4

Gene Trapping: a Multi-Purpose Tool for Functional Genomics

ALESSANDRA GENTILE, LORENZA D'ALESSANDRO AND ENZO MEDICO*

Institute for Cancer Research and Treatment, University of Torino Medical School, S.P. 142, km. 3.95, 10060 Candiolo (TO), Italy

Introduction

Biomedical research is currently facing a historical change in the perspectives and modality of gathering information about biological processes and gene function. Genome sequencing projects have been completed, or are in advanced progress for many organisms, including *Homo sapiens* (Lander *et al.*, 2001; Venter *et al.*, 2001; http://www.ncbi.nlm.nih.gov/genome/guide/human/) and *Mus musculus*, for which a genome physical map has just been completed (Gregory *et al.*, 2002; http://www.ncbi.nlm.nih.gov/genome/guide/mouse/). This has reversed the conventional approach to biomedical discovery, in which understanding a certain biological function required identification of one or more genes involved in that function. The current situation is that thousands of genes have been sequenced, but still await functional information to be assigned to them. The fact that genes of unknown function represent over 70% of all genes suggests that our current comprehension of most biological and pathological processes is far from complete. In this perspective, systematic exploration of gene function is likely to yield a huge amount of information in the forthcoming years.

*To whom correspondence may be addressed: (enzo.medico@ircc.it)

Abbreviations: β-gal, β-galactosidase; CDS, coding sequence; ES, embryonic stem; FACS, fluorescence-activated cell sorting; GFP, green fluorescent protein; GM-CSF, granulocyte-macrophage colony stimulating factor; HGF, hepatocyte growth factor; HIV, human immunodeficiency virus; HSV-TK, herpes simplex virus thymidine kinase; IRES, internal ribosome entry site; LTR, long terminal repeat; MN, metronidazole; neo, neomycin phosphotransferase; NTR, nitroreductase; PCR, polymerase chain reaction; PGK, phosphoglycerate kinase; PLAP, placental alkaline phosphatase; poly(A), polyadenylation; puro, puromycin N-acetyltransferase; RA, retinoic acid; RACE, rapid amplification of cDNA ends; ROSA, reverse-oriented splice acceptor; RT, reverse transcription; SA, splice acceptor; SD, splice donor; SS, signal sequence; UAS, upstream activating sequence; UTR, un-translated region; X-gal, 5-bromo-4-chloro-3-indolyl β-D-galactoside.

Biotechnology and Genetic Engineering Reviews - Vol. 20, December 2003
0264-8725/03/20/77-100 $20.00 + $0.00 © Intercept Ltd, PO Box 716, Andover, Hampshire, SP10 1YG, UK

There are several ways of gathering information about gene function, some of which have been evolving at an incredibly high pace. For example, the relatively recent development of DNA microarray technology currently has enabled parallel mRNA expression analysis for thousands of genes (Schena *et al.*, 1995; Lockhart *et al.*, 1996). Indeed, expression (at least at the RNA level) is an essential prerequisite for a gene to exert its function, and by studying the sites and pathways of regulation of a particular gene expression it is possible to putatively assign it to a broad functional group (Eisen *et al.*, 1998). In this view, genes with restricted, tissue-specific expression are likely to play key roles in the biochemical and biological processes occurring specifically at the expression sites.

Another powerful approach to gene functional characterization is the exploration of the consequences of gene loss-of-function in various model organisms, ranging from unicellular microorganisms to invertebrates, vertebrates, and mammals. In particular, generation of mutations in murine ES cells by targeted and random approaches offers a powerful tool for loss-of-function studies in the mouse. ES cells can be grown *in vitro* as a continuous cell line, genetically modified, and subsequently returned to the embryo, where they can generate chimeric mice, and eventually contribute to the germ line. Mouse ES cells are now widely used for gene disruption by homologous recombination (Thomas and Capecchi, 1987), or chemically induced mutagenesis (Justice, 2000), to create mutant mice that lack or express an altered form of a specific gene. Recently, an International Mouse Mutagenesis Consortium has been established, with the long-term goal of producing at least one heritable mutation, in either ES cells or mice, for every gene in the mouse genome (Nadeau *et al.*, 2001). In many cases, however, functional redundancy, or subtle phenotypes, may impair functional characterization of the targeted genes. Moreover, this approach is aimed at defining gene function in the context of the organism, but will be hard to direct at exploring basic biological and biochemical functions at the cellular level. This latter type of information can be achieved by systematic screening programmes exploring particular features of the gene protein product, like subcellular localization (Gonzalez and Bejarano, 2000), biochemical activity (Zhu *et al.*, 2000), interactions (Suzuki *et al.*, 2001), and others. Recently, the development of small interfering RNA (siRNA)-based approaches has rendered loss-of-function studies more easily practicable in cell lines and higher organisms (Brummelkamp *et al.*, 2002; Lewis *et al.*, 2002). Finally, genes can be characterized by gain-of-function approaches, relying on overexpression of cloned genes in cells and organisms (Medico *et al.*, 2001a), or on random activation of gene expression (RAGE; Harrington *et al.*, 2001).

From this brief outline of the major strategies for gene functional characterization, it is clear that a crucial issue in functional genomics is the development of technologies for high throughput functional analysis. In this perspective, gene traps are emerging as powerful tools, and for three principal reasons: a) gene trapping does not rely on previous knowledge or cloning of the target genes; b) it enables efficient, systematic gene targeting; and c) a single gene trap vector may allow exploration of multiple gene parameters, for instance expression level, protein product localization, and loss-of-function effects. Gene trapping has been performed extensively on mouse ES cells (reviewed in Stanford *et al.*, 2001), as well as on many organisms, including yeasts (Kumar *et al.*, 2002), plants (Lindsey *et al.*, 1993), *Drosophila* (O'Kane and Gehring, 1987), zebrafish (Chen *et al.*, 2002), and *Xenopus* (Bronchain *et al.*, 1999).

Table 4.1. Gene elements and vector modules exploited for trapping

Gene element	Trapping module
Enhancer	Minimal promoter
Promoter	Promoterless reporter, eventually preceded by a splice acceptor site
Coding sequence (CDS)	Reporter/selection marker lacking translation initiation site
Signal sequence (SS)	Reporter/selection marker functionally disrupted by a transmembrane domain
Poly(A) signal	Splice donor (SD), eventually followed by a destabilizing sequence

The present review is mainly focused on gene trapping applied to mouse ES cells and other mammalian cell lines.

Principles of gene trapping

The founding principle of gene trapping is stable, random integration throughout the cellular genome of a DNA construct capable of highlighting integration within functional genes. To this aim, gene trap vectors transduce specific modules composed of reporters, or selection marker, lacking a specific functional feature, and therefore normally inactive. Full functionality of these modules is only acquired when they have properly integrated into genes that provide the missing function. Following this principle, we will propose a classification of the gene traps based on the gene functional elements exploited for trapping and, as a consequence, on the gene trap module specifically designed for the task. The gene functional elements exploited for trapping are: 1) enhancers; 2) promoter; 3) splice acceptor/donor (SA/SD); 4) coding sequence (CDS); 5) signal sequence (SS); and 6) polyadenylation site. Accordingly, we define all types of trapping vectors as 'gene traps', because they exploit one or more gene elements, and subdivide the trapping modules according to the respective trapped element: 1) enhancer traps; 2) promoter traps; 3) CDS traps; 4) SS traps; and 5) poly(A) traps. This classification is summarized in *Table 4.1*, and we will now explain this in some detail.

GENE ELEMENTS AND VECTOR MODULES EXPLOITED FOR TRAPPING (*FIGURE 4.1*)

Enhancers

Gene expression is tightly regulated by a series of small sequence stretches, located in proximity of promoters, named enhancers. In most cases, a minimal promoter without enhancer elements has the ability to promote gene transcription, but its activity is extremely low. An enhancer trap is therefore constituted of a minimal or weak promoter, followed by a reporter gene and a polyadenylation site (*Figure 4.1b*). To work properly, enhancer traps have to integrate near regulatory elements, generally located at the 5'-end of the genes. An additional cassette composed of a selection marker driven by a constitutive promoter may follow the trap reporter, to enable selection of integrants.

