed insertions in both *NF1* alleles, and expressed only truncated NF1 transcripts and no stable, full-length neurofibromin. The same results were obtained with cell lines derived from the MLV-induced tumours with a proviral insertion in only one of the *NF1* alleles, indicating that the wt allele has become inactivated by other means in these cell lines (Largaespada *et al*., 1995).

Although other cases of gene disruption by viral insertions have been reported (e.g. by the Feline Immunodeficiency Virus, Beatty *et al*., 2002), inactivating mutations represent a minority of the oncogenic mutations identified in insertion mutagenesis screens, because both gene copies of a tumour suppressor must become inactivated to abolish protein activity. However, this does not hold for insertions that induce haploinsufficiency, or inactivation of one allele while retaining a wild-type allele (reviewed by Cook and McCaw, 2000; Quon and Berns, 2001). To our knowledge, no data have been reported on proviral inactivation of haploinsufficient tumour suppressors; however, results from a genetic screen in zebrafish suggest that screens for haploinsufficient tumour suppressors may yield interesting results (Amsterdam *et al*., 2004). In this paper, zebrafish were constructed to be heterozygous for recessive embryonic lethal mutations and screened for tumour development. Zebrafish heterozygous for ribosomal protein genes showed elevated cancer incidence, while tumours isolated from these heterozygous zebrafish did not show loss-of-heterozygosity or mutations in the second allele of the ribosomal genes. However, tumours expressed lower levels of the normal transcript than wt cells, indicating that reduced levels of ribosomal proteins may be oncogenic, and suggesting that these genes do not represent classic tumour-suppressors that require loss-of-heterozygosity but rather may act as haplo-insufficient tumour suppressors (Amsterdam *et al*., 2004).

For tumours carrying hemizygously deleted tumour suppressors, the absence of detectable mutations in the second tumour suppressor allele has often been attributed to other inactivating mechanisms as, for example, epigenetic silencing. However, this may also indicate that these tumour suppressors show haplo-insufficient effects while retaining the wild-type allele. Therefore, many tumour suppressor genes that show haplo-insufficient effects in tumorigenesis could still be

unidentified as such (reviewed by Quon and Berns, 2001).

Retroviral insertion mutagenesis screens may be very useful for identification of haplo-insufficient tumour suppressor genes, since it allows efficient identification of recurrent retroviral insertions within genes. However, to verify potential haplo-insufficient tumour suppressors, tumours with mutated tumour suppressors must be analysed to determine whether an intact second allele is producing wild-type transcripts.

Activation by protein truncation: Notch1 and Tpl-2

Transcript truncation by proviral insertions may also produce a functional protein with oncogenic properties by deletion of either NH_2- or COOH-terminal domains. The proto-oncogene *Notch1*, which is mutated in many human T-cell acute lymphoblastic leukaemias (Ellisen *et al*., 1991; Weng *et al*., 2004), has been identified as a recurrent insertion site in MLV-induced T-cell leukaemia (Girard *et al*., 1996). *Notch1* encodes a transmembrane receptor involved in normal T-cell development, which is cleaved upon binding of its ligand, thereby generating an intracellular Notch1 protein that translocates to the nucleus to activate transcription (reviewed by Radtke *et al*., 2004). Proviral insertions result in two different classes of Notch mutants that closely mimic the effect of mutations that have been found in human T-ALLs (Girard *et al*., 1996; Hoemann *et al*., 2000; Weng *et al*., 2004). Moreover, it has been shown that overexpression of one of the provirally induced mutations accelerates lymphomagenesis in cancer-predisposed mice (Feldman *et al*., 2000).

Mutations in the *Tpl2* gene in MLV-induced lymphoma provide another example of insertions creating a truncated oncogenic version of a cellular protein (Makris *et al*., 1993; Patriotis *et al*., 1993; Mikkers *et al*., 2002a, b; Lund *et al*., 2002). *Tpl-2* encodes a MAP kinase kinase kinase that can activate the ERK, JNK and p38 MAPK pathways and is required for continuous upregulation of genes involved in innate immune responses (Dumitru *et al*., 2000; Eliopoulos *et al*., 2002). Proviral insertions frequently mutate the *Tpl-2* gene in MLV-induced lymphoma and always target the last intron of the gene resulting in strongly increased levels of *Tpl-2* mRNA levels, most likely due to removal of destabilizing motifs in the 3′UTR (Figure 5) (Ceci *et al*., 1997). In addition, the insertions create a carboxy-terminally truncated Tpl-2 that is dramatically more potent than its wild-type counterpart, because negative regulation of Tpl-2 kinase activity by binding of the NF-kB1 p105 protein to its C-terminal activity is abrogated (Beinke *et al*., 2003). Overexpression in mice of this C-terminally deleted Tpl-2, but not of wild-type Tpl-2, produced large-cell lymphoblastic lymphomas of T-cell origin, thus showing that truncating mutants of Tpl-2 are oncogenic *in vivo* (Ceci *et al*., 1997).

Advances in cloning of insertion sites

Recent technical developments have increased not only the number of insertion sites that can be cloned per tumour but also improved the speed, efficiency and reliability with which they are cloned. In the earliest screens, insertion sites were isolated and characterized

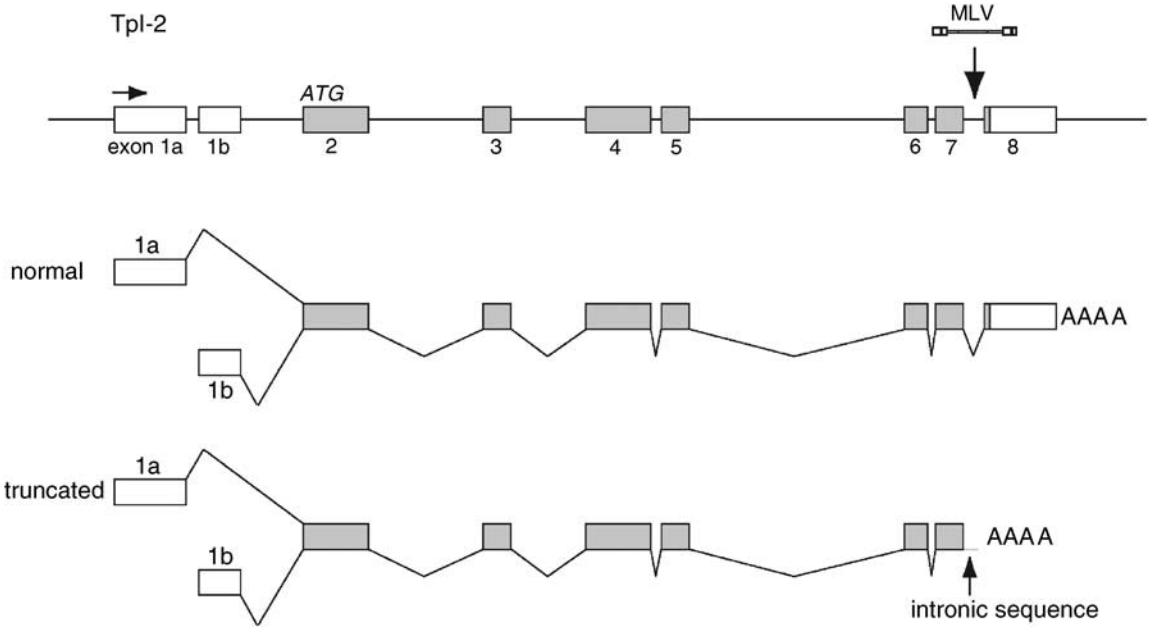

Figure 5 Oncogenic truncation of *Tpl-2* by proviral insertions. MLV insertions between the last two exons of the *Tpl-2* gene (exons 7 and 8) terminate gene transcription and thus induce the formation of a truncated mRNA. The truncated mRNA encodes a mutant Tpl-2 protein with strongly increased kinase activity. The *Tpl-2* gene consists of 7 coding (exons 2–8) and two noncoding exons (exons 1a and 1b). As depicted, splicing will incorporate either of the two noncoding exons in *Tpl-2* mRNA. Exons are drawn as boxes (coding sequences in grey). The transcriptional and translational start site of the *TpL-2* gene is indicated (horizontal arrow and ATG, respectively). The vertical arrow between exons 7 and 8 indicates the MLV insertion site

by a combination of Southern blot analysis and genomic library screening (Hayward *et al*., 1981; Tsichlis *et al*., 1983; Corcoran *et al*., 1984; Cuypers *et al*., 1984; Selten *et al*., 1984). Lambda libraries were constructed from tumour DNA and screened using viral sequences as probes. This approach was subsequently improved by engineering a tRNA amber suppressor selectable marker into the virus genome. The virus could then be used in combination with a lambda vector carrying amber mutations, to enrich libraries for virus encoding DNA fragments (Reik *et al*., 1985). Once insertion sites had been identified within a tumour panel, these and other known proto-oncogene loci could then be screened in a more efficient manner, that is, by Southern blotting of tumour DNAs to detect insertions within these loci.

