expert reviews
in molecular medicine



# Searching the unknown with gene trapping

**Majlinda Lako and Nicholas Hole**

**A dramatic increase in sequence information, both in the form of complementary DNA (cDNA) and genomic DNA, has created a huge gap between the discovery of genes and the process of identifying gene function. To fill this gap, the 'gene-trapping' approach has been developed; this combines into a single process the three stages of gene cloning, the study of the pattern of gene expression and the analysis of the respective mutant phenotype. Recent results indicate that gene trapping can be used successfully to clone specific genes that are involved in the development of the central nervous system, limbs and haematopoietic system. Continuous improvements in the design of trapping vectors, faster sequencing of cDNA clones and more-efficient in vitro pre-screening will certainly aid the large-scale trapping of mammalian genomes.**

Large quantities of DNA sequence data are rapidly being accumulated in public databases (Ref. 1). Despite this tremendous progress, and the fact that it provides an important foundation for the design of key experiments, there is increasing concern that this information will not be sufficient to determine the biological function of proteins from sequence data alone. In the past, one of the most successful approaches for determining gene function has been saturation screening. In these functional screens, mutagenic agents are used to produce a large number of lines of an organism, which are then inbred for several generations. By matching phenotype to genetic lesion, many of the genes that are involved in the developmental process have been identified. This approach has been carried out successfully in the fruitfly *Drosophila melanogaster* (Ref. 2), the nematode *Caenorhabditis elegans* (Ref. 3) and the flowering plant *Arabidopsis thaliana* (Ref. 4). However, although great insights have been derived from the study of these organisms, the mouse remains the model of choice for studying the function of human genes and the mechanisms underlying human genetic disorders.

## Strategies for identifying gene function

### *Gene knock-outs*

In mice, phenotype-driven approaches have been developed that employ mutagenesis to recover novel phenotypes without prior assumptions being made about the underlying gene or pathway that has been disrupted (Ref. 5). Unfortunately, these approaches have not proven

Nicholas Hole (corresponding author)
Lecturer, Biological Sciences, University of Durham, South Road, Durham, DH1 3LE, UK.
Tel: +44 (0)191 374 2414; Fax: +44 (0)191 374 2417; E-mail: Nicholas.Hole@durham.ac.uk

Majlinda Lako
Postdoctoral Fellow, Biological Sciences, University of Durham, South Road, Durham, DH1 3LE, UK. Tel: +44 (0)191 374 2414; Fax: +44 (0)191 374 2417; E-mail: Majlinda.Lako@durham.ac.uk



**Accession information:** (00)00182-4a.pdf (short code: txt001nhd); 6 July 2000
ISSN 1462-3994 ©2000 Cambridge University Press

https://doi.org/10.1017/S1462399400001824 Published online by Cambridge University Press

as successful as those in simpler eukaryotes. Reasons for this include the long generation time, the large space requirements and the laborious handling required to maintain and breed the mice. However, the situation has changed dramatically with the advent of embryonic stem (ES) cell lines and the means to generate and select for their genetic alteration (Ref. 6).

ES cells can be maintained in culture as totipotent cells (i.e. cells that can give rise to all types of differentiated cells) under proper growth conditions and can be genetically altered with comparative ease (for a review, see Ref. 7). When genetically altered ES cells from mice are injected into a host blastocyst (embryo), they have the capacity to contribute to all tissues of the 'chimaeric' animal (i.e. an animal whose body contains cell populations derived from different zygotes). Furthermore, they can contribute to germ cells and transmit the genetic alteration through the germline, producing mutant lines of mice. This process has been used extensively in combination with directed (or targeted) genomic mutagenesis to produce mice that have a null allele for specific gene loci (i.e. an allele that is no longer capable of producing the native gene product). Such heterozygous mutant animals can be crossed to produce homozygous 'gene knock-out' mice (Ref. 7).

Although the gene knock-out approach has produced remarkable insights into specific areas of biological research, its use is compromised by difficulties. Clearly, directed mutagenesis, typically by homologous recombination, requires information on the genomic sequence of the gene that is to be targeted. Furthermore, because targeting constructs must, in general, be isogenic with the ES cells that are to be targeted, the nature of the genomic DNA that such constructs can be composed of is restricted; thus, most genomic libraries are not suitable. The production of germline-competent chimaeric mice (i.e. mice whose germ cells are ultimately derived from ES cells) can take several months, before they can be used to establish a heterozygous mutant line, which in turn needs to be crossed to produce homozygous animals. All these stages take considerable time, and there is no guarantee that the resultant homozygous animals will have an interpretable phenotype. Indeed, animals that are homozygous for a null allele may die in utero. In reality, the large-scale phenotypic screening of knock-out animals that have been produced by directed mutagenesis is often too expensive, both in terms of time and resources, to pursue.