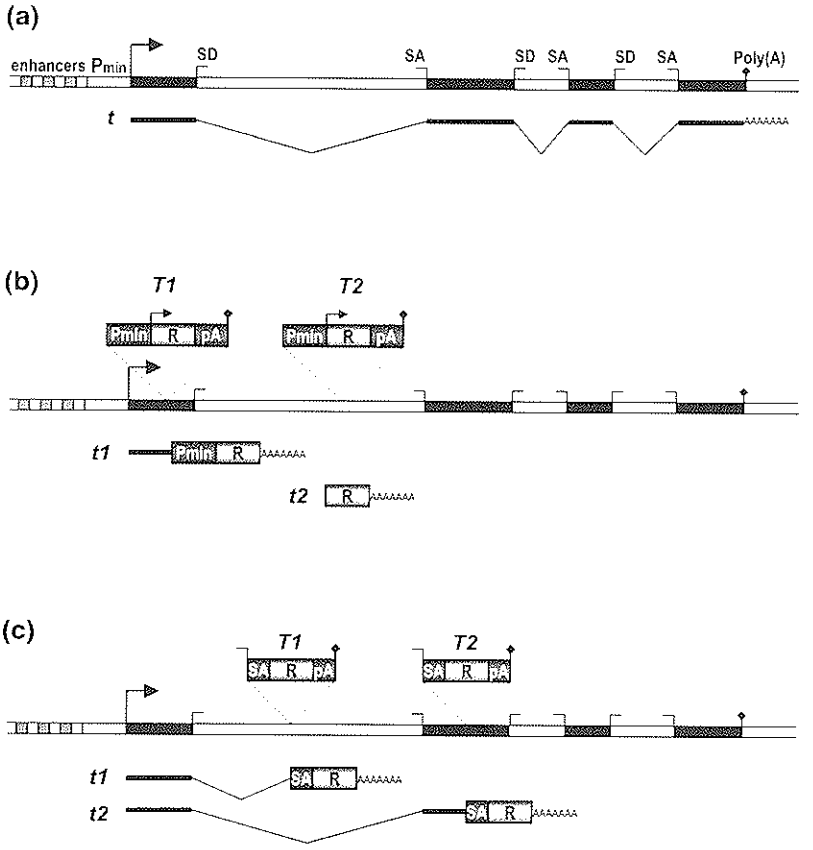

Figure 4.1. Principles of gene trapping (a–c). (a) General structure of a gene and the mRNA resulting from its transcription. Each gene has its *regulatory regions* comprising a minimal promoter (Pmin) that drives gene expression (the arrow indicates the transcription start site), one or more enhancer sequences that can be before or after the promoter (here represented only before the promoter), and a polyadenylation signal (poly(A)) that interrupts transcription and promotes RNA maturation and polyadenylation. The transcribed portion of the gene is composed of introns (filled in white) and exons (filled in black), delimited by splice donor (SD) and splice acceptor (SA) signals. The primary transcript (*t*) illustrates the splicing process, in which the exonic portions (thicker line) are joined together after removal of introns during mRNA maturation. (b) A prototype enhancer trap, composed of a minimal promoter (Pmin), a reporter gene (R), and a polyadenylation site (pA), illustrated in two of the many possible integration sites: the first exon (T1) and the first intron (T2). When the endogenous gene is transcribed and the trap is integrated in the first exon, a chimeric mRNA is generated (*t1*), composed of a 5' portion derived from the endogenous transcript, and a 3' portion derived from the trap. Upon intronic integration, which is the most frequent event, the trap promoter drives transcription of the trap reporter alone (*t2*). However, the *t2* type of transcript can also be generated by exonically integrated traps. (c) Promoter trapping by a construct composed of a splice acceptor site (SA), a reporter (R), and a polyadenylation site (pA). The trap is respectively integrated within the first intron (T1) and the second exon (T2). In both cases, trapping generates fusion transcripts (*t1* and *t2*) containing trap-derived sequence at the 3' (*t1* and *t2*). Note that in the absence of an SA site, the T1-type of integration would result in splicing out of the trap from the mature transcript. Promoter traps lacking the SA domain would only generate fusion mRNAs upon exonic integration.

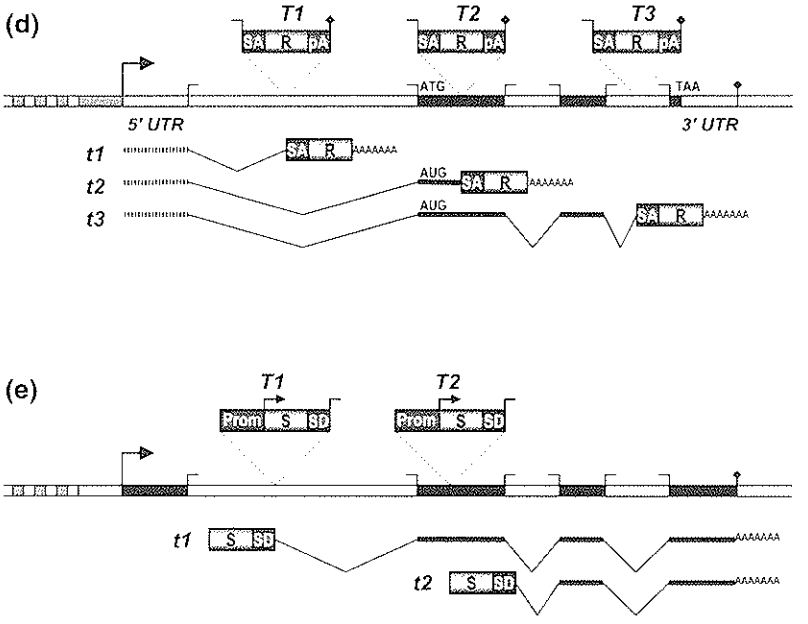

Figure 4.1. Principles of gene trapping (d–e). (d) Coding sequence trap, composed of a SA site, an ATG-less reporter (R), and a poly(A) signal, represented in three possible integrations: upstream of the endogenous gene coding sequence (T1), within the first translated exon (T2), or between two translated exons (T3). In the T1-type of integrations, trapping is inactive, because the resulting fusion transcript (*t1*) does not contain an initiating AUG upstream of the reporter, which therefore is not translated. Conversely, integrations within or between translated exons result in translation of a fusion protein composed of an N-terminal portion derived from the endogenous gene, and a C-terminal trap-derived portion. Also in this case, the SA site avoids trap splicing out upon intronic integration (T3). (e) A poly(A) trap, composed of a constitutively active promoter (Prom), followed by a selection marker (S), and a splice donor site (SD). Poly(A) trap generate fusion transcripts driven by the trap promoter (*t1* and *t2*) that therefore contain trap-derived sequence at their 5'. These transcripts are stabilized by the poly(A) signal provided by the trapped gene. The SD domain allows removal of long intronic sequences that may destabilize transcripts originating from intronic trap integration (T1).

Promoters

A promoter trap carries, in its minimal configuration, a reporter gene not preceded by any promoter sequence, and followed by a polyadenylation site. As a consequence, the reporter is transcribed only when it integrates in the right orientation downstream from an active endogenous promoter. In this case, the promoter and enhancer sequences of the endogenous gene will drive transcription of a chimeric mRNA, composed of a 5' portion deriving from the endogenous gene and a 3' portion deriving from the trap, including the reporter gene. As the transcription is interrupted at the end of the reporter gene by the trap-encoded polyadenylation site, the endogenous gene is disrupted in many cases. However, when the integration occurs in an intron, all of the trap sequence is spliced out of the mature mRNA. As a consequence, the basic promoter trap illustrated above has a very low frequency of effective gene trapping

(up to 200-fold lower than enhancer traps). To overcome this problem, most promoter traps carry a splice acceptor (SA) site immediately upstream of the reporter gene (*Figure 4.1c*), which increases trapping efficiency up to 50-fold (Friederich and Soriano, 1991).

Splice donors/acceptors

The transcribed portion of most genes is composed of segments that are retained in the mature mRNA, named exons, and segments that are lost during RNA maturation, named introns. The process of intron elimination from the primary RNA transcript is called splicing. The intron/exon boundaries are defined by specific sites, named splice acceptor (SA) and splice donor (SD) sites (*Figure 4.1a*). An SA site defines the 5'-end of an exon, and an SD site defines its 3'-end. In other words, an exon begins with an SA and ends with an SD. SA and SD sites are used extensively in gene trap vectors, to achieve proper gene trapping even when vector integration occurs in intronic regions, which is indeed much more frequent than exonic integration. In many cases, therefore, gene trap modules are artificial exons, in that they contain either an SA site at their 5'- or an SD site at their 3'-end.

Coding sequence

CDS traps are very similar to promoter traps, the main difference being that CDS traps transduce promoterless reporters lacking the translation-initiating ATG codon (*Figure 4.1d*). Translation of such reporters may only take place when the trapped gene provides the initiating AUG in the correct reading frame. To achieve proper CDS trapping upon intronic integration, the ATG-less reporter is generally preceded by an SA site. In this case, trapping requires the downstream reporter being in the same reading frame of the upstream coding exon. To this aim, three differently framed vectors can be used in parallel trapping procedures (Skarnes, 2000). Alternatively, vectors may incorporate a splice acceptor site capable of splicing in all three reading frames simultaneously, like the one derived from the murine leukaemia virus *env* gene (Brenner *et al.*, 1989), though splice acceptors of this type are generally weak.

Signal sequence

The coding sequence of most secreted and transmembrane proteins initiates with a cleavable, N-terminal signal sequence (SS). In the presence of a transmembrane (TM) domain, the protein is inserted in the cell membrane in a precise orientation, with the portion comprised between the SS and the TM domain located extracellularly. Conversely, when the TM is not preceded by an SS, the protein localization is reversed, with the C-terminal portion exposed at the outer cell surface. SS-trapping modules take advantage of this feature by containing an ATG-less ORF in which a TM domain precedes a reporter specifically sensitive to its intracellular/extracellular localization. For instance, β-gal is only active when localized intracellularly (Skarnes *et al.*, 1995). Therefore, β-gal fusion with an upstream TM domain determines extracellular localization of the β-gal domain, which results in an inactive reporter. Conversely, when the trapped gene provides an SS to the fusion protein, β-gal is

'swapped' to the cytoplasmic face of the membrane, which results in positive X-gal staining. Another strategy for SS trapping can be based on a fusion protein composed of an extracellular cell-surface antigen, followed by a TM domain and an intracellular selection marker. In the absence of an SS, the protein is inversely collocated, with the selection marker outside the cell and the antigen inside, inaccessible to antibodies. Upon SS trapping, the protein is swapped, and the extracellularly exposed antigen can be used for antibody-mediated staining and selection by panning (Gebauer *et al.*, 2001).

Polyadenylation signal

At the end of their transcribed portion, genes contain a poly(A) signal, which interrupts transcription and subsequently promotes RNA polyadenylation and stabilization. In the absence of a poly(A) signal, transcript maturation is compromised, and the RNA is promptly degraded. A poly(A) trapping vector carries a selection marker driven by a constitutive promoter and followed by a splice donor (SD) site, but lacking the polyadenylation site (*Figure 4.1e*). Therefore, the selection marker is always transcribed, but the transcript is not polyadenylated and is highly unstable. Stabilization occurs only when trap integration takes place upstream of an exon containing the poly(A) site, eventually preceded by other exons. Being driven by a constitutive promoter, poly(A) trapping occurs independently of the expression of the trapped gene. The resulting transcript generates a chimeric mRNA, composed of a 5' portion deriving from the trap and a 3' portion deriving from the endogenous gene, including the poly(A) tail. To increase the stringency of the poly(A) trapping selection, an additional RNA destabilizing sequence may be added immediately downstream to the trap splice donor site (see below). In this case, the further reduced half-life of unspliced, trap-derived transcripts will greatly diminish the probability that incorrect trapping events may give rise to basal activity (Ishida and Leder, 1999).