Over the past two decades, these strategies have been superceded by PCR-based methods that were developed for insertional screening in a range of organisms. These methods enrich for insertion site DNA by specifically amplifying sequences flanking the insertion site. One of these, inverse PCR, uses restriction digested DNA template, which has been ligated to itself at low concentrations to form circular templates. Those circles that contain the insertion sequence of interest are then preferentially amplified using primers, which point outward from the insertion sequence (Ochman *et al*., 1988; Triglia *et al*., 1988) (illustrated in Figure 6a). Many different insertions can be amplified within the same reaction, although the products of inverse PCRs are limited to a 'window size' of fragments short enough to be efficiently PCRed but long enough to efficiently circularize. Dilution of the template before ligation can also result in inefficient amplification.

Linker mediated PCRs avoid the use of dilute, circular template DNA by instead using linkers ligated to the ends of digested DNA. A PCR is carried out with one primer against the insertion, and another against the linker. Numerous variants on this theme have been developed using different linkers: vectorette PCR (Riley *et al*., 1990; Arnold and Hodgson, 1991), splinkerette PCR (Devon *et al*., 1995), T-linker PCR (Yuanxin *et al*., 2003) and LAM-PCR (Schmidt *et al*., 2002; Hematti *et al*., 2004). Each of these methods has its own strategy to avoid amplification of fragments that have linkers at both ends but no insert sequence (for an explanation of splinkerette PCR see Figure 6b). Both inverse and linker mediated PCRs are dependent upon the presence of appropriate restriction sites close to the insertion site, hence both give improved coverage when employed using multiple restriction enzymes.

Methods that do not require ligation of template DNA have also been developed. Thermal Asymmetric Interlaced (TAIL) PCR enriches for target sequences

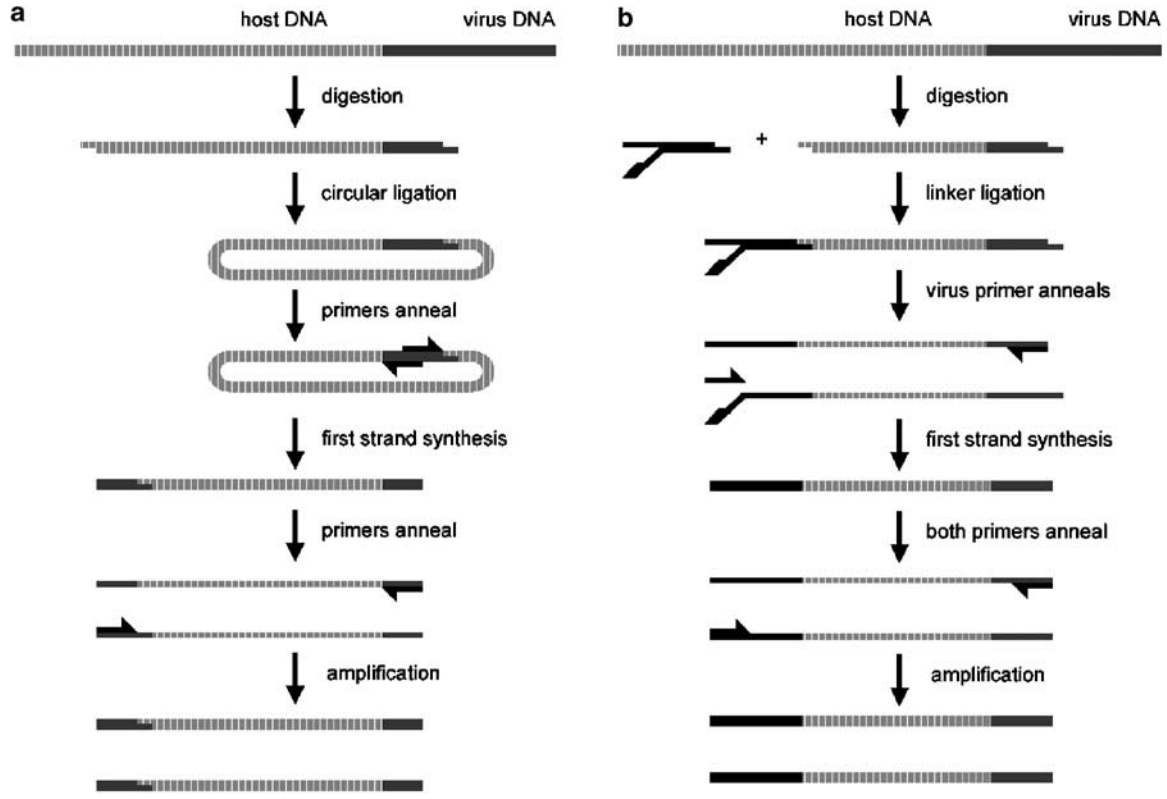

Figure 6 Two PCR based methods of insertion site cloning. (**a**) Inverse PCR. Digested DNA is ligated into circles and insert specific primers pointing outward are used to amplify the regions external to the insertion site. (**b**) Splinkerette PCR. Digested DNA is ligated to a linker containing a noncomplementary hairpin at one end. In the first round of annealing and sythesis, only the viral primer is capable of binding a template. This newly synthesized strand is complementary to the linker top strand and in the second round of synthesis, the linker primer is able to anneal to the first strand synthesis. From this point onward amplification in exponential

using a specific primer at one end and a degenerate primer at the other. Amplification occurs by alternating rounds of low, medium and high stringency (Liu and Whittier, 1995). A similar approach using a combination of specific and degenerate primers (Sorensen *et al*., 1993; Sorensen *et al*., 1996) has been combined with a secondary step using specific biotinylated primers for enrichment of products on streptavidin beads.

While the above methods are all capable of amplifying small numbers of insertions per sample, not all are equally appropriate for MLV and MMTV induced tumours since these tumours may carry a large number of insertions of varying clonality. Ideally, methods for retroviral tumour screens should efficiently amplify multiple fragments, maintaining not only the complexity of the starting material but also the relative proportions of insertions from the starting DNA. Few if any studies have been published comparing different methods, but linker mediated approaches appear to dominate the recent literature. Further improvements in specificity and yield might also be achieved by combining different portions of the above methods or adapting some of these to use non-PCR methods of amplification (such as strand displacement amplification (SDA), rolling-circle amplification (RCA) and multiple-displacement amplification (MDA) (Walker *et al*., 1992; Lizardi *et al*., 1998; Dean *et al*., 2002).

Fragment separation

Regardless of the strategy used, PCR fragments cannot be sequenced until they are separated from a mixture, usually on agarose or polyacrylamide gels. The eluted products are sequenced directly or following a re-amplification step. Similarly sized fragments that comigrate on a gel cannot be separated from each other, thus multiple PCR templates prepared using different enzymes should be analysed for each tumour. This selective gel extraction has the advantage of enriching for sequences that represent the majority of the PCR, but as a consequence lacks the sensitivity required to find insertions that are either rare in the original DNA or else poorly PCR amplified. Gel extraction is also somewhat laborious and selection of fragments for elution introduces a subjective bias. Separation alternatives such as HPLC or capillary electrophoresis might be one way of increasing the throughput and reducing the bias of this approach.

Another approach to fragment separation is to subclone the PCR products directly into a vector, either by cloning gel eluted fragments or by shotgun subcloning the entire mixture (Suzuki *et al*., 2002; Johansson *et al*., 2004). Plasmid DNA is isolated from single bacterial colonies and individual clones sequenced, thus allowing complex mixtures of fragments, including those of similar size, to be readily separated. With enough sequencing capacity, subcloning the entire mixture allows reproducible identification of sequences that represent only a tiny proportion of the PCR (our unpublished observations). Shotgun subcloning is also more amenable to automation than separation on gels, creating opportunities for increasing the scale and throughput of screens. Depending on the complexity of starting material, however, significant redundancy of sequencing may occur, thus more sequencing may be required to get the same level of coverage. Including a subcloning step might also introduce a bias against fragments that are not readily propagated in bacteria.

Other possible approaches

Different insertion sites could also be separated from each other during the PCR stage rather than after. Instead of using one or two PCRs to clone all insertions, amplification could be carried out by a larger series of separate PCRs, each of which only amplifies one of many potential templates. This might be achieved by using a series of restriction enzymes (or a single enzyme generating degenerate overhangs) to generate DNA fragments with different overhangs that can be ligated to a series of linkers with matching overhangs. Alternately, a series of PCR primers with primers that differ at their 3′ end might be used to specifically amplify subsets of insertions independently from the same ligation mixture.