### *Gene trapping*

During the early 1990s, alternative strategies for identifying gene function were explored. In particular, the possibility was investigated of whether the directed screening of libraries of random genomic mutants could be used to identify classes of genes that were relevant to particular biological systems. This approach has been termed 'gene trapping'.

The principle behind gene trapping is essentially the random insertion of a DNA vector that is designed to signal its presence via the activation of a reporter gene, which both mimics the expression of the endogenous gene and potentially mutates the locus (Ref. 8). By selecting in cell culture those ES cells in which the reporter gene has been activated and transferring them to a host embryo, mutant mice can be generated far more quickly than by conventional gene mutation because there is no need to target each gene separately. Huge libraries of clones of ES cells that bear random integrations can therefore be isolated and stored indefinitely. The sequence of the 'trapped' gene can also be identified using techniques that are based on the polymerase chain reaction (PCR), and this can lead to the isolation of novel genes regardless of their level of expression in vivo (Ref. 9). The entire approach of cloning a new gene, studying its pattern of expression in vivo and simultaneously creating a mutant mouse has made gene trapping one of the most attractive approaches available to developmental biologists. Its applications extend from the random disruption and identification of genes in mouse ES cells (Refs 10, 11) to the capture of genes that encode both membrane and secreted proteins (Ref. 12), genes that respond to retinoic acid (Ref. 13), genes that are specifically activated in differentiated mouse ES cells (Ref. 14) and genes that are important in the development of the mammalian nervous system (Ref. 15).

## Design of gene-trap vectors

Two main types of entrapment vectors have been described: (1) 'enhancer-trap' vectors, which must be integrated near a cellular enhancer to activate the reporter gene that is fused to a minimal promoter (Ref. 16) and (2) gene-trap vectors, which contain a splice acceptor (SA) site positioned at the 5′ end of the reporter gene,



**Accession information:** (00)00182-4a.pdf (short code: txt001nhd); 6 July 2000
ISSN 1462-3994 ©2000 Cambridge University Press

https://doi.org/10.1017/S1462399400001824 Published online by Cambridge University Press

allowing fusion transcripts to be generated after the integration has occurred within the introns of an actively transcribed gene (Refs 10, 13, 17, 18, 19, 20). Gene-trap vectors have been widely used in mouse ES cells, and are discussed more fully in the rest of this review.

For any gene-trap vector, it is important that three criteria are fulfilled: (1) the vector must provide a suitable reporter for the analysis of gene expression; (2) the 'trap event' must mutate the gene and (3) the sequence of both the trapped cDNA and the genomic site of integration must be straightforward. Gene-trap vectors are introduced into the target cells (normally ES cells) either by electroporation (Ref. 21) or using retroviral vectors (Ref. 22). Electroporation facilitates the entry of foreign DNA into living cells by temporarily permeabilising cell membranes. Although this technique has been used extensively, its efficiency is quite low; retroviral vectors provide a higher transfection frequency and integrate as an intact single copy.

## First-generation vectors

The earliest vectors were used to trap genes that were active in undifferentiated ES cells (Refs 10, 22). The essential feature of these early gene-trap vectors is the placement of an SA site in front of a promoterless reporter gene such as *lacZ* (which encodes the enzyme β-galactosidase; Ref. 10) or *β-geo* [which is formed from the β-galactosidase gene (*β-gal*) and the neomycin-resistance gene (*neo*) and encodes a fusion protein; Ref. 22]. The integration of the vector into the intron of an expressed gene in the correct orientation results in a fusion messenger RNA (mRNA) transcript. Furthermore, if the reading frames of the reporter gene and the endogenous gene are the same, an active protein is produced that contains the *N*-terminal portion of the native protein fused to the reporter protein. Recent improvements to this type of vector include the insertion of an internal ribosome entry site (IRES) from the encephalomyocarditis virus between the SA site and the reporter gene sequence (see Fig. 1; Ref. 23). The use of the IRES allows di-cistronic translation (i.e. two proteins can be translated from one mRNA), and therefore the reporter gene can be translated even when it is not fused to the trapped gene. It is important to note that the level of expression of the reporter gene remains dependent on the rate of transcription of the mRNA of the trapped gene.