TRAPPING VECTOR BACKBONES

Gene trap vectors can have a plasmidic or retroviral backbone. One possible issue in using one type of vector versus the other could be the presence of backbone-specific preferential integration sites in the genome. However, recent results from large-scale gene trap screenings indicate that both types of vectors integrate in a mostly random fashion, and that the few insertional 'hot spots' identified are shared between the two vector types (Chowdury *et al.*, 1997; Wiles *et al.*, 2000). A lentiviral vector-based gene trap has also recently been described (Lai *et al.*, 2002), which may take advantage of the fact that HIV-derived vectors efficiently transduce non-dividing cells, also *in vivo* (Naldini *et al.*, 1996), and tend to integrate in expressed genes (Schroder *et al.*, 2002). An advantage of viral vectors is that they can be titrated to easily optimize high throughput, low multiplicity infections yielding single-copy integrants. Also, retroviruses tend to insert at the 5'-end of the genes, even when the trapped element is the poly(A) site (Zambrowicz *et al.*, 1998), which renders gene disruption more likely. Finally, proviral DNA always retains the LTRs during the integration process, which facilitates cloning of trap insertion sites. Conversely, plasmid-derived linear DNA is exposed to endogenous exonucleases before integration,

which may complicate procedures for obtaining trap-flanking genomic sequences. For this series of reasons, gene trap vectors with a retroviral backbone are generally preferred, although it is likely that full genome coverage by gene trapping will require the use of both types of vectors.

TRAPPED GENE LOSS OF FUNCTION

Most gene trap vectors have been designed to be mutagenic upon proper integration within functional genes. Production of a fusion transcript containing trap-derived and gene-derived sequence is generally considered a good indicator of effective gene targeting. However, a reporter cell generated by gene trapping may reliably track the activity of the upstream promoter without completely disrupting the trapped gene. In some cases, a functionally active transcript may still be expressed, which results in partial or absent loss of gene function.

The events responsible for ineffective gene targeting may differ according to the type of trapping construct, or to the integration site. The most frequent cases are illustrated in *Figure 4.2*. Intronic integration of an enhancer trap, which does not contain any SA or SD sequence, when combined with poor efficiency of the trap poly(A) site, may result in splicing out of the trap from the endogenous transcript (*Figure 4.2a*). A similar event may also occur when the SA of a promoter trap, integrated within an intron, is not fully active in a particular transcript, which may result in the concomitant generation of two transcripts, one containing the trap as its last exon, and the other excluding the trap and continuing with the downstream exons (*Figure 4.2b*). A third event may occur upon exonic integration of any kind of trap: if the exon physiologically undergoes alternative splicing, and the trap poly(A) site is not adequately efficient, the alternatively spliced mRNA retains its full functionality (*Figure 4.2c*). In all of the above cases, however, inefficient gene targeting by the trap requires multiple, concomitant adverse events that may lead to a partial loss of function deriving from dosage reduction of the WT transcripts. Finally, gene trap insertion in the 3' portion of a gene may result in the translation of a totally or partially functional protein from the endogenous ORF (*Figure 4.2d*). More aberrant events, like trans-splicing, duplications, and recombinations, may also lead to inefficient gene targeting, but these cannot be adequately schematized.

IDENTIFICATION OF TRAPPED GENES

All gene trapping approaches rely on random integration throughout the host genome. Therefore, once interesting traps have been selected, it is necessary to map the sites at which trap integration occurred, to identify the trapped genes, define how integration could disrupt the gene, and verify that the trap reporter may reliably reflect the gene transcriptional activity. Trapped gene identification can be achieved through protocols based on genomic DNA amplification, or on mRNA reverse-transcription-PCR (RT-PCR). Genomic DNA amplification allows precise mapping of the integration site, even when it occurs in an intronic region of a gene. RT-PCR protocols allow characterization of the fusion transcript derived from the trapping event. Here follows a brief overview of the most widely used techniques for trapped gene identification.

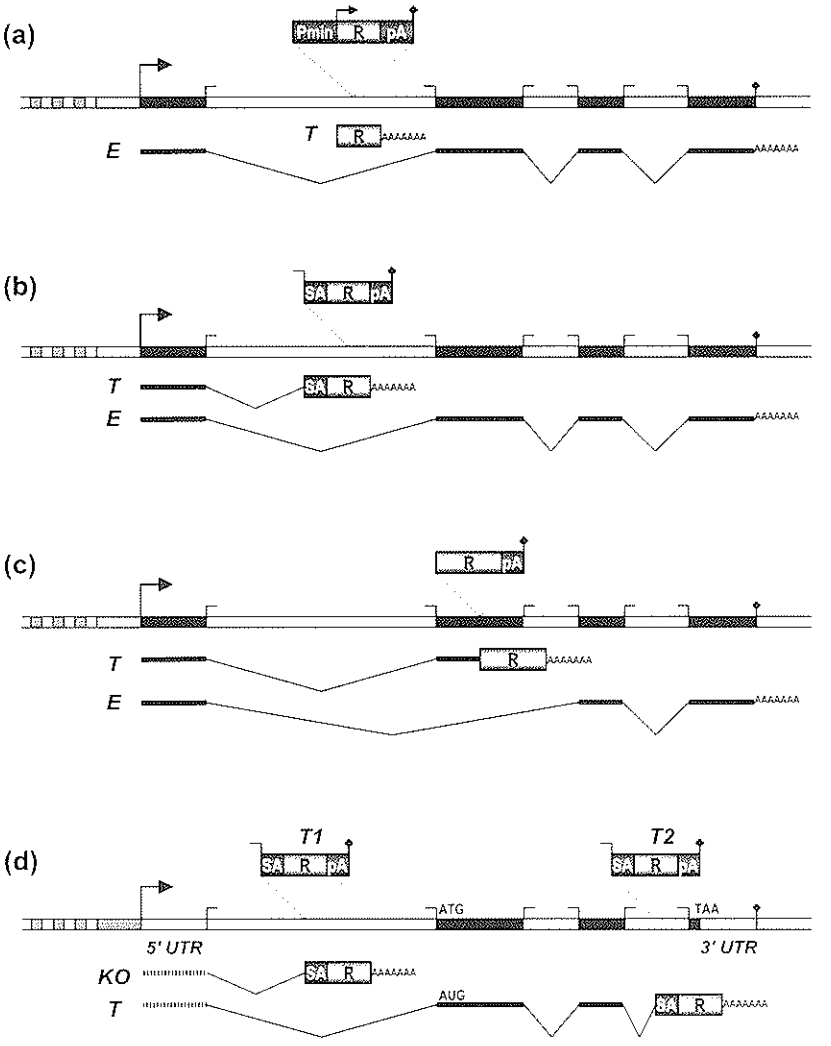

Figure 4.2. Possible causes of trapped gene partial loss of function. This figure illustrates how different types of gene trap vectors may cause incomplete gene loss of function. This generally results in the concomitant presence of two transcripts originating from the same gene, one reflecting proper trapping (*T*), and the other containing most (if not all) of the endogenous sequence (*E*). (a) Intronic integration of an enhancer trap with a weak poly(A) site may result in splicing out of the trap from the endogenous transcript. (b) Intronic integration of a promoter trap, associated with weak activity of its SA and poly(A) domains, may result in the concomitant generation of two transcripts, one containing the trap as its last exon (*T*), and the other splicing out the trap and continuing with the downstream exons (*E*). (c) Trap integration within an exon that physiologically undergoes alternative splicing: if the trap poly(A) site is not adequately efficient, the alternatively spliced mRNA retains its full functionality. (d) While trap insertion in the 5' portion of a gene (T1) generally results in efficient gene targeting (*KO*), integration in the 3' portion (T2) may result in the translation of a totally or partially functional protein from the endogenous ORF (*WT*).

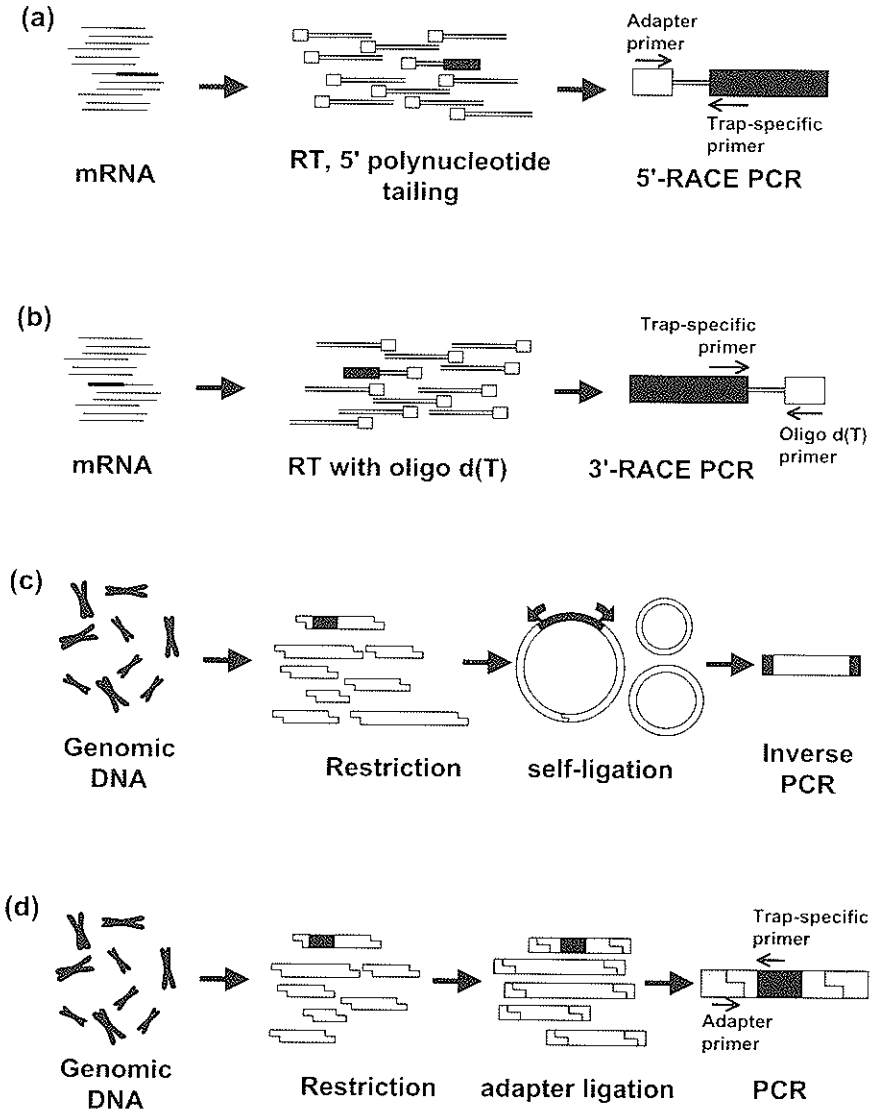

Figure 4.3. Identification of trapped genes. (a) 5'-RACE. (b) 3'-RACE. (c) Inverse PCR. (d) Adapter PCR. (See text for details.)