Non-PCR methods such as embedding selectable markers within the virus genome might eventually replace PCR-based enrichment. This approach has been successfully employed using the *SupF* marker to construct lambda libraries from tumour DNA (Reik *et al*., 1985). Using markers that are selectable as plasmids rather than phage might allow a shotgun subcloning based-approach that skips the PCR stage. The concept of a plasmid-based shuttle vector is by no means novel (Cepko *et al*., 1984); however, its use for retroviral screens has been hindered by the fact that large inserts tend to interfere with replication competence of the virus. Even the 200bp tRNA amber suppressor fragment of the *SupF* virus has a tendency to be lost during *in vivo* viral replication. Presumably, the modified virus recombines with endogenous retroviruses, giving rise to non-tRNA carrying offspring strains that compete with the parent strain.

Replication defective viruses that express oncogenes are in some cases also capable of inducing tumours if they are transduced into haematopoietic precursors that are subsequently used to reconstitute depleted bone marrow (discussed later). Some replication defective strains have also been used to induce tumours by coinfection with a replication competent strain (Jaenisch, 1980; Johansson *et al*., 2004). The replication competent strain acts as a 'helper' strain, although predictably the competent strain tends to dominate the resulting viremia in these animals, meaning that cloning insertions of the replication defective virus would only yield a subset of the total insertion sites. Alternately, propagating replication defective viruses *in vivo* might be achieved by transgenic expression of any or all of the viral *gag pol*

and *env* transcripts, although it is perhaps difficult to predict how such transgenes might affect viremia kinetics and/or interact with endogenous retroviral like sequences.

Mapping and analysis

Improvements in techniques for insert cloning and sequencing has enabled studies which clone hundreds of insertion sites (Li *et al*., 1999; Hansen *et al*., 2000; Hwang *et al*., 2002; Joosten *et al*., 2002; Lund *et al*., 2002; Johansson *et al*., 2004; Suzuki *et al*., 2002; Mikkers *et al*., 2002a, b). Mapping larger numbers of insertions is best handled by applications that have been optimized for finding high-quality matches for large numbers of sequences relatively rapidly such as BLAT (Kent, 2002) or SSAHA (Ning *et al*., 2001). Once mapped the viral/genome junction is used to determine the exact position and orientation of proviral insertions. The precision with which individual sequences can be mapped varies, and is dependent upon the length of the PCR and quality of sequencing. Theoretically, fragments less than 20 base pairs can be mapped uniquely, in practice however, many high-quality longer sequences cannot be placed unambiguously because they map equally well to multiple repetitive elements throughout the genome.

The volume of data produced by some of these larger scale screens has prompted establishment of the Retroviral Tagged Cancer Gene Database (RTCGD, http://RTCGD.ncifcrf.gov), containing a collection of miscellaneous retroviral insertion screens (Akagi *et al*., 2004). At the time of its publication, this database contained approximately 3000 insertions derived from 1000 tumours, and this tally is periodically updated to include tumour sets from more recent publications. This resource not only creates the opportunity for multiple data sets to be analysed as one, but also facilitates comparison between screens conducted in different strains and genotypes. Many of these data sets have also been annotated onto the mouse genome browsers at UCSC (http://genome.ucsc.edu/ also linked from RTCGD) and also the Ensembl genome browser (http://www.ensembl.org/Mus_musculus/) of the European Bioinformatics Institute and Wellcome Trust Sanger Institute.

There is, however, still potential for improvement of these resources. Insertion sites could be coupled with more thorough and comprehensive data from each tumour (such as detailed tumour histology, mouse characteristics, virus characteristics, literature of prior studies validating the locus). Global comparisons between different sets of insertions or specific searches using multiple tumour characteristics simultaneously might also be useful (for example, searching to compare the difference between insertions found in male and female mice with B-cell lymphomas on a wild-type background). Also mapping insertions to the syntenic regions of the human genome and coupling this information to the location of known human oncogenes should facilitate identification of which genes are affected by proviral insertion sites.

Defining common insertion sites

Common insertion sites (also called common integration sites or CISs) are loci that have been found to contain proviral insertions in multiple tumours. When insertions cloned from multiple independent tumours are found in the same locus, this overlap is unlikely to be due to chance but rather a result of selective expansion of tumour cells carrying insertions at this locus. If a moderate number of insertions were distributed randomly over the genome without any selection, it is unlikely that two or more insertions would occur within the vicinity of the same gene. As the number of insertions in a study increases, however, the chance that two insertions might be found within the same locus also increases. To date, the largest published sets of insertions number in the hundreds, and combining insertion sets from mutliple sources (as is done in the RTCGD), allows indentification of novel rare CISs and improves the resolution of known common insertion sites. However, with so many insertions, deciding whether a CIS might have occurred due to chance requires increasingly sophisticated statistical methods.

Parameters required for these analyses are genome size, the total number of insertions and the number of tumours from which they were derived. These parameters can be used to model a random distribution of insertions over the genome, and experimental deviations from this distribution can have a *P*-value attributed to them. However, these analyses are perhaps overly simplistic since they presume that retrovirus insertions will be randomly distributed throughout the genome prior to selection.

Measuring insertional biases

Discriminating CISs that have arisen by oncogenic selection from those representing insertional hotspots requires some knowledge of bias present in the starting population of insertions. Retroviruses, retrotransposons and transposons have all been shown to posess local and regional preferences for where they integrate in the genome (reviewed in Bushman, 2003). At least some of these trends are thought to have evolved as a means of optimizing integration, expression, replication and/or excision.

Retroviral insertion biases have been estimated by a variety of methods. Early *in vivo* studies showed a preference of MLV for insertion close to DNAseI sensitive and/or hypomethylated regions (van der Putten *et al*., 1982; Vijaya *et al*., 1986; Rohdewohld *et al*., 1987) suggesting that oncogenic retroviruses had a tendency to insert within actively transcribed regions of the genome. More recently, high-throughput cloning of insertion

sites from cell lines and primary tissues has examined insertion site preference with greater resolution. Insertions from MLV and HIV-1 infected cell lines (Schroder *et al*., 2002; Wu *et al*., 2003; Mitchell *et al*., 2004) and MLV and SIV infected primate haematopoietic precursors (Hematti *et al*., 2004) indicate that while both viruses preferentially insert into transcribed regions of the genome, MLV has a tendency to insert near gene promoters whereas lentiviruses integrate further downstream within transcriptional units. HIV-1 also appears to have a bias against CpG islands. By contrast, avian sarcoma-leukosis virus displays a weak preference for active genes and no preference for transcription start regions (Mitchell *et al*., 2004).

These trends can be correlated with (and in some cases attributed to) the target DNA's local characteristics, including DNA sequence, conformation and methylation status, gene density, chromatin conformation, host DNA associated proteins and local transcriptional activity. For example, *in vitro* studies using Moloney MLV indicate that local DNA target availability is partially affected by nucleosome structure (Pryciak and Varmus, 1992). Similarly, the preintegration complexes of mammalian retroviruses are known to associate with host expressed proteins such as mobility group protein A1 (HMG-A1), barrier-to-autointegration factor (BAF) and lamina-associated polypeptide 2 alpha (LAP2 alpha) (Li *et al*., 2000; Lin and Engelman, 2003; Suzuki *et al*., 2004).

It seems likely that each combination of virus and host cell type will have its own unique insertional 'fingerprint' that can only be determined empirically by sequencing large numbers of independent unselected insertions. In the absence of this data, however, estimations could be based on each virus' known local insertion preferences extrapolated to the entire genome using information such as predicted gene boundaries, measurements of transcription levels, chromatin structure and distributions of chromatin associated proteins. Once a distribution pattern is measured and estimated, this could be used to adjust significance estimates of common insertion sites. Theoretical distributions used for modelling random insertions could then be modified to account for local background insertion density estimates, rather than estimates that presume random spread of insertions over the entire genome. Alternately, background insertion density could be used to define a minimum density threshold for common insertion sites to be considered significant.

Interpreting common insertion sites

Identifying which gene within a CIS is playing a causal role in oncogenesis is not simple, particularly in cases where expression of multiple genes is affected by an insertion (Hanlon *et al*., 2003). Tumours harbouring common insertion sites are characterized initially by examining the expression and/or transcript size of candidate genes, and by verifying insert clonality by southern blot to see if the insertion is present within a substantial proportion of the tumour. *In vitro* assays for proliferation, apoptosis suppression, transformation, colony forming and anchorage independent growth can give an indication of a CIS candidate gene's role in oncogenesis, but ultimately the most convincing proof is the use of animal models of cancer, and/or by identification of functionally equivalent mutations in human cancers.