The expression of the neomycin-selectable marker gene (*neo*) confers resistance to the selective drug G418, and can therefore be used to facilitate the identification of clones of ES cells in which gene trapping has occurred. Moreover, by monitoring *lacZ* gene activity in mouse embryos, it is possible to visualise the expression pattern of the endogenous gene during development. β-galactosidase activity identifies those tissues that would normally express the gene, and helps identify whether the protein exerts its effects solely on the cells in which it is expressed. In vectors in which *β-geo* is the reporter gene, the combined β-galactosidase and neomycin-resistance genes ensure that clones that are selected by resistance to G418 will also express the marker gene activity (i.e. β-galactosidase or β-gal activity). β-galactosidase activity can be detected using the chromogenic substrate X-gal, which yields a blue colour. Selection on the basis of resistance to G418 has been shown to be more sensitive than X-gal staining. Consequently, at least some of the clones that are negative for X-gal (i.e. stain white rather than blue) may contain weakly expressed genes. Thus, the combined *β-geo* reporter gene increases the efficiency of gene trapping and ensures that genes that are expressed at low levels can also be trapped (Ref. 22).

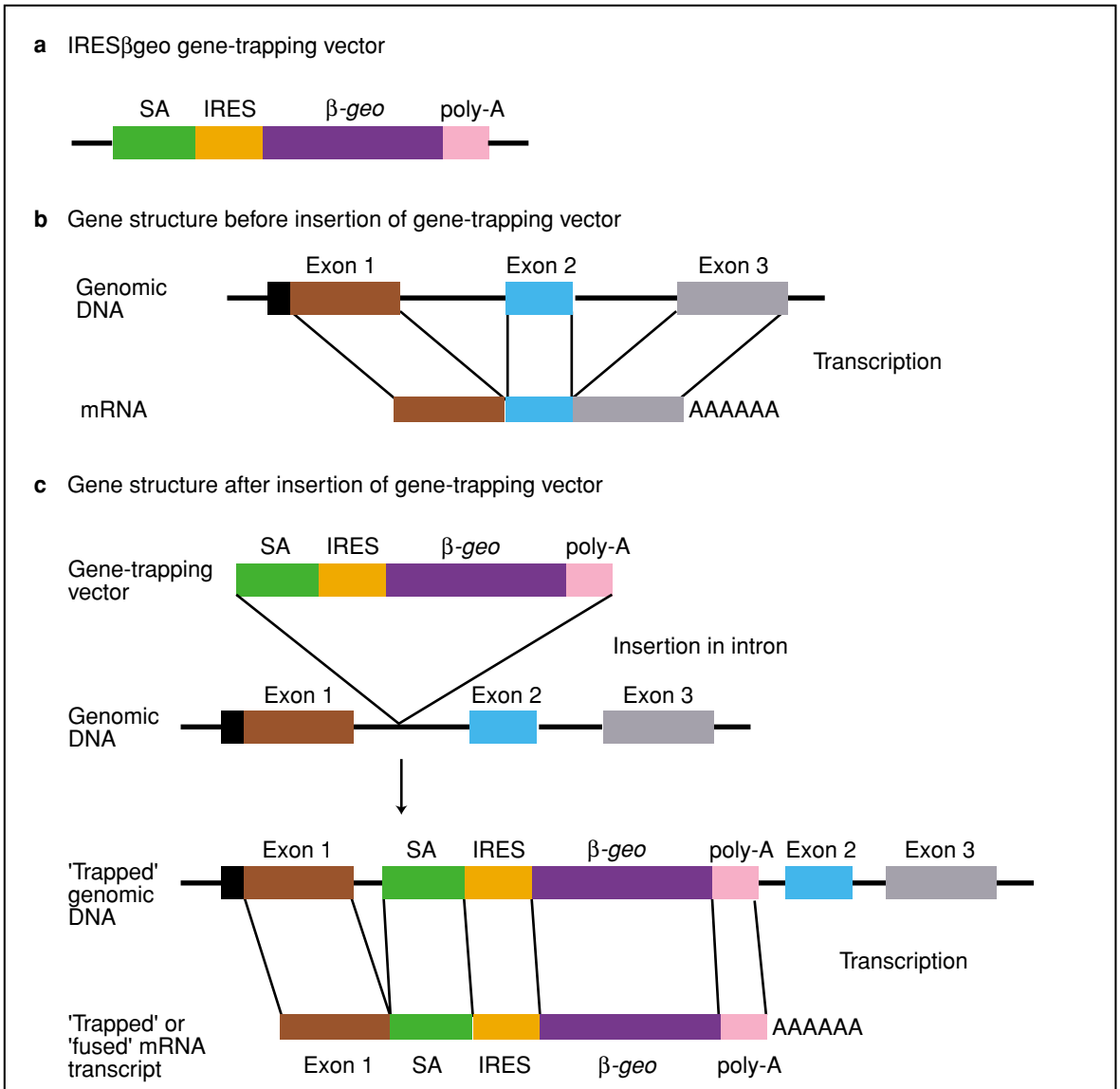
The most notable limitation of these early vectors is that they target only those genes that are actively transcribed; this restricts the set of genes that can be mutated in any particular cell type. Thus, if transcription in ES cells is poor or absent, for example, drug selection might not select all of the clones that contain trapped genes.