Rapid amplification of cDNA ends (RACE)

As a consequence of effective promoter, CDS, or poly(A) trapping, trapped cells express a fusion transcript consisting of a portion derived from the endogenous gene, and the remaining part derived from the trap. In promoter and CDS traps, the trap-derived portion is located at the 3'-end of the transcript. Specific RT-PCR amplification of these fusion transcripts may therefore be achieved using one trap-specific antisense primer and one primer annealing to an adaptor sequence previously added to the 5'-

end of all the cDNAs. This procedure is therefore called 5'-RACE (*Figure 4.3a*). The common 5' sequence may be added either by a terminal deoxynucleotidyltransferase (tdt) reaction, which adds a 3'-poly(mononucleotide) tail, e.g. poly(A), to the antisense cDNA (Frohman *et al.*, 1988), or by direct adaptor linking during reverse transcription (Matz *et al.*, 1999). In the case of poly(A)-traps, the trap-derived portion is at the 5'-end of the transcript. Amplification of these fusion transcripts may be achieved using one trap-specific sense primer and one oligo(dT) primer annealing to the 3' poly(A) tail. This procedure is therefore called 3'-RACE (*Figure 4.3b*). As in both 5'- and 3'-RACE, only one of the primers is specific for the fusion transcript; these procedures frequently require multiple PCR steps with nested primers. The RACE products can then be cloned and sequenced, or also directly sequenced, when the amplification procedure is particularly efficient, as in the case of 3'-RACE of poly(A) traps. In fact, in this case, the poly(A) tail exploited for PCR is already present in all mature transcripts and does not need to be added during or after cDNA synthesis (Zambrowicz *et al.*, 1998). Indeed, a method has also been developed for direct, solid phase sequencing of 5'-RACE products, which yielded satisfactory sequence information from every successful 5'-RACE reaction (Townley *et al.*, 1997).

Gene identification by 5'- or 3'-RACE has been widely used in gene trapping, for two main reasons: 1) before completion of the genome sequencing projects, mRNA-derived sequences are more likely to be represented in databases, which increases the probability of successful gene identification with respect to sequences derived from gene intronic regions; and 2) a successful RACE reaction reflects generation of a fusion transcript, which is a good indicator of proper trapping and more probable effective targeting of the trapped gene (Townley *et al.*, 1997). The main drawback of this approach is that precise mapping of trap insertion sites in intronic regions is not possible.

Inverse PCR

This technique allows amplification of small genomic regions immediately 5' or 3' to the integrated trap (*Figure 4.3c*). It first involves digestion of genomic DNA with a restriction enzyme of known, rare cutting frequency. The presence of one or more corresponding sites in the trap does not impair the procedure, but it has to be known in advance for proper primer design. Subsequently, the restriction fragments undergo a ligation reaction calibrated to favour DNA circularization by self-ligation. Among all DNA circles, the one containing the trap-derived sequence is PCR-amplified by using two primers designed on the two extremities of the calculated trap insert, oppositely directed towards the flanking DNA. The PCR products are then cloned and sequenced (van Lohuizen *et al.*, 1991). A similar technique, based on DNA restriction and circularization, has been developed by Hicks and colleagues (1997). In this case, the trapping vector contains a plasmid rescue module, which allows direct transformation of bacteria with the circularized DNA to recover the trap and flanking genomic sequences.

Adapter PCR

Like inverse PCR, this technique allows the amplification of trap-flanking genomic regions (*Figure 4.3d*). Genomic DNA is cut by restriction enzymes. Fragments

obtained are ligated to adaptors, and then amplified using one primer specific for the linker and one specific for the trap sequence. The products obtained are then cloned and sequenced (Schroder *et al.*, 2002).

In general, genomic DNA-based techniques are more labour intensive than RACE, but provide precise information on the trap integration site when it also occurs in an intron, and independently from the generation of fusion transcripts.

TRAP SELECTION

Gene trapping involves parallel handling of multiple cell clones, each originating from a different integration event. As a consequence, the throughput of a gene trap screening is limited by the number of independent clones possibly handled. To maximize the screening multiplicity and reduce cell handling at the same time, a series of selection procedures have been devised. We will subdivide trap selection procedures into two main classes, using the cell cloning step as the differentiating event. In fact, selection before cloning, which we define *a priori* selection, allows us to start from a high multiplicity of integrants, to reduce the cells to those carrying the desired type of integration, and thereafter facilitates the handling of a limited number of clones. Conversely, *a posteriori* selection occurs after cloning, and thus requires the initial handling of a large number of clones, most of which are subsequently discarded. Examples of *a priori* selection include pharmacological selection (Friedrich and Soriano, 1991) and cell sorting (Reddy *et al.*, 1992); examples of *a posteriori* selection include expression analysis (Wurst *et al.*, 1995; Forrester *et al.*, 1996), trapped gene identification (Townley *et al.*, 1997), studies of subcellular localization (Skarnes *et al.*, 1995; Tate *et al.*, 1998), all performed on cell clones, or also direct phenotypic analysis in chimeric animals (Gossler *et al.*, 1989). In many cases, *a priori* and *a posteriori* selection procedures may be combined, for example, *a priori* drug-based selection of traps in expressed genes may be combined with *a posteriori* sequence-based selection of traps efficiently targeting novel genes, to achieve high efficiency in novel gene characterization (Hicks *et al.*, 1997).

Gene trap vectors and their application

ENHANCER TRAPS

Enhancer trapping was originally developed to probe the mouse genome for active chromosomal regions involved in developmental processes (Allen *et al.*, 1988). The trapping vector was simply composed of β-gal driven by the minimal HSV-TK promoter, which was directly used to generate transgenic mice (*Figure 4.4a*). A similar construct, with β-gal expression driven by a mouse weak promoter derived from the hsp68 gene, was used to study chromosomal domains involved in the development of neural tube in mice (Kothary *et al.*, 1988). These studies demonstrated that activity of weak promoters is largely dependent on their chromosomal position, and that enhancer traps may serve as useful tools for identifying tissue-specific enhancers and their associated genes. A second, refined enhancer trap vector was subsequently constructed to take advantage of mouse ES cells for a generation of transgenic mice (Gossler *et al.*, 1989). The trap consisted of β-gal driven by the hsp68

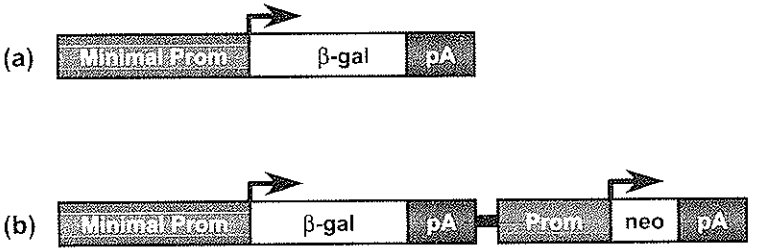

Figure 4.4. Enhancer trap vectors. (a) Basic enhancer trap construct originally developed for studies in transgenic animals (Allen *et al.*, 1988; Kothary *et al.*, 1988). The vector is composed of β-gal driven by a minimal promoter (Minimal Prom) and followed by a polyadenylation site (pA). (b) Selectable enhancer trap plasmid vector (Gossler *et al.*, 1989). The enhancer trap module is followed by a selection cassette composed of a constitutive promoter (Prom), followed by neo and a poly(A) site.

weak promoter, followed by a constitutively expressed neo cassette (*Figure 4.4b*). The construct was introduced in mouse ES cells, which were subsequently G418 selected and introduced in mouse blastocysts. This facilitated the direct study of the expression of the trapped genes in chimeric embryos, without the need for generating transgenic mice. This further confirmed the hypothesis that transcriptional activity of minimal promoters is strictly dependent on endogenous, *cis*-acting elements, the enhancers. As discussed, enhancer traps may, in some cases, disrupt the genes in which they integrate, which permits concomitant monitoring of promoter activity and the investigation of the consequences of gene loss. A retroviral version of the enhancer trap has also been constructed by deleting the LTR enhancer element and using the *neo* gene as selector and reporter (Sablitzky *et al.*, 1993). More recently, enhancer trap vectors have been used in *Drosophila*. Also in this case, enhancer trap lines have been generated to identify genes with cell type- or tissue-specific expression during development (Gustafson and Boulianne, 1996). In this case, the reporter gene is itself a transcription factor, the yeast GAL4, which activates the upstream activating sequence (UAS) element. In this way, detection of GAL4 expression patterns has been achieved by crossing individual GAL4 lines with flies carrying the reporter gene lacZ under the transcriptional control of the UAS.