CIS interactions

Once CISs have been defined (either as loci or as specific genes), they can be correlated with other tumour characteristics such as host genotype, background strain, tumour histology and the presence or absence of other CISs within the same tumour. Tumours bearing a *c-Myc* insertion have a strong tendency to carry mutations in the *Pim1*, *Gfi1*, *Ahi1*/*Myb*, *Rras2* and *Cyclin D1* loci, suggesting the contribution of these loci toward tumorigenesis are complementary to *c-Myc* overexpression. Conversely, tumours carrying *c-Myc* insertions do not apppear to carry insertions within the *Sox4*, *Hhex* and *Rasgrp1* loci. The significance of the co-occurrence of insertion sites can be verified using transgenic mice showing these genes collaborate in oncogenesis. Cooperation of *c-Myc* with *Pim1* (van Lohuizen *et al*., 1989; Verbeek *et al*., 1991) or Gfi-1 (Schmidt *et al*., 1998) has been demonstrated this way.

This phenomenon of insertion site cooperation has also been observed in a dramatic case of inadvertent insertional mutagenesis in humans. MLV infected AKxD mice have two known insertions in the *LMO2* locus, one of which is associated with an insertion near the common gamma chain of the interleukin 2 receptor (*IL2RG*)(Dave *et al*., 2004). A similar association has been observed between these genes during a gene therapy clinical trial for X-linked severe combined immunodeficiency (X-linked SCID). In this trial, a series of 10 infant X-SCID patients had their haematopoietic precursors infected with retroviral based gene therapy vector expressing the interleukin receptor common gamma chain (Hacein-Bey-Abina *et al*., 2002; Cavazzana-Calvo *et al*., 2005; Schmidt *et al*., 2005). This treatment effectively cured nine of the 10 patients; however, within 3 years several of the patients were found to have developed T-cells leukaemia, where a copy of the retroviral vector had inserted near the *LMO2* promoter (Marshall, 2003; Hacein-Bey-Abina *et al*., 2003a, b). This case demonstrates the need for caution in the design of gene therapy vectors, but also suggests that the safety of such vectors might be testable using animal models.

One potential application for identifying novel oncogene interactions might be in finding candidates for development of cancer therapeutics. Mutation of the *c-Myc* oncogene is one of the most common events in human lymphoma and leukaemia, but it is currently not considered a druggable target. Loss of alleles of *c-Myc*

collaborators such as the polycomb group protein *Bmi1* or the kinase *Pim-1* causes a delay in tumour onset induced by a *c-Myc* transgene and/or MLV (van der Lugt *et al*., 1995; Jacobs *et al*., 1999a, b). By inference, if a *c-Myc* collaborator was found to be a druggable target, such a drug might have clinically useful effects on lymphomas that depend on *c-Myc* for disease progression. Conversely, if an oncogene found in retroviral screens is known to be a druggable target, identifying frequent collaborators may help in the profiling of patient tumour samples to determine which targeted therapeutics are appropriate, and may even suggest novel indications for existing therapies.

Using slow transforming retroviruses to generate other tumour types

The utility of slow transforming retroviruses for the study of cancer is limited by the range of tumours they induce. Deviations from MLVs propensity for T-lymphomagenesis can be induced by expression of oncogenic transgenes. The *Emu-Myc* transgene (*c-Myc* expressed from a CMV promoter/Emu enhancer) gives rise to lymphomas of a B-cell lineage and expands the pre-B-cell compartment (Adams *et al*., 1985; Langdon *et al*., 1986). When these transgenics are infected with Moloney MLV, lymphomagenesis is accelerated and 90% of these lymphomas are B cell rather than the T cell lymphomas typical of Moloney MLV (Haupt *et al*., 1991; van Lohuizen *et al*., 1991). MLVs can also be influenced by the background strain of the host, the recombinant inbred strain BXH-2 giving rise to myeloid leukaemias (Bedigian *et al*., 1984) and some AKxD strains showing a bias toward B-cell lymphomas (Gilbert *et al*., 1993).

MLVs can also induce tumours in non lymphoid/myeloid cell types by a combination of tissue specific virus delivery and viral driven oncogene expression (Uhrbom *et al*., 1998). Tumours with the characteristics of human glioblastoma can be induced by expressing the platelet derived growth factor B-chain (PDGFB) from a replication incompetent Moloney MLV based retrovirus and delivering this virus to the brains of neonatal mice by intracerebral injection (Uhrbom *et al*., 1998; Johansson *et al*., 2004). Autocrine activation of PDGF receptors is an important event in the genesis of human glioblastoma. Propagation of this PDGFB virus *in vivo* is enhanced by coinjection/coinfection with replication competent Moloney MLV. Cloning insertion sites from these tumours (using the LTR of both viruses) has identified loci known to be involved in human cancer as well as many previously identified from retrovirally induced leukaemias and lymphomas (Johansson *et al*., 2004).

A similar strategy might also be employed using retroviruses that express short hairpin RNA inhibitors of tumour suppressor genes in combination with MLV based mutagenesis. Short hairpin RNAi retroviral vectors have been used for stable germline and somatic knockdown of gene expression (Rubinson *et al*., 2003) and can be targeted against tumour supressors to accellerate lymphoma development (Hemann *et al*., 2003). The small size of siRNA transcripts might even allow development of replication competent viruses, eliminating the need for coinfection with helper strains.

Mutations of retroviral genomes (either spontaneously arising or artificially engineered) can also have a substantial influence on the tumour types that arise. This specificity is frequently attributable the LTR regions (DesGroseillers and Jolicoeur, 1984a, b) although in some cases subtle mutations can also alter the tropism of the virus envelope proteins (Bahrami *et al*., 2004). A series of variants of the SL3-3 murine leukaemia virus that normally induces T-cell lymphomas have been shown to have altered oncogenic capabilities. A range of LTR mutations alter tumour latency (Ethelberg *et al*., 1997a, b; Ma *et al*., 2003), alteration of SL3-3 enhancer factor 1 sites skews tumour formation away from thymic lymphoma (Hallberg *et al*., 1991) and mutation of Runx sites severely extends latency, inhibiting T-lymphomas and causing myeloid and to a lesser extent B-lymphoid and erythroid tumours (Sorensen *et al*., 2004).

Manipulated viruses with reduced replication competence have a tendency to lose the changes made to them by recombination with related endogenous retrovirus like sequences. Thus, while it would seem there is untapped potential in the range of tumour types that manipulated slow transforming viruses can induce, the inherent tropism of these viruses and the limited replication competence and stability of manipulated viruses illustrates a need for alternate insertional mutagens to study other tumour types.

Transposable elements as insertional mutagens

Obvious candidates for developing novel insertional mutagens are transposable elements. These elements fall into two groups, those which use reverse transcriptase to transpose via an RNA intermediate and those which transpose in a 'cut and paste' manner as a DNA intermediate. Both types have been used as insertional mutagens in a range of organisms and recent developments have increased their utility in mammalian systems (for reviews see Hamer *et al*., 2001; Largaespada, 2003).

Retrotransposons

Retrotransposons can be further subdivided into two main classes. The retroviral-like group possesses LTR and reverse transcriptase (although they encode no env protein) and their mode of replication and insertion is thought to be similar to retroviruses. The transcript is reverse transcribed and converted to ds DNA, and this fragment integrates into the genome, that is, reverse transcription and integration are separate events. The other group, known as Non-LTR retrotransposons or Poly(A) retrotransposons initiate reverse transcription

from a single-strand break in recipient DNA which is used as a primer for the RNA transcript to be reverse transcribed. Thus, the reverse transcription event is coupled to integration.

Of this second group, long interspersed elements (LINEs) are the most prevalent in mammals, with LINE-1 (L1) elements comprising about 17% of the human genome. Recently, L1 elements have been investigated for their potential as germ line insertional mutagens in mammals (Ostertag *et al*., 2002) although the frequency of transposition events has been limited by the fact that endogenous L1 elements are inefficiently transcribed (Han *et al*., 2004). Codon optimization of the L1 coding regions has removed this inhibition of transcription, resulting in increased L1 protein levels and a 200 fold increase in transposition activity (Han and Boeke, 2004). Whether this magnitude of increase will be sufficient for the use of L1 transposons as oncogenic somatic mutagens remains to be seen, however human L1 elements have recently been shown to be capable of mobilization in rodent neuronal precursor cells, in some cases altering gene expression and cell fate (Muotri *et al*., 2005).

DNA Transposons

Transposable elements that move via a DNA intermediate (excising themselves from one site and integrating at another) have been used as insertional mutagens in plants, *Drosophila* and *Caenorhabditis elegans*. Most consist of transposase gene flanked by terminal inverted repeats. Some demonstrate a requirement for species specific host factors while others are active between species and in some cases between kingdoms (reviewed in Plasterk *et al*., 1999). Endogenous cut and paste transposons have been found in vertebrates though most if not all appear to have degenerated due to loss of function mutations. This lack of active endogenous cut and paste transposons has prompted the engineering and enhancement of a number of elements for use in verterbrates. A number of Tc1/mariner family transposons have been shown to have activity in mammalian cells, including Tc1 and Tc3 from *C. elegans*, MosI from *D. mauritania*, Himar I from *Haematobia irritans* and Minos from *D. hydei* (reviewed in Largaespada, 2003). Similarly the *piggyBac* DNA transposon of lepidopterans has been demonstrated to mobilize in human and mouse cell lines and in the mouse germ line (Ding *et al*., 2005).