The sequence information gathered from the analysis of trapped genes often corresponds to 5′-untranslated regions (UTRs); furthermore, gene identification becomes difficult because such information is under-represented in expressed databases of sequence tags (ESTs) and is less conserved between species. ESTs are partial DNA sequences (sequence tags) that are derived from cDNA clones; they are used in genome sequencing studies to identify transcribed sequences in genomic DNA sequences. In addition, the allele may be mutated, but the translation of the reporter gene may be compromised. This can occur when there are problems in either appropriate splicing (i.e. splice around) or transcriptional termination (i.e. read through). Thus, a second generation of gene-trapping vectors has been developed to circumvent these problems.



**Accession information:** (00)00182-4a.pdf (short code: txt001nhd); 6 July 2000
ISSN 1462-3994 ©2000 Cambridge University Press

https://doi.org/10.1017/S1462399400001824 Published online by Cambridge University Press




The IRESβgeo gene-trapping vector and gene-trapping strategy

Expert Reviews in Molecular Medicine © 2000 Cambridge University Press

**Figure 1. The IRESβgeo gene-trapping vector and gene-trapping strategy.** (a) The IRESβgeo gene-trapping vector, showing the splice acceptor (SA) site, the internal ribosomal entry site (IRES), the *β-geo* gene (a fusion gene formed from the β-galactosidase gene and the neomycin-resistance gene) and the polyadenylation (poly-A) site. (b) Conventional exon and intron gene structure, showing the splicing together of the exons to produce messenger RNA (mRNA). (c) The IRESβgeo gene-trapping strategy. On transcription to form mRNA, the SA allows the vector to be spliced to exon 1. Transcription is terminated before exon 2 because of the existence of a poly-A site on the IRESβgeo vector and the absence of a splice donor site on the vector. The expression of the resultant fusion protein is dependent on the rate of transcription of the mRNA of the trapped gene. The presence of the IRES ensures di-cistronic translation, allowing the translation of the *β-geo* marker (i.e. selection) gene without the need to produce a protein fused to the trapped gene. Because *β-geo* is a combination of the β-galactosidase gene and a neomycin-resistance gene, clones that are selected by resistance to neomycin will express the marker gene activity **(fig001nhd)**.



**Accession information:** (00)00182-4a.pdf (short code: txt001nhd); 6 July 2000
ISSN 1462-3994 ©2000 Cambridge University Press

https://doi.org/10.1017/S1462399400001824 Published online by Cambridge University Press

### Second-generation vectors

Second-generation vectors do not have a poly-A site that directs the addition of a poly-A tail (i.e. a polyadenylic sequence of varying length at the 3' end of mRNA) at the end of the selectable marker gene; thus, a poly-A signal must be provided by the endogenous mouse gene to produce stable mRNA (see Fig. 2; Refs 11, 14, 24). Because the expression of the selectable marker gene is under the control of a constitutive promoter (usually mouse phosphoglycerate kinase-1 promoter), a gene can be trapped without being transcribed. The sequence of the trapped gene can be identified using RACE (rapid amplification of mRNA ends by PCR; a process that produces amplified DNA using only one sequence-specific primer). The identified sequence often corresponds to that of coding regions or 3′ UTRs, which are more useful than 5′ UTRs for gene identification. Furthermore, because the reporter gene is under the transcriptional control of the trapped locus, it is possible to isolate genes that change their pattern of expression during differentiation (Ref. 14).

Such so-called splice acceptor (SA) vectors must integrate within an intron, and consequently favour genes that contain longer introns because they represent a larger target for insertion. It therefore seems logical to develop a gene-trap system that functions independently of the splicing machinery (which removes introns, producing mRNA that contains only exons). In 1997, Hicks and colleagues reported the feasibility of a large gene-trap screen using 'SA-less' exon-trap vectors (Ref. 25). This was achieved using a retroviral shuttle vector, U3NeoSV1, which carried a promoterless *neo* resistance gene in the U3 region of the long terminal repeat (LTR). Because the reporter gene is expressed only when it is integrated into an exon, resistance to neomycin will select only those cells in which retroviral integration has occurred into an expressed gene.

## Improving the technique

The simultaneous identification and mutation of previously unidentified genes in the mouse has made gene trapping an attractive alternative for the functional analysis of the mouse genome (Ref. 8). Experience gathered from small-scale experiments and improvements in trapping vectors suggest that large-scale gene trapping is quite feasible (Refs 19, 26). However, to optimise such large-scale trapping, several areas of the strategy require attention.

### Cloning and sequencing