Altogether, enhancer traps offer the advantage of high sensitivity towards gene regulatory regions, without the need for stringently determined integration sites and orientation. As a consequence however, a possible problem with this approach may be the difficulty in identifying the gene responsible for the enhancer activity, due to the possible distance between the vector insertion site and the gene itself.

PROMOTER TRAPS

Promoter traps have been originally developed to monitor the activity of endogenous promoters and identify growth arrest-induced genes in NIH-3T3 cells (Brenner *et al.*, 1989). The trapping vector designed for this task, Mo-MuLV*lac*, consisted of a self-inactivating retrovirus containing, in reverse orientation with respect to the LTRs, a splice acceptor site followed by LacZ and an SV40-derived polyadenylation site

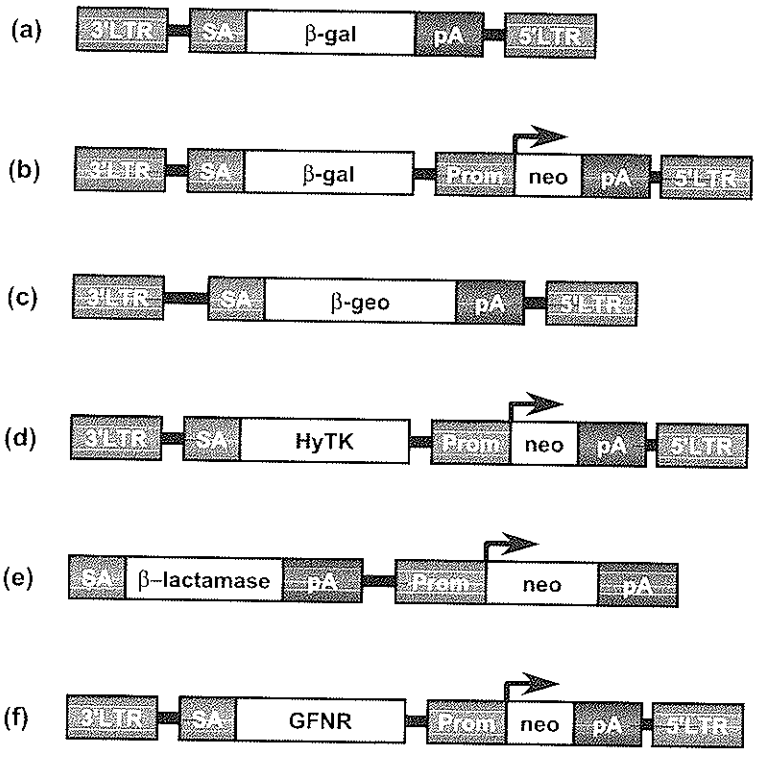

Figure 4.5. Promoter trap vectors. (a) Retroviral promoter trap composed of a splice acceptor site (SA), followed by β-gal and poly(A), with the viral LTRs disposed in reverse orientation (Brenner *et al.*, 1989). (b) Selectable retroviral promoter trap (Friedrich and Soriano, 1991) in which the SA-β-gal promoter trapping module is followed by a constitutive neo cassette. (c) Retroviral promoter trap incorporating the β-geo selectable reporter (Friedrich and Soriano, 1991). (d) Retroviral promoter trap incorporating the HyTK double selection marker (Gogos *et al.*, 1996), followed by a constitutive neo cassette. (e) Plasmid promoter trap incorporating β-lactamase (β-lac) as a reporter in the promoter trap module, followed by a constitutive neo cassette (Whitney *et al.*, 1998). (f) Retroviral promoter trap incorporating the GFNR selectable reporter, followed by a constitutive neo cassette (Medico *et al.*, 2001b). Enhancer deletion in the 3' LTRs of all retroviral promoter trap vectors is not illustrated.

(*Figure 4.5a*). Reverse orientation of the trap versus the viral transcript has been chosen to avoid interference of the trap SA and poly(A) signals with the correct processing of retroviral RNA, and to concomitantly eliminate any possible influence of the remaining LTR portions on the transcription of the trap reporter gene. Traps in active genes were selected by fluorescence-activated cell sorting (FACS), and subsequently screened for increased LacZ activity upon serum starvation.

A similar retroviral vector was built to take advantage of the high mutagenic activity of SA-based promoter traps for identifying genes involved in mouse development (Friedrich and Soriano, 1991). This trap was called reverse oriented splice acceptor β-gal (ROSA-β-gal, *Figure 4.5b*), and contained a PGK promoter-driven neo selection cassette downstream of the SA-LacZ module (Friedrich and Soriano, 1991). This vector allowed low infection multiplicity, followed by G418 selection, to

achieve a vast majority of single-copy integrants. Mouse ES cells were trapped, G418 selected, cloned, and stained with X-gal to identify traps in active genes. X-gal-positive clones were subsequently used to generate mice carrying homozygous trap insertion, and therefore likely to be null mutants for the trapped gene. This facilitated the identification of novel genes with essential developmental roles, and the concomitant *in vivo* monitoring of the activity of their promoters. To improve trap selection efficiency, Friedrich and Soriano have also built a selectable reporter deriving from the fusion of β-gal and *neo*, called β-geo. In ROSA-β-geo, the selectable reporter is preceded by an SA site and followed by a poly(A) site (*Figure 4.5c*). G418 selection allows direct isolation of cells carrying traps in active genes. As G418 selection is more sensitive than X-gal staining, ROSA-β-geo trapping may allow selection of genes expressed at very low levels (Friedrich and Soriano, 1991).

The reliability of promoter traps as reporters of gene transcriptional activity has been further exploited to select and monitor integrations downstream from regulated promoters. Forrester and colleagues (1996) have used an SA-β-gal plasmid trap, including a *neo* cassette to select integrants, to screen for ES cell traps in retinoic acid (RA)-induced genes. The screening involved replica plating of all G418-resistant clones in the presence of RA, followed by X-gal staining. Selected positive clones had to be subsequently screened for reduction of β-gal activity in the absence of RA. Generation of homozygous mice carrying traps in RA-induced genes has allowed identification of RA targets essential for embryonic development.

Selection of traps in expressed and/or regulated genes by *a posteriori* screening of replica plates is quite labour intensive, and this limits the number of distinct trapping events that can be reasonably explored. In this regard, *a priori* selection of 'interesting' traps by pharmacological treatment or cell sorting could allow more extensive surveys. With this aim in mind, several traps that take advantage of selection markers, or *a priori* selectable reporters, have been constructed. To identify genes regulated by MyoD in a mouse fibroblast cell line carrying inducible MyoD expression, Gogos and colleagues (1997) have derived a promoter trap from ROSA-β-gal in which β-gal was substituted by a fusion between hygromycin phosphotransferase and thymidine kinase, named ROSA-HyTK (*Figure 4.5d*). When expressed, HyTK renders cells resistant to hygromycin, but also sensitive to gancyclovir. Therefore, ROSA-HyTK has allowed sequential pharmacological treatments, starting with G418+gancyclovir in the MyoD off-state, followed by hygromycin in the MyoD on-state, to directly select traps in MyoD-induced genes. However, this trapping vector did not provide any reporter downstream of the endogenous promoter, so gene regulation by MyoD had to be validated at the RNA level (Gogos *et al.*, 1997). To study genes transcriptionally regulated during Jurkat T cell activation by phytohaemagglutinin (PHA), a plasmid gene trap vector (GAS-1; *Figure 4.5e*) was built to include β-lactamase (β-lac) as a selectable reporter. β-lac is preceded by an SA and followed by a poly(A) site and a neo cassette (Whitney *et al.*, 1998). To detect β-lac activity, cells have to be loaded with a fluorogenic substrate containing a β-lactamic ring, CCF2/AM. Cleavage of the CCF2/AM β-lactamic ring leads to a marked shift in its fluorescence emission spectrum, which can be used to precisely quantify β-lac expression and sort positive or negative cells (Zlokarnik *et al.*, 1998). The sensitivity of the system is quite high, and this allows identification of traps in genes expressed at very low levels. Traps in PHA-induced genes were selected by three subsequent

rounds of cell sorting: 1) β-lac-positive cells in the presence of PHA; 2) β-lac-negative cells in the absence of PHA; and 3) β-lac-positive cells in the presence of PHA. The procedure was reversed to select traps in PHA-suppressed genes: 1) β-lac-positive cells in the absence of PHA; 2) β-lac-negative cells in the presence of PHA; and 3) β-lac-positive cells in the absence of PHA. After sorting, individual clones were generated and tested for PHA-induced or suppressed β-lac expression. One possible problem with this very efficient selection method is that clones are derived at the end of the entire procedure, which may result in multiple redundant clones derived from the same trapping event.

A combination between cell sorting and pharmacological counter-selection can be efficiently applied for a promoter trap vector we developed to increase the throughput of a systematic screening (Medico *et al.*, 2001b). The trap is a derivative of ROSA-β-gal, where the β-gal coding sequence is substituted by a fusion between enhanced green fluorescent protein (eGFP; *Zhang et al.*, 1996) and *Escherichia coli* nitroreductase (NTR; Bridgewater *et al.*, 1995), named green-fluorescent nitroreductase (GFNR). GFNR-expressing cells are fluorescent, which allows direct quantification of GFNR abundance and positive selection by FACS; at the same time, the NTR component of GFNR renders cells sensitive to pro-drugs like metronidazole (MN; Bridgewater *et al.*, 1997), which kill rapidly and efficiently only GFNR-positive cells. The GFNR-based retroviral trap vector, named ROSA-GFNR (*Figure 4.5f*), allows sequential positive selection by FACS, and negative selection by MN, to enrich for traps in transcriptionally regulated genes. We have used this vector to isolate traps in genes induced or suppressed by hepatocyte growth factor (HGF) in epithelial cells (Medico *et al.*, 2001b). The selection procedure for HGF-induced traps consisted of three steps: 1) infection of cells and concomitant pharmacological selection with G418 and MN, which allowed selection of all integrants and exclusion of traps in constitutively expressed genes; 2) removal of MN, stimulation with HGF, and selection of fluorescent cells by FACS, followed by immediate cloning, which allowed the isolation of clones carrying traps in genes preferentially expressed in the presence of HGF; and 3) further counter-selection by MN treatment of growing clones in the absence of HGF, which allowed a more efficient elimination of traps in constitutive genes that survived the first counter-selection.