An alternative to adapting invertebrate transposons for use in vertebrates is to create synthetic transposons based on remnant degenerate transposon sequences found dispersed throughout vertebrate genomes. Sleeping Beauty (SB) is a synthetic transposon derived from ancient Tc1/mariner-like sequences of salmonid fish species. A consensus sequence was deduced from the inactive remnant transposon sequences of fish genomes, and a functional version was synthesized and named Sleeping Beauty 10 (SB10) (Ivics *et al*., 1997). The newly awakened transposon was shown to be active not only in fish cells, but also human and mouse cells. More recently a similar approach has also been used to reconstruct 'Frog Prince', a related transposon from ancient *Xenopus* sequences (Miskey *et al*., 2003).

The SB element consists of a transposon open reading frame flanked by inverted repeat/direct repeat (IR/DR) elements. The transposase protein functions as a tetramer, with a dimer binding each of the IR/DRs. Point mutations which increase transposition activity have been identified either by inference from related transposases or by alanine scanning substitutions (SB11 (Geurts *et al*., 2003), SB12 (Zayed *et al*., 2004), HSB1, HSB2, HSB3, HSB4 (Yant *et al*., 2004)). Increase transposition frequency has also been achieved by specific point mutations (pT2 (Cui *et al*., 2002)) or by direct replacement of the right IRDR with the left IRDR (pT3 (Yant *et al*., 2004)).

These improvements in transposase activity and transposon mobility have increased the utility of SB for several applications. Nonviral gene therapy vectors based on SB have been developed and tested for somatic expression in mouse liver (Yant *et al*., 2000, 2002) and lung endothelial cells (Liu *et al*., 2004). SB based vectors have been developed for transgenesis in zebrafish (Davidson *et al*., 2003) and Medaka (Grabher *et al*., 2003). Germ line mutagenesis has also been attempted by coupling expression of the transposase with transgenes carrying a modified transposon (Fischer *et al*., 2001) and a gene trap transposon (Carlson *et al*., 2003). A gene trap germ line mutagensis approach has also been developed in zebrafish (Clark *et al*., 2004).

Two recent papers report successful use of SB as an oncogenic insertional mutagen in somatic tissues (Collier *et al*., 2005; Dupuy *et al*., 2005). In both papers, an SB transposase transgene was expressed in mice, which also carry concatamers of gene trap transposons. These gene traps are designed to be activating and inactivating mutagens much like slow transforming retroviruses. They contain an MSCV promoter coupled to a *FoxF2* splice donor in order to create fusion transcripts, which will drive expression of genes downstream of the insertion site. The transposons also contain splice acceptor sites coupled to a bidirectional SV40 polyadenylation signal, thus allowing for premature termination of transcripts running into the transposon. It is not yet known whether the IR/DR elements themselves might have transcriptional or enhancer activity on surrounding genes, although none has yet been reported.

In one of these papers (Collier *et al*., 2005) the SB10 transposase is expressed constitutively as a transgene driven by the CAGGs promoter. These mice do not develop tumours spontaneously; however, mobilization of the transposase appears to accelerate the onset of sarcomas in *p19 ARF* null mice. Insert analyses of these tumours reveal the presence of a prevalent common insertion site on chromosome 6 in the region of *BRAF*. These insertions create a truncated transcript, which while not equivalent to BRAF mutations seen in human melanoma (Davies *et al*., 2002), is nonetheless capable of inducing foci formation in NIH 3T3 cells.

In the other paper (Dupuy *et al*., 2005) a more active version of the SB transposase, SB11, was integrated into the Rosa-26 locus. These mice were crossed onto a strain carrying close to 100 copies of a gene trap transposon. This combination of transposase and concatemer is apparently more potent since offspring of these crosses frequently die *in utero*, presumably because of the mutational load they are subject to during development. Those mice which survive until birth, succumb to cancer rapidly, in most cases of T-cell lymphomas but also medulloblastomas and some pituitary and intestinal neoplasias. A common insertion site was identified from the lymphomas, these insertions introducing truncation mutations of the *Notch-1* transcript.

It remains to be seen the extent to which SB common insertion sites from haematopoietic and mammary tumours will either complement or substantially overlap with those induced by retroviruses. Transposons, like retroviruses, are not random in their integration, with specificity determined by host or element encoded factors (reviewed in Craig, 1997). Several studies have examined the insertional biases of SB and found that it has a slight bias toward sites matching a consensus ATATATAT, and a preference for DNA with a bendable structure as well as several other specific conformational characteristics (Vigdal *et al*., 2002; Liu *et al*., 2005). Cloning larger numbers of insertion sites from cell lines has also demonstrated a slight preference for integration within the promoters of genes, similar to MLV but less pronounced (Yant *et al*., 2005). SB also demonstrated a strong bias toward microsatellite repeats, but no observable correlation with transcriptional activity.

When comparing this system to the use of slow transforming retroviruses the main advantage is its potential flexibility. The use of controlled tissue specific transposase expression should allow direction of tumorigenesis to nonlymphoid and nonmammary tissues. Altering the contents of the transposon (promoters, splice donors, splice acceptors, introns, RNA stability motifs, Kozak sequences) should also influence the types of mutations that can be created. Using different numbers of transposons will likely alter the extent to which individual insertions are selected for sequentially, or multiple cooperating insertions are selected together from the outset of tumorigenesis.

Screens based on DNA transposons are, however, prone to complications not present in retroviral screens. Excision of the SB transposon leaves behind 'foot print' mutations, thus transposons which hop more than once will contribute a background of mutations. Furthermore, insertion events, which are selected for and hence clonally expanded, are not immune to remobilization and may be lost in some cells during tumour growth. This could give rise to heterogenous tumours and may complicate analysis of insertion sites.

More troubling, however, is the observation that like other Tc1/mariner transposons, Sleeping Beauty has a strong preference for jumping in *cis*. This property complicates the analysis of CISs, with those occurring on the same chromosome as the donor concatamer needing to fulfil more stringent statistical criteria in order to be considered significant. The use of multiple mouse lines carrying concatamers on different chromosomes provides a simple solution to this problem; however, a means of distributing transposition events more evenly throughout the genome would improve the utility of this system.

Conclusion

A large number of the proto-oncogene and tumour suppressor loci found from retroviral screens have also been shown to be causally involved in human tumour development (Table 1). This provides some validation for the use of these screens to model human oncogenesis, and suggests that common insertion sites are useful candidates to pursue for mutation detection studies of human tumours. Even when equivalent mutations of CISs are not present in human tumours, understanding these CISs should still help clarify oncogenic branches of signaling networks and may even provide valuable targets for intervention downstream of mutated oncogenes.

As insertion site cloning methods improve and screens achieve greater saturation, this should allow better identification of not only oncogene gain of function mutations, but also rarer tumour suppressor and haploinsufficient tumour suppressor loss of function mutations. Screening of nonhaematopoietic tissues will be particularly valuable for the study of tumour types who's genetic aetiology is still poorly understood. Thus, although the principle of retroviral insertion mutagenesis has been around for quite a while now, its future holds promise.

Acknowledgements
Thanks to Colin Pritchard and Jos Jonkers for critically reading the manuscript and providing helpful suggestions.

References

Adams JM, Harris AW, Pinkert CA, Corcoran LM, Alexander WS, Cory S, Palmiter RD and Brinster RL. (1985). *Nature*, **318,** 533–538.

Aho TL, Sandholm J, Peltola KJ, Mankonen HP, Lilly M and Koskinen PJ. (2004). *FEBS Lett.*, **571,** 43–49.

Akagi K, Suzuki T, Stephens RM, Jenkins NA and Copeland NG. (2004). *Nucleic Acids Res.*, **32,** D523–D527.

Amsterdam A, Sadler KC, Lai K, Farrington S, Bronson RT, Lees JA and Hopkins N. (2004). *PLoS Biol.*, **2,** E139.

Arnold C and Hodgson IJ. (1991). *PCR Methods Appl.*, **1,** 39–42.

Bahrami S, Duch M and Pedersen FS. (2004). *J. Virol.*, **78,** 9343–9351.

Beatty J, Terry A, MacDonald J, Gault E, Cevario S, O'Brien SJ, Cameron E and Neil JC. (2002). *Cancer Res.*, **62,** 7175–7180.
Bedigian HG, Johnson DA, Jenkins NA, Copeland NG and Evans R. (1984). *J. Virol.*, **51,** 586–594.
Beinke S, Deka J, Lang V, Belich MP, Walker PA, Howell S, Smerdon SJ, Gamblin SJ and Ley SC. (2003). *Mol. Cell. Biol.*, **23,** 4739–4752.
Ben David Y, Lavigueur A, Cheong GY and Bernstein A. (1990). *New Biol.*, **2,** 1015–1023.
Ben David Y, Prideaux VR, Chow V, Benchimol S and Bernstein A. (1988). *Oncogene*, **3,** 179–185.
Bushman FD. (2003). *Cell*, **115,** 135–138.
Cameron ER and Neil JC. (2004). *Oncogene*, **23,** 4308–4314.
Carlson CM, Dupuy AJ, Fritz S, Roberg-Perez KJ, Fletcher CF and Largaespada DA. (2003). *Genetics*, **165,** 243–256.
Cavazzana-Calvo M, Lagresle C, Hacein-Bey-Abina S and Fischer A. (2005). *Annu. Rev. Med.*, **56,** 585–602.
Ceci JD, Patriotis CP, Tsatsanis C, Makris AM, Kovatch R, Swing DA, Jenkins NA, Tsichlis PN and Copeland NG. (1997). *Genes Dev.*, **11,** 688–700.
Cepko CL, Roberts BE and Mulligan RC. (1984). *Cell*, **37,** 1053–1062.
Cho BC, Shaughnessy Jr JD, Largaespada DA, Bedigian HG, Buchberg AM, Jenkins NA and Copeland NG. (1995). *J. Virol.*, **69,** 7138–7146.
Clark KJ, Geurts AM, Bell JB and Hackett PB. (2004). *Genesis*, **39,** 225–233.
Clausse N, Baines D, Moore R, Brookes S, Dickson C and Peters G. (1993). *Virology*, **194,** 157–165.
Collier LS, Carlson CM, Ravimohan S, Dupuy AJ and Largaespada DA. (2005). *Nature*, **436,** 272–276.
Cook WD and McCaw BJ. (2000). *Oncogene*, **19,** 3434–3438.
Corcoran LM, Adams JM, Dunn AR and Cory S. (1984). *Cell*, **37,** 113–122.
Craig NL. (1997). *Annu. Rev. Biochem.*, **66,** 437–474.
Cui Z, Geurts AM, Liu G, Kaufman CD and Hackett PB. (2002). *J. Mol. Biol.*, **318,** 1221–1235.
Cullen BR, Lomedico PT and Ju G. (1984). *Nature*, **307,** 241–245.
Cuypers HT, Selten G, Quint W, Zijlstra M, Maandag ER, Boelens W, van Wezenbeek P, Melief C and Berns A. (1984). *Cell*, **37,** 141–150.
Dave UP, Jenkins NA and Copeland NG. (2004). *Science*, **303,** 333.
Davidson AE, Balciunas D, Mohn D, Shaffer J, Hermanson S, Sivasubbu S, Cliff MP, Hackett PB and Ekker SC. (2003). *Dev. Biol.*, **263,** 191–202.
Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W, Davis N, Dicks E, Ewing R, Floyd Y, Gray K, Hall S, Hawes R, Hughes J, Kosmidou V, Menzies A, Mould C, Parker A, Stevens C, Watt S, Hooper S, Wilson R, Jayatilake H, Gusterson BA, Cooper C, Shipley J, Hargrave D, Pritchard-Jones K, Maitland N, Chenevix-Trench G, Riggins GJ, Bigner DD, Palmieri G, Cossu A, Flanagan A, Nicholson A, Ho JW, Leung SY, Yuen ST, Weber BL, Seigler HF, Darrow TL, Paterson H, Marais R, Marshall CJ, Wooster R, Stratton MR and Futreal PA. (2002). *Nature*, **417,** 949–954.
Dean FB, Hosono S, Fang L, Wu X, Faruqi AF, Bray-Ward P, Sun Z, Zong Q, Du Y, Du J, Driscoll M, Song W, Kingsmore SF, Egholm M and Lasken RS. (2002). *Proc. Natl. Acad. Sci. USA*, **99,** 5261–5266.
DesGroseillers L and Jolicoeur P. (1984a). *J. Virol.*, **52,** 448–456.
DesGroseillers L and Jolicoeur P. (1984b). *J. Virol.*, **52,** 945–952.
Devon RS, Porteous DJ and Brookes AJ. (1995). *Nucleic Acids Res.*, **23,** 1644–1645.
Ding S, Wu X, Li G, Han M, Zhuang Y and Xu T. (2005). *Cell*, **122,** 473–483.
Dhanasekaran SM, Barrette TR, Ghosh D, Shah R, Varambally S, Kurachi K, Pienta KJ, Rubin MA and Chinnaiyan AM. (2001). *Nature*, **412,** 822–826.
Domen J, Von Lindern M, Hermans A, Breuer M, Grosveld G and Berns A. (1987). *Oncogene Res.*, **1,** 103–112.
Dudley JP, Mertz JA, Rajan L, Lozano M and Broussard DR. (2002). *Leukemia*, **16,** 1086–1098.
Dumitru CD, Ceci JD, Tsatsanis C, Kontoyiannis D, Stamatakis K, Lin JH, Patriotis C, Jenkins NA, Copeland NG, Kollias G and Tsichlis PN. (2000). *Cell*, **103,** 1071–1083.
Dupuy AJ, Akagi K, Largaespada DA, Copeland NG and Jenkins NA. (2005). *Nature*, **436,** 221–226.
Eischen CM, Weber JD, Roussel MF, Sherr CJ and Cleveland JL. (1999). *Genes Dev.*, **13,** 2658–2669.
Eliopoulos AG, Dumitru CD, Wang CC, Cho J and Tsichlis PN. (2002). *EMBO J.*, **21,** 4831–4840.
Ellisen LW, Bird J, West DC, Soreng AL, Reynolds TC, Smith SD and Sklar J. (1991). *Cell*, **66,** 649–661.
Ellwood-Yen K, Graeber TG, Wongvipat J, Iruela-Arispe ML, Zhang J, Matusik R, Thomas GV and Sawyers CL. (2003). *Cancer*, **4,** 223–238.
Erkeland SJ, Valkhof M, Heijmans-Antonissen C, Delwel R, Valk PJ, Hermans MH and Touw IP. (2003). *Blood*, **101,** 1111–1117.
Erkeland SJ, Valkhof M, Heijmans-Antonissen C, Hoven-Beijen A, Delwel R, Hermans MH and Touw IP. (2004). *J. Virol.*, **78,** 1971–1980.
Ethelberg S, Hallberg B, Lovmand J, Schmidt J, Luz A, Grundstrom T and Pedersen FS. (1997a). *J. Virol.*, **71,** 1196–1206.
Ethelberg S, Lovmand J, Schmidt J, Luz A and Pedersen FS. (1997b). *J. Virol.*, **71,** 7273–7280.
Feldman BJ, Hampton T and Cleary ML. (2000). *Blood*, **96,** 1906–1913.
Fischer SE, Wienholds E and Plasterk RH. (2001). *Proc. Natl. Acad. Sci. USA*, **98,** 6759–6764.
Fox CJ, Hammerman PS, Cinalli RM, Master SR, Chodosh LA and Thompson CB. (2003). *Genes Dev.*, **17,** 1841–1854.
Futreal PA, Coin L, Marshall M, Down T, Hubbard T, Wooster R, Rahman N and Stratton MR. (2004). *Nat. Rev. Cancer*, **4,** 177–183.
Geurts AM, Yang Y, Clark KJ, Liu G, Cui Z, Dupuy AJ, Bell JB, Largaespada DA and Hackett PB. (2003). *Mol. Ther.*, **8,** 108–117.
Gilbert DJ, Neumann PE, Taylor BA, Jenkins NA and Copeland NG. (1993). *J. Virol.*, **67,** 2083–2090.
Girard L, Hanna Z, Beaulieu N, Hoemann CD, Simard C, Kozak CA and Jolicoeur P. (1996). *Genes Dev.*, **10,** 1930–1944.
Grabher C, Henrich T, Sasado T, Arenz A, Wittbrodt J and Furutani-Seiki M. (2003). *Gene*, **322,** 57–66.
Graves BJ, Eisenman RN and McKnight SL. (1985). *Mol. Cell. Biol.*, **5,** 1948–1958.
Grimes HL, Gilks CB, Chan TO, Porter S and Tsichlis PN. (1996). *Proc. Natl. Acad. Sci. USA*, **93,** 14569–14573.
Gunther CV and Graves BJ. (1994). *Mol. Cell. Biol.*, **14,** 7569–7580.
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, Fischer A, Davies EG, Kuis W, Leiva L and Cavazzana-Calvo M. (2002). *N. Engl. J. Med.*, **346,** 1185–1193.