The reversed selection procedure for HGF-suppressed genes also consisted of three steps: 1) infection of cells and treatment with G418, which allowed selection of all integrants; 2) FACS sorting and cloning of fluorescent cells, to isolate cells carrying traps in genes expressed in the absence of HGF; and 3) HGF stimulation, followed by MN counter-selection, to eliminate traps in genes not suppressed by HGF treatment. To efficiently eliminate traps in constitutive genes, step 3 had to be performed at least twice. ROSA-GFNR trapping combines two efficient methods of *a priori* selection, drug treatment and cell sorting, minimizes redundancy of trapped clones, and brings a direct reporter (GFP) downstream from trapped promoters. A possible issue with this system, which is shared with β-lac-based trapping, is that it relies on a state-of-the-art cell sorting facility, which may not be available to many labs. Other strategies to trap transcriptionally regulated genes have been developed, relying on specifically designed selectable reporter systems like a triple fusion between β-gal, BLA-S deaminase, and HSV-TK (Akiyama *et al.*, 2000), and promoterless Cre-activated excision of a TK–neo fusion (Russ *et al.*, 1996).

CDS TRAPS

The CDS trap design allows the selection of traps in protein coding portions of functional genes. As a consequence, cells expressing the CDS trap-derived reporter are likely to carry a proper trapping event, with disruption of gene function. Moreover, CDS traps allow exploration of subcellular localization signals eventually provided by the fused CDS, which allows extensive protein localization studies. As an example, systematic search for CDS trapping events resulting in nuclear localization of fusion proteins allowed identification of novel mouse genes encoding chromosomal and other nuclear proteins (Tate *et al.*, 1998).

In the absence of an SA site, the efficiency of CDS trapping is extremely low, requiring integration within a coding exon, in the right orientation and reading frame. Upon addition of an SA site upstream of the ATG-less reporter, CDS traps have a 12-fold higher frequency of reporter-positive transformants (Gossler *et al.*, 1989). *Figure 4.6a* illustrates the components of a plasmidic CDS trap vector, pGT4.5, developed by Gossler and colleagues in the first example of gene trapping of mouse ES cells (Gossler *et al.*, 1989), and extensively used to concomitantly disrupt genes and localize their protein products (Skarnes *et al.*, 1992). An incomplete loss of function with CDS trapping may occur when the N-terminal portion of the fusion protein is long enough to retain part, or all, of its function (*Figure 4.2d*).

SS TRAPS

SS traps can be considered as a subclass of CDS traps specifically designed for exploration of genes involved in cell–cell and cell–matrix signalling. *Figure 4.6b* illustrates the first example of an SS trap, the plasmidic pGT1.8TM vector (Skarnes *et al.*, 1995). The SS trapping module consists of an SA, followed by an ORF encoding a fusion between an ATG-less, CD4-derived TM domain and β-geo. In the absence of an SS provided by the trapped gene, β-geo is exposed at the outer cell surface, which results in loss of β-gal activity. Conversely, traps in genes encoding membrane or secreted proteins acquire an SS upstream of the TM domain, and display a typical membrane/secretory X-gal pattern derived from active β-gal located in the cytoplasmic face of the membrane (Skarnes *et al.*, 1995). Extensive gene trap screenings have been carried out using pGT1.8TM and its derivatives, and are still ongoing (Skarnes, 2000). A very interesting derivative of pGT1.8TM is a vector specifically designed to explore brain wiring patterns, the 'PLAP secretory trap' (*Figure 4.6c*). In this vector, β-geo is followed by an IRES and by PLAP, a GPI-linked cell-surface protein that, when expressed transgenically in neurons, labels axons completely (Leighton *et al.*, 2001). This elegant approach allows X-gal staining of the cellular bodies of neurons in which the trapped promoter is active, and concomitant PLAP staining to map their axonic projections. Finally, a retroviral SS trap has been developed by Gebauer and colleagues to trap secreted and membrane proteins that are transcriptionally regulated by IGF-1 signalling (Gebauer *et al.*, 2001). The trap is named U3Ceo (*Figure 4.6d*), and consists of a fusion between the extracellular and transmembrane domain of the CD2 cell surface antigen and neo. The fusion, named Ceo, lacks both the initiating ATGs and the SS, and is inserted in the U3 element of a retroviral, enhancerless 3′ LTR. Upon infection and reverse transfection, the 3′-LTR

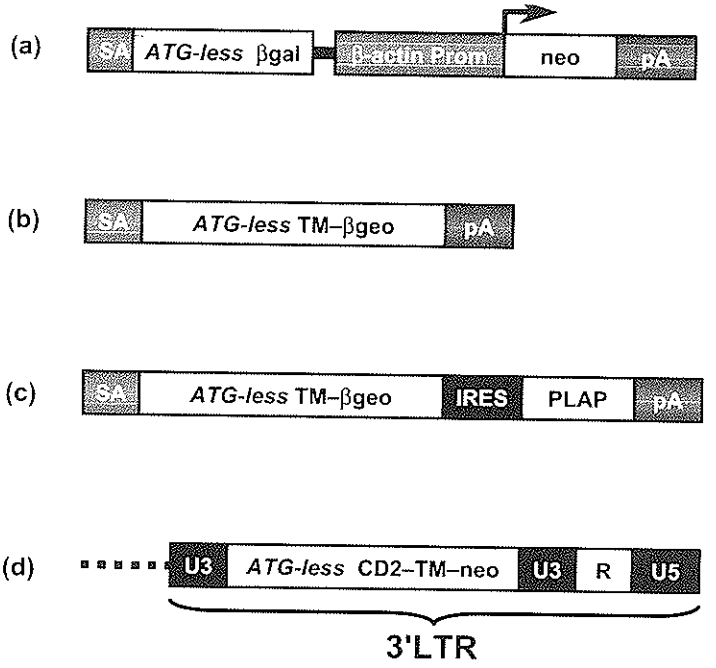

Figure 4.6. CDS and SS trap vectors. (a) Plasmidic CDS trap vector, composed of an ATG-less β-gal, preceded by an SA site, and followed by a constitutive neo cassette (Gossler *et al.*, 1989). (b) Plasmidic SS trap vector, composed of an ATG-less fusion between a type II transmembrane domain (TM) and β-geo, preceded by an SA, and followed by a polyadenylation (pA) site (Skarnes *et al.*, 1995). (c) PLAP secretory trap vector (Leighton *et al.*, 2001), composed of an SS-trapping module identical to that of panel B, followed by an IRES, and the PLAP coding sequence (see text). (d) 3' LTR of a retroviral SS trap vector, in which the U3 region of the 3' LTR carries an ATG-less fusion between the CD2 cell surface antigen extracellular and TM domains, and neo (*ATG-less* CD2–TM–neo; Gebauer *et al.*, 2001). Enhancer deletion in the LTR is not illustrated.

is replicated at the 5'-end of the provirus which, upon integration, brings Ceo just 30 nucleotides from the flanking endogenous DNA. Only upon entrapment of an SS-providing gene does the CD2 antigen get exposed at the outer surface of the cell membrane, and can therefore be used for sequential positive and negative selection by panning. This ultimately enables *a priori* selection of traps in transcriptionally regulated genes encoding secreted or membrane proteins (Gebauer *et al.*, 2001).

POLY(A) TRAPS

The first gene trap vector with a poly(A) trapping module simply lacked the poly(A) signal downstream from the *neo* gene to increase gene trapping efficiency in ES cells (Niwa *et al.*, 1993). Subsequent poly(A) trapping vectors would include an SD site (Yoshida *et al.*, 1995), and would be specifically developed to disrupt genes in mouse ES cells independently of their expression state (*Figure 4.7a*; Salminen *et al.*, 1998; Zambrowicz *et al.*, 1998). However, endogenous gene disruption is not assured by the

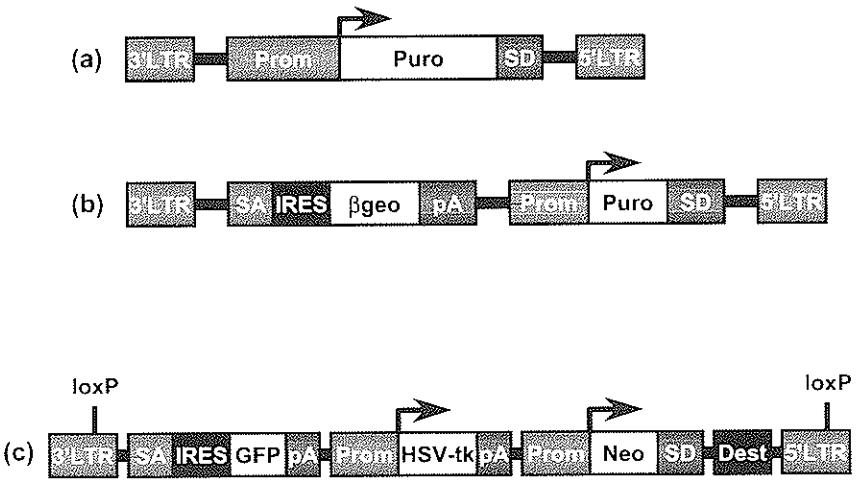

Figure 4.7. Poly(A) trap vectors. (a) Retroviral poly(A) trap vector, in which the trapping module is composed of a constitutive promoter (Prom), followed by puro and a splice donor site (SD) (Zambrowicz *et al.*, 1998). (b) Retroviral combined promoter- and poly(A)-trapping vector (Zambrowicz *et al.*, 1998). The promoter trap module is based on the β-geo selectable reporter, preceded by SA-IRES and followed by pA. The poly(A) trap module is identical to the one described in panel A. (c) 'Removable exon trap' (Ishida and Leder, 1999), composed as follows in 5'–3' order: 1) a promoter trap module, SA-IRES-GFP-pA; 2) a constitutive HSV-tymidine kinase (tk) cassette; and 3) the poly(A) trap, made of constitutively expressed neo, followed by an SD and an RNA destabilizing sequence (Dest). The three modules are put in reverse orientation, in a self-inactivating retroviral vector containing loxP sites in the LTRs. Enhancer deletion in the 3' LTRs of all retroviral vectors is not illustrated.