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 and Fischer A. (2003a). *N. Engl. J. Med.*, **348,** 255–256.
Hacein-Bey-Abina S, Von Kalle C, Schmidt M, McCormack MP, Wulffraat N, Leboulch P, Lim A, Osborne CS, Pawliuk R, Morillon E, Sorensen R, Forster A, Fraser P, Cohen JI, de Saint Basile G, Alexander I, Wintergerst U, Frebourg T, Aurias A, Stoppa-Lyonnet D, Romana S, Radford-Weiss I, Gross F, Valensi F, Delabesse E, Macintyre E, Sigaux F, Soulier J, Leiva LE, Wissler M, Prinz C, Rabbitts TH, Le Deist F, Fischer A and Cavazzana-Calvo M. (2003b). *Science*, **302,** 415–419.
Hallberg B, Schmidt J, Luz A, Pedersen FS and Grundstrom T. (1991). *J. Virol.*, **65,** 4177–4181.
Hamer L, DeZwaan TM, Montenegro-Chamorro MV, Frank SA and Hamer JE. (2001). *Curr. Opin. Chem. Biol.*, **5,** 67–73.
Han JS and Boeke JD. (2004). *Nature*, **429,** 314–318.
Han JS, Szak ST and Boeke JD. (2004). *Nature*, **429,** 268–274.
Hanlon L, Barr NI, Blyth K, Stewart M, Haviernik P, Wolff L, Weston K, Cameron ER and Neil JC. (2003). *J. Virol.*, **77,** 1059–1068.
Hansen GM, Skapura D and Justice MJ. (2000). *Genome Res.*, **10,** 237–243.
Haupt Y, Alexander WS, Barri G, Klinken SP and Adams JM. (1991). *Cell*, **65,** 753–763.
Hayward WS, Neel BG and Astrin SM. (1981). *Nature*, **290,** 475–480.
Hemann MT, Fridman JS, Zilfou JT, Hernando E, Paddison PJ, Cordon-Cardo C, Hannon GJ and Lowe SW. (2003). *Nat. Genet.*, **33,** 396–400.
Hematti P, Hong BK, Ferguson C, Adler R, Hanawa H, Sellers S, Holt IE, Eckfeldt CE, Sharma Y, Schmidt M, von Kalle C, Persons DA, Billings EM, Verfaillie CM, Nienhuis AW, Wolfsberg TG, Dunbar CE and Calmels B. (2004). *PLoS Biol.*, **2,** e423.
Herman SA and Coffin JM. (1986). *J. Virol.*, **60,** 497–505.
Hirai H, Izutsu K, Kurokawa M and Mitani K. (2001). *Cancer Chemother. Pharmacol.*, **48** (Suppl 1)**,** S35–S40.
Hoemann CD, Beaulieu N, Girard L, Rebai N and Jolicoeur P. (2000). *Mol. Cell. Biol.*, **20,** 3831–3842.
Hwang HC, Martins CP, Bronkhorst Y, Randel E, Berns A, Fero M and Clurman BE. (2002). *Proc. Natl. Acad. Sci. USA*, **99,** 11293–11298.
Ivics Z, Hackett PB, Plasterk RH and Izsvak Z. (1997). *Cell*, **91,** 501–510.
Jacobs JJ, Kieboom K, Marino S, DePinho RA and van Lohuizen M. (1999). *Nature*, **397,** 164–168.
Jacobs JJ, Scheijen B, Voncken JW, Kieboom K, Berns A and van Lohuizen M. (1999). *Genes Dev.*, **13,** 2678–2690.
Jaenisch R. (1980). *Cell*, **19,** 181–188.
Johansson FK, Brodd J, Eklof C, Ferletta M, Hesselager G, Tiger CF, Uhrbom L and Westermark B. (2004). *Proc. Natl. Acad. Sci. USA*, **101,** 11334–11337.
Johnson C, Lobelle-Rich PA, Puetter A and Levy LS. (2005). *J. Virol.*, **79,** 57–66.
Joosten M, Vankan-Berkhoudt Y, Tas M, Lunghi M, Jenniskens Y, Parganas E, Valk PJ, Lowenberg B, van den Akker E and Delwel R. (2002). *Oncogene*, **21,** 7247–7255.
Kent WJ. (2002). *Genome Res.*, **12,** 656–664.
Kreider BL, Orkin SH and Ihle JN. (1993). *Proc. Natl. Acad. Sci. USA*, **90,** 6454–6458.
Langdon WY, Harris AW, Cory S and Adams JM. (1986). *Cell*, **47,** 11–18.
Largaespada DA, Shaughnessy Jr JD, Jenkins NA and Copeland NG. (1995). *J. Virol.*, **69,** 5095–5102.
Largaespada DA. (2003). *Reprod. Biol. Endocrinol.*, **1,** 80.
Lazo PA, Lee JS and Tsichlis PN. (1990). *Proc. Natl. Acad. Sci. USA*, **87,** 170–173.
Li J, Shen H, Himmel KL, Dupuy AJ, Largaespada DA, Nakamura T, Shaughnessy Jr JD, Jenkins NA and Copeland NG. (1999). *Nat. Genet.*, **23,** 348–353.
Li L, Yoder K, Hansen MS, Olvera J, Miller MD and Bushman FD. (2000). *J. Virol.*, **74,** 10965–10974.
Lin CW and Engelman A. (2003). *J. Virol.*, **77,** 5030–5036.
Lipsick JS and Wang DM. (1999). *Oncogene*, **18,** 3047–3055.
Liu G, Geurts AM, Yae K, Srinivasan AR, Fahrenkrug SC, Largaespada DA, Takeda J, Horie K, Olson WK and Hackett PB. (2005). *J. Mol. Biol.*, **346,** 161–173.
Liu L, Sanz S, Heggestad AD, Antharam V, Notterpek L and Fletcher BS. (2004). *Mol. Ther.*, **10,** 97–105.
Liu YG and Whittier RF. (1995). *Genomics*, **25,** 674–681.
Lizardi PM, Huang X, Zhu Z, Bray-Ward P, Thomas DC and Ward DC. (1998). *Nat. Genet.*, **19,** 225–232.
Lund AH and van Lohuizen M. (2002). *Cancer Cell*, **1,** 213–215.
Lund AH, Turner G, Trubetskoy A, Verhoeven E, Wientjens E, Hulsman D, Russell R, DePinho RA, Lenz J and van Lohuizen M. (2002). *Nat. Genet*, **32,** 160–165.
Ma SL, Lovmand J, Sorensen AB, Luz A, Schmidt J and Pedersen FS. (2003). *Virology*, **313,** 638–644.
Makris A, Patriotis C, Bear SE and Tsichlis PN. (1993). *J. Virol.*, **67,** 4283–4289.
Marshall E. (2003). *Science*, **299,** 320.
Martin-Hernandez J, Sorensen AB and Pedersen FS. (2001). *J. Virol.*, **75,** 11907–11912.
Mikkers H, Allen J and Berns A. (2002a). *Oncogene*, **21,** 6559–6566.
Mikkers H, Allen J, Knipscheer P, Romeijn L, Hart A, Vink E and Berns A. (2002b). *Nat. Genet.*, **32,** 153–159.
Mikkers H and Berns A. (2003). *Adv. Cancer Res.*, **88,** 53–99.
Mikkers H, Nawijn M, Allen J, Brouwers C, Verhoeven E, Jonkers J and Berns A. (2004). *Mol. Cell. Biol.*, **24,** 6104–6115.
Miskey C, Izsvak Z, Plasterk RH and Ivics Z. (2003). *Nucleic Acids Res.*, **31,** 6873–6881.
Mitani K. (2004). *Oncogene*, **23,** 4263–4269.
Mitchell RS, Beitzel BF, Schroder AR, Shinn P, Chen H, Berry CC, Ecker JR and Bushman FD. (2004). *PLoS Biol.*, **2,** E234.
Morishita K, Parganas E, Matsugi T and Ihle JN. (1992). *Mol. Cell. Biol.*, **12,** 183–189.
Morishita K, Parker DS, Mucenski ML, Jenkins NA, Copeland NG and Ihle JN. (1988). *Cell*, **54,** 831–840.
Mowat M, Cheng A, Kimura N, Bernstein A and Benchimol S. (1985). *Nature*, **314,** 633–636.
Mucenski ML, Taylor BA, Ihle JN, Hartley JW, Morse HC, III, Jenkins NA and Copeland NG. (1988). *Mol. Cell. Biol.*, **8,** 301–308.
Muotri AR, Chu VT, Marchetto MCN, Deng W, Moran JV and Gage FH. (2005). *Nature*, **435,** 903–910.
Nesbit CE, Tersak JM and Prochownik EV. (1999). *Oncogene*, **18,** 3004–3016.
Nielsen AA, Sorensen AB, Schmidt J and Pedersen FS. (2005). *J. Virol.*, **79,** 67–78.
Ning Z, Cox AJ and Mullikin JC. (2001). *Genome Res.*, **11,** 1725–1729.