poly(A) trap alone: when integration takes place in an intron, the endogenous transcript may splice out the entire trap, which results in a fully wild-type mRNA (Zambrowicz *et al.*, 1998). To overcome this problem, a promoter trap module has been added 5' of poly(A) traps (Salminen *et al.*, 1998; Zambrowicz *et al.*, 1998; Ishida and Leder, 1999). One retroviral poly(A) trap used for systematic gene tagging in ES cells by the Lexicon Genetics Company (www.lexgen.com) is VICTR20 (*Figure 4.7b*). This trap is constituted of two units: a mutagenic component (promoter trap) at the 5', and a poly(A) trap at the 3'. The mutagenic component consists of an SA-IRES-β-geo-pA cassette. The poly(A) trap consists of a puro-SD cassette driven by the PGK promoter. Cells infected with this trap are selected with puromycin, so that only clones bearing poly(A) trapping events can survive. It is then possible to monitor the activity of the trapped genes' promoters by X-gal staining.

When a gene trap generates a mutant mouse phenotype, a formal proof that the observed phenotype specifically depends on the trapping event may come from phenotypic reversion upon trap removal. To render trap removal possible and efficient, Ishida and Leder (1999) have generated a 'removable exon trap' (RET; *Figure 4.7c*). RET is composed of three modules, in 5'–3' order: 1) a promoter trap, SA-IRES-GFP-pA; 2) a constitutive HSV-tymidine kinase (tk) cassette, for virus titration in tk^- cells, and for RET counter-selection by gancyclovir; and 3) a poly(A) trap, made of RNApolII-driven *neo* followed by an SD and an RNA destabilizing

sequence derived from the 3'-UTR of the GM-CSF mRNA. The three modules are put in reverse orientation in a self-inactivating retroviral vector containing loxP sites in the LTRs. Therefore, after phenotypic characterization of mice carrying homozygously trapped genes, the integrated traps can be excised from the host genome upon transient or stable expression of the Cre recombinase (Gu *et al.*, 1993). Selection of cells that have lost integrated RET can be easily achieved by gancyclovir treatment. Upon Cre-mediated trap excision, the only remnant is an enhancer-negative LTR that has no detectable effects on the transcription and splicing of the host gene (Ishida and Leder, 1999).

Concluding remarks

Although originally developed before the 'genomic era', gene trap vectors have evolved to become precious tools for current functional genomics projects in different model organisms. Gene trapping efficiency and flexibility, together with the possibility to perform systematic, multi-parametric analysis, are distinctive features that enable functional screening projects. Indeed, large-scale projects of gene trapping and subsequent functional analysis are currently ongoing in major research centres, as illustrated in the following website list that concludes this hopefully comprehensive overview.

LINKS TO GENE TRAPPING WEBSITES

University of Manitoba Institute of Cell Biology, Winnipeg, Canada: http://www.escells.ca/
Extensive Tagged-Sequence Mutagenesis to disrupt genes expressed in mouse ES cells and to characterize each mutation by direct DNA sequencing. The initial goal is to develop an Embryonic Stem Cell Library of 20–40 000 defined gene mutations.

The German Gene Trap Consortium (GGTC): http://tikus.gsf.de/
The GGTC aim is to generate a reference library of gene trap sequence tags (GTST) from insertional mutations generated in mouse embryonic stem (ES) cells. Gene traps are generated in ES cells using various types of vectors, which are delivered either by electroporation or retroviral infection.

The BayGenomics Gene Trap Project: http://baygenomics.ucsf.edu/
The BayGenomics project is to generate gene traps in mouse ES cells to inactivate 2500 random genes per year.

The Centre for Modelling Human Disease (CMHD) Gene Trap Project, Toronto, Canada: http://www.cmhd.ca/
The project goal is to simultaneously generate and annotate by expression analysis and sequence tags an Embryonic Stem (ES) library of 5000 gene trap clones per year. The gene trap library will be screened for insertions in genes expressed by haematopoietic, endothelial, cardiomyocyte, mesenchymal, and neural lineages, as well as genes induced by physiological stimuli including ionizing irradiation, retinoic acid, hypoxia, and members of the TGF-β family.

Lexicon Genetics 'OmniBank': http://www.lexgen.com/omnibank/omnibank.htm
This company has developed a high throughput gene trapping technology to create OmniBank, a library of more than 200 000 genetically modified mouse embryonic stem cell clones. Each OmniBank mouse clone contains a gene trap event in a single gene.

Nancy Hopkins' lab: http://web.mit.edu/biology/www/facultyareas/facresearch/hopkins.shtml
This lab focuses on identification of the genetic basis of early developmental processes and simple behaviours in the zebrafish Danio rerio *using large-scale insertional mutagenesis screening. Mutants with defects in about half the 2400 genes have been visualized to date.*

ADDITIONAL LINKS

Harwell mouse ENU mutagenesis programme: http://www.mut.har.mrc.ac.uk/
The purpose of this mutagenesis programme is to generate and make available approximately 100 F1 progeny of mutagenized animals per week for genome-wide dominant phenotypic screens.

Oak Ridge National Laboratory's Mouse Genetics and Genomics Program: http://bio.lsd.ornl.gov/mgd/
The programme operates at the intersection between genetics and genomics, building upon successful past and current strengths in genomics, genetic analysis, mutagenesis, phenotype screening, and germ-cell biology to exploit the mouse fully as a model system for post-genome biology.

The Jackson Laboratory: http://jaxmice.jax.org/index.shtml
A well known information resource describing 2500+ strains of genetically defined mice for biomedical research.

References

AKIYAMA, N., MATSUO, Y., SAI, H., NODA, M. AND KIZAKA-KONDOH, S. (2000). Identification of a series of transforming growth factor beta-responsive genes by retrovirus-mediated gene trap screening. *Molecular and Cellular Biology* **20**, 3266–3273.

ALLEN, N.D., CRAN, D.G., BARTON, S.C., HETTLE, S., REIK, W. AND SURANI, M.A. (1988). Transgenes as probes for active chromosomal domains in mouse development. *Nature* **333**, 852–855.

BRENNER, D.G., LIN-CHAO, S. AND COHEN, S.N. (1989). Analysis of mammalian cell genetic regulation *in situ* by using retrovirus-derived 'portable exons' carrying the *Escherichia coli* lacZ gene. *Proceedings of the National Academy of Sciences of the United States of America* **86**, 5517–5521.

BRIDGEWATER, J.A., SPRINGER, C.J., KNOX, R.J., MINTON, N.P., MICHAEL, N.P. AND COLLINS, M.K. (1995). Expression of the bacterial nitroreductase enzyme in mammalian cells renders them selectively sensitive to killing by the prodrug CB1954. *European Journal of Cancer* **31A**, 2362–2370.

BRIDGEWATER, J.A., KNOX, R.J., PITTS, J.D., COLLINS, M.K. AND SPRINGER, C.J. (1997). The bystander effect of the nitroreductase/CB1954 enzyme/prodrug system is due to a cell-permeable metabolite. *Human Gene Therapy* **8**, 709–717.

BRONCHAIN, O.J., HARTLEY, K.O. AND AMAYA, E. (1999). A gene trap approach in *Xenopus*. *Current Biology* **9**, 1195–1198.

BRUMMELKAMP, T.R., BERNARDS, R. AND AGAMI, R. (2002). A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296**, 550–553.

CHEN, W., BURGESS, S., GOLLING, G., AMSTERDAM, A. AND HOPKINS, N. (2002). High-throughput selection of retrovirus producer cell lines leads to markedly improved efficiency of germ line-transmissible insertions in zebra fish. *Journal of Virology* **76**, 2192–2198.

CHOWDHURY, K., BONALDO, P., TORRES, M., STOYKOVA, A. AND GRUSS, P. (1997). Evidence for the stochastic integration of gene trap vectors into the mouse germ line. *Nucleic Acids Research* **25**, 1531–1536.

EISEN, M.B., SPELLMAN, P.T., BROWN, P.O. AND BOTSTEIN, D. (1998). Cluster analysis and display of genome-wide expression patterns. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 14863–14868.

FORRESTER, L.M., NAGY, A., SAM, M. *ET AL.* (1996). An induction gene trap screen in embryonic stem cells: identification of genes that respond to retinoic acid *in vitro*. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 1677–1682.

FRIEDRICH, G. AND SORIANO, P. (1991). Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes and Development* **5**, 1513–1523.

FROHMAN, M.A., DUSH, M.K. AND MARTIN, G.R. (1988). Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proceedings of the National Academy of Sciences of the United States of America* **85**, 8998–9002.

GEBAUER, M., VON MELCHNER, H. AND BECKERS, T. (2001). Genome-wide trapping of genes that encode secreted and transmembrane proteins repressed by oncogenic signalling. *Genome Research* **11**, 1871–1877.

GOGOS, J.A., THOMPSON, R., LOWRY, W., SLOANE, B.F., WEINTRAUB, H. AND HORWITZ, M. (1996). Gene trapping in differentiating cell lines: regulation of the lysosomal protease cathepsin B in skeletal myoblast growth and fusion. *Journal of Cell Biology* **134**, 837–847.

GOGOS, J.A., LOWRY, W. AND KARAYIORGOU, M. (1997). Selection for retroviral insertions into regulated genes. *Journal of Virology* **71**, 1644–1650.