Ochman H, Gerber AS and Hartl DL. (1988). *Genetics*, **120,** 621–623.
Ostertag EM, DeBerardinis RJ, Goodier JL, Zhang Y, Yang N, Gerton GL and Kazazian Jr HH. (2002). *Nat. Genet.*, **32,** 655–660.
Palmer S, Brouillet JP, Kilbey A, Fulton R, Walker M, Crossley M and Bartholomew C. (2001). *J. Biol. Chem.*, **276,** 25834–25840.
Patriotis C, Makris A, Bear SE and Tsichlis PN. (1993). *Proc. Natl. Acad. Sci. USA*, **90,** 2251–2255.
Pelengaris S, Khan M and Evan G. (2002). *Nat. Rev. Cancer*, **2,** 764–776.
Plasterk RH, Izsvak Z and Ivics Z. (1999). *Trends Genet.*, **15,** 326–332.
Pryciak PM and Varmus HE. (1992). *Cell*, **69,** 769–780.
Quon KC and Berns A. (2001). *Genes Dev.*, **15,** 2917–2921.
Radtke F, Wilson A, Mancini SJ and MacDonald HR. (2004). *Nat. Immunol.*, **5,** 247–253.
Reik W, Weiher H and Jaenisch R. (1985). *Proc. Natl. Acad. Sci. USA*, **82,** 1141–1145.
Reisman D. (1990). *FEBS Lett.*, **277,** 209–211.
Riley J, Butler R, Ogilvie D, Finniear R, Jenner D, Powell S, Anand R, Smith JC and Markham AF. (1990). *Nucleic Acids Res.*, **18,** 2887–2890.
Roh M, Gary B, Song C, Said-Al-Naief N, Tousson A, Kraft A, Eltoum IE and Abdulkadir SA. (2003). *Cancer Res.*, **63,** 8079–8084.
Rohdewohld H, Weiher H, Reik W, Jaenisch R and Breindl M. (1987). *J. Virol.*, **61,** 336–343.
Rosenberg N and Jolicoeur P. (1997). *Retroviruses*, Coffin JM, Hughes SH and Varmus HE (eds). Cold Spring Harbor Laboratory Press: Cold Spring Harbor, NY, pp. 475–585.
Rubinson DA, Dillon CP, Kwiatkowski AV, Sievers C, Yang L, Kopinja J, Rooney DL, Ihrig MM, McManus MT, Gertler FB, Scott ML and Van Parijs L. (2003). *Nat. Genet.*, **33,** 401–406.
Schmidt M, Hacein-Bey-Abina S, Wissler M, Carlier F, Lim A, Prinz C, Glimm H, Andre-Schmutz I, Hue C, Garrigue A, Le Deist F, Lagresle C, Fischer A, Cavazzana-Calvo M and von Kalle C. (2005). *Blood*, **105,** 2699–2706.
Schmidt M, Zickler P, Hoffmann G, Haas S, Wissler M, Muessig A, Tisdale JF, Kuramoto K, Andrews RG, Wu T, Kiem HP, Dunbar CE and von Kalle C. (2002). *Blood*, **100,** 2737–2743.
Schmidt T, Karsunky H, Gau E, Zevnik B, Elsasser HP and Moroy T. (1998). *Oncogene*, **17,** 2661–2667.
Schroder AR, Shinn P, Chen H, Berry C, Ecker JR and Bushman F. (2002). *Cell*, **110,** 521–529.
Selten G, Cuypers HT and Berns A. (1985). *EMBO J.*, **4,** 1793–1798.
Selten G, Cuypers HT, Zijlstra M, Melief C and Berns A. (1984). *EMBO J.*, **3,** 3215–3222.
Shore SK, Tantravahi RV and Reddy EP. (2002). *Oncogene*, **21,** 8568–8576.
Sorensen AB, Duch M, Amtoft HW, Jorgensen P and Pedersen FS. (1996). *J. Virol.*, **70,** 4063–4070.
Sorensen AB, Duch M, Jorgensen P and Pedersen FS. (1993). *J. Virol.*, **67,** 7118–7124.
Sorensen KD, Quintanilla-Martinez L, Kunder S, Schmidt J and Pedersen FS. (2004). *J. Virol.*, **78,** 13216–13231.
Starkey CR, Lobelle-Rich PA, Granger S, Brightman BK, Fan H and Levy LS. (1998). *J. Virol.*, **72,** 1078–1084.
Stewart M, Mackay N, Cameron ER and Neil JC. (2002). *J. Virol.*, **76,** 4364–4369.
Stewart M, Terry A, Hu M, O'Hara M, Blyth K, Baxter E, Cameron E, Onions DE and Neil JC. (1997). *Proc. Natl. Acad. Sci. USA*, **94,** 8646–8651.
Sun W, Graves BJ and Speck NA. (1995). *J. Virol.*, **69,** 4941–4949.
Suzuki T, Shen H, Akagi K, Morse HC, Malley JD, Naiman DQ, Jenkins NA and Copeland NG. (2002). *Nat. Genet.*, **32,** 166–174.
Suzuki Y, Yang H and Craigie R. (2004). *EMBO J.*, **23,** 4670–4678.
Swain A and Coffin JM. (1993). *J. Virol.*, **67,** 6265–6269.
Triglia T, Peterson MG and Kemp DJ. (1988). *Nucleic Acids Res.*, **16,** 8186.
Tsichlis PN, Strauss PG and Hu LF. (1983). *Nature*, **302,** 445–449.
Uhrbom L, Hesselager G, Nister M and Westermark B. (1998). *Cancer Res.*, **58,** 5275–5279.
van der Lugt NM, Domen J, Verhoeven E, Linders K, van der Gulden H, Allen J and Berns A. (1995). *EMBO J.*, **14,** 2536–2544.
van der Putten H, Quint W, Verma IM and Berns A. (1982). *Nucleic Acids Res.*, **10,** 577–592.
van Lohuizen M, Breuer M and Berns A. (1989). *EMBO J.*, **8,** 133–136.
van Lohuizen M, Verbeek S, Krimpenfort P, Domen J, Saris C, Radaszkiewicz T and Berns A. (1989). *Cell*, **56,** 673–682.
van Lohuizen M, Verbeek S, Scheijen B, Wientjens E, van der Gulden H and Berns A. (1991). *Cell*, **65,** 737–752.
Verbeek S, van Lohuizen M, van der Valk M, Domen J, Kraal G and Berns A. (1991). *Mol. Cell Biol.*, **11,** 1176–1179.
Vigdal TJ, Kaufman CD, Izsvak Z, Voytas DF and Ivics Z. (2002). *J. Mol. Biol.*, **323,** 441–452.
Vijaya S, Steffen DL and Robinson HL. (1986). *J. Virol.*, **60,** 683–692.
Voronova AF, Adler HT and Sefton BM. (1987). *Mol. Cell. Biol.*, **7,** 4407–4413.
Walker GT, Little MC, Nadeau JG and Shank DD. (1992). *Proc. Natl. Acad. Sci. USA*, **89,** 392–396.
Weng AP, Ferrando AA, Lee W, Morris JP, Silverman LB, Sanchez-Irizarry C, Blacklow SC, Look AT and Aster JC. (2004). *Science*, **306,** 269–271.
West AG and Fraser P. (2005). *Hum. Mol. Genet.*, **14** (Suppl 1), R101–R111.
Westaway D, Payne G and Varmus HE. (1984). *Proc. Natl. Acad. Sci. USA*, **81,** 843–847.
Wotton S, Stewart M, Blyth K, Vaillant F, Kilbey A, Neil JC and Cameron ER. (2002). *Cancer Res.*, **62,** 7181–7185.
Wu X, Li Y, Crise B and Burgess SM. (2003). *Science*, **300,** 1749–1751.
Yant SR, Ehrhardt A, Mikkelsen JG, Meuse L, Pham T and Kay MA. (2002). *Nat. Biotechnol.*, **20,** 999–1005.
Yant SR, Meuse L, Chiu W, Ivics Z, Izsvak Z and Kay MA. (2000). *Nat. Genet.*, **25,** 35–41.
Yant SR, Park J, Huang Y, Mikkelsen JG and Kay MA. (2004). *Mol. Cell. Biol.*, **24,** 9239–9247.
Yant SR, Wu X, Huang Y, Garrison B, Burgess SM and Kay MA. (2005). *Mol. Cell. Biol.*, **25,** 2085–2094.
Yuanxin Y, Chengcai A, Li L, Jiayu G, Guihong T and Zhangliang C. (2003). *Nucleic Acids Res.*, **31,** e68.
Zayed H, Izsvak Z, Walisko O and Ivics Z. (2004). *Mol. Ther.*, **9,** 292–304.
Zindy F, Eischen CM, Randle DH, Kamijo T, Cleveland JL, Sherr CJ and Roussel MF. (1998). *Genes Dev.*, **12,** 2424–2433.