GONZALEZ, C. AND BEJARANO, L.A. (2000). Protein traps: using intracellular localization for cloning. *Trends in Cell Biology* **10**, 162–165.

GOSSLER, A., JOYNER, A.L., ROSSANT, J. AND SKARNES, W.C. (1989). Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Science* **244**, 463–465.

GREGORY, S.G., SEKHON, M., SCHEIN, J. *ET AL.* (2002). A physical map of the mouse genome. *Nature* **418**, 743–750.

GU, H., ZOU, Y.R. AND RAJEWSKY, K. (1993). Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting. *Cell* **73**, 1155–1164.

GUSTAFSON, K. AND BOULIANNE, G.L. (1996). Distinct expression patterns detected within individual tissues by the GAL4 enhancer trap technique 45. *Genome* **39**, 174–182.

HARRINGTON, J.J., SHERF, B., RUNDLETT, S. *ET AL.* (2001). Creation of genome-wide protein expression libraries using random activation of gene expression. *Nature Biotechnology* **19**, 440–445.

HICKS, G.G., SHI, E.G., LI, X.M., LI, C.H., PAWLAK, M. AND RULEY, H.E. (1997). Functional genomics in mice by tagged sequence mutagenesis. *Nature Genetics* **16**, 338–344.

ISHIDA, Y. AND LEDER, P. (1999). RET: a poly A-trap retrovirus vector for reversible disruption and expression monitoring of genes in living cells. *Nucleic Acids Research* **27**, e35.

JUSTICE, M.J. (2000). Capitalizing on large-scale mouse mutagenesis screens. *Nature Reviews in Genetics* **1**, 109–115.

KOTHARY, R., CLAPOFF, S., BROWN, A., CAMPBELL, R., PETERSON, A. AND ROSSANT, J. (1988). A transgene containing lacZ inserted into the dystonia locus is expressed in neural tube. *Nature* **335**, 435–437.

KUMAR, A., HARRISON, P.M., CHEUNG, K.H. *ET AL.* (2002). An integrated approach for finding overlooked genes in yeast. *Nature Biotechnology* **20**, 58–63.

LAI, Z., HAN, I., PARK, M. AND BRADY, R.O. (2002). Design of an HIV-1 lentiviral-based gene-trap vector to detect developmentally regulated genes in mammalian cells. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 3651–3656.

LANDER, E.S., LINTON, L.M., BIRREN, B. *ET AL.* (2001). Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921.

LEIGHTON, P.A., MITCHELL, K.J., GOODRICH, L.V. *ET AL.* (2001). Defining brain wiring patterns and mechanisms through gene trapping in mice. *Nature* **410**, 174–179.

LEWIS, D.L., HAGSTROM, J.E., LOOMIS, A.G., WOLFF, J.A. AND HERWEIJER, H. (2002). Efficient delivery of siRNA for inhibition of gene expression in postnatal mice. *Nature Genetics* **32**, 107–108.

LINDSEY, K., WEI, W., CLARKE, M.C., MCARDLE, H.F., ROOKE, L.M. AND TOPPING, J.F. (1993). Tagging genomic sequences that direct transgene expression by activation of a promoter trap in plants. *Transgenic Research* **2**, 33–47.

LOCKHART, D.J., DONG, H., BYRNE, M.C. *ET AL.* (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnology* **14**, 1675–1680.

MATZ, M., SHAGIN, D., BOGDANOVA, E. *ET AL.* (1999). Amplification of cDNA ends based on template-switching effect and step-out PCR. *Nucleic Acids Research* **27**, 1558–1560.

MEDICO, E., GENTILE, A., LO CELSO, C. *ET AL.* (2001a). Osteopontin is an autocrine mediator of hepatocyte growth factor-induced invasive growth. *Cancer Research* **61**, 5861–5868.

MEDICO, E., GAMBAROTTA, G., GENTILE, A., COMOGLIO, P.M. AND SORIANO, P. (2001b). A gene trap vector system for identifying transcriptionally responsive genes. *Nature Biotechnology* **19**, 579–582.

NADEAU, J.H., BALLING, R., BARSH, G. *ET AL.* (2001). Sequence interpretation. Functional annotation of mouse genome sequences. *Science* **291**, 1251–1255.

NALDINI, L., BLOMER, U., GALLAY, P. *ET AL.* (1996). *In vivo* gene delivery and stable transduction of non-dividing cells by a lentiviral vector. *Science* **272**, 263–267.

NIWA, H., ARAKI, K., KIMURA, S., TANIGUCHI, S., WAKASUGI, S. AND YAMAMURA, K. (1993). An efficient gene-trap method using poly A trap vectors and characterization of gene-trap events. *Journal of Biochemistry (Tokyo)* **113**, 343–349.

O'KANE, C.J. AND GEHRING, W.J. (1987). Detection *in situ* of genomic regulatory elements in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America* **84**, 9123–9127.

REDDY, S., RAYBURN, H., VON MELCHNER, H. AND RULEY, H.E. (1992). Fluorescence-activated sorting of totipotent embryonic stem cells expressing developmentally regulated lacZ fusion genes. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 6721–6725.

RUSS, A.P., FRIEDEL, C., BALLAS, K. *ET AL.* (1996). Identification of genes induced by factor deprivation in haematopoietic cells undergoing apoptosis using gene-trap mutagenesis and site-specific recombination. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 15279–15284.

SABLITZKY, F., JONSSON, J.I., COHEN, B.L. AND PHILLIPS, R.A. (1993). High frequency expression of integrated proviruses derived from enhancer trap retroviruses. *Cell Growth and Differentiation* **4**, 451–459.

SALMINEN, M., MEYER, B.I. AND GRUSS, P. (1998). Efficient poly A trap approach allows the capture of genes specifically active in differentiated embryonic stem cells and in mouse embryos. *Developmental Dynamics* **212**, 326–333.

SCHENA, M., SHALON, D., DAVIS, R.W. AND BROWN, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270**, 467–470.

SCHRODER, A., SHINN, P., CHEN, H., BERRY, C., ECKER, J. AND BUSHMAN, F. (2002). HIV-1 integration in the human genome favours active genes and local hotspots. *Cell* **110**, 521.

SKARNES, W.C. (2000). Gene trapping methods for the identification and functional analysis of cell surface proteins in mice. *Methods in Enzymology* **328**, 592–615.

SKARNES, W.C., AUERBACH, B.A. AND JOYNER, A.L. (1992). A gene trap approach in mouse

embryonic stem cells: the lacZ reported is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice. *Genes and Development* **6**, 903–918.

SKARNES, W.C., MOSS, J.E., HURTLEY, S.M. AND BEDDINGTON, R.S. (1995). Capturing genes encoding membrane and secreted proteins important for mouse development. *Proceedings of the National Academy of Sciences of the United States of America* **92**, 6592–6596.

STANFORD, W.L., COHN, J.B. AND CORDES, S.P. (2001). Gene-trap mutagenesis: past, present and beyond. *Nature Reviews in Genetics* **2**, 756–768.

SUZUKI, H., FUKUNISHI, Y., KAGAWA, I. *ET AL.* (2001). Protein–protein interaction panel using mouse full-length cDNAs. *Genome Research* **11**, 1758–1765.

TATE, P., LEE, M., TWEEDIE, S., SKARNES, W.C. AND BICKMORE, W.A. (1998). Capturing novel mouse genes encoding chromosomal and other nuclear proteins. *Journal of Cell Science* **111**, 2575–2585.

THOMAS, K.R. AND CAPECCHI, M.R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**, 503–512.

TOWNLEY, D.J., AVERY, B.J., ROSEN, B. AND SKARNES, W.C. (1997). Rapid sequence analysis of gene trap integrations to generate a resource of insertional mutations in mice. *Genome Research* **7**, 293–298.

VAN LOHUIZEN, M., VERBEEK, S., SCHEIJEN, B., WIENTJENS, E., VAN DER, G.H. AND BERNS, A. (1991). Identification of cooperating oncogenes in E mu-myc transgenic mice by provirus tagging. *Cell* **65**, 737–752.

VENTER, J.C., ADAMS, M.D., MYERS, E.W. *ET AL.* (2001). The sequence of the human genome. *Science* **291**, 1304–1351.

WHITNEY, M., ROCKENSTEIN, E., CANTIN, G. *ET AL.* (1998). A genome-wide functional assay of signal transduction in living mammalian cells. *Nature Biotechnology* **16**, 1329–1333.

WILES, M.V., VAUTI, F., OTTE, J. *ET AL.* (2000). Establishment of a gene-trap sequence tag library to generate mutant mice from embryonic stem cells. *Nature Genetics* **24**, 13–14.

WURST, W., ROSSANT, J., PRIDEAUX, V. *ET AL.* (1995). A large-scale gene-trap screen for insertional mutations in developmentally regulated genes in mice. *Genetics* **139**, 889–899.

YOSHIDA, M., YAGI, T., FURUTA, Y. *ET AL.* (1995). A new strategy of gene trapping in ES cells using 3'RACE. *Transgenic Research* **4**, 277–287.

ZAMBROWICZ, B.P., FRIEDRICH, G.A., BUXTON, E.C., LILLEBERG, S.L., PERSON, C. AND SANDS, A.T. (1998). Disruption and sequence identification of 2000 genes in mouse embryonic stem cells. *Nature* **392**, 608–611.

ZHANG, G., GURTU, V. AND KAIN, S.R. (1996) An enhanced green fluorescent protein allows sensitive detection of gene transfer in mammalian cells. *Biochemical and Biophysical Research Communications* **227**, 707–711.

ZHU, H., KLEMIC, J.F., CHANG, S. *ET AL.* (2000). Analysis of yeast protein kinases using protein chips. *Nature Genetics* **26**, 283–289.

ZLOKARNIK, G., NEGULESCU, P.A., KNAPP, T.E. *ET AL.* (1998). Quantitation of transcription and clonal selection of single living cells with beta-lactamase as reporter. *Science* **279**, 84–88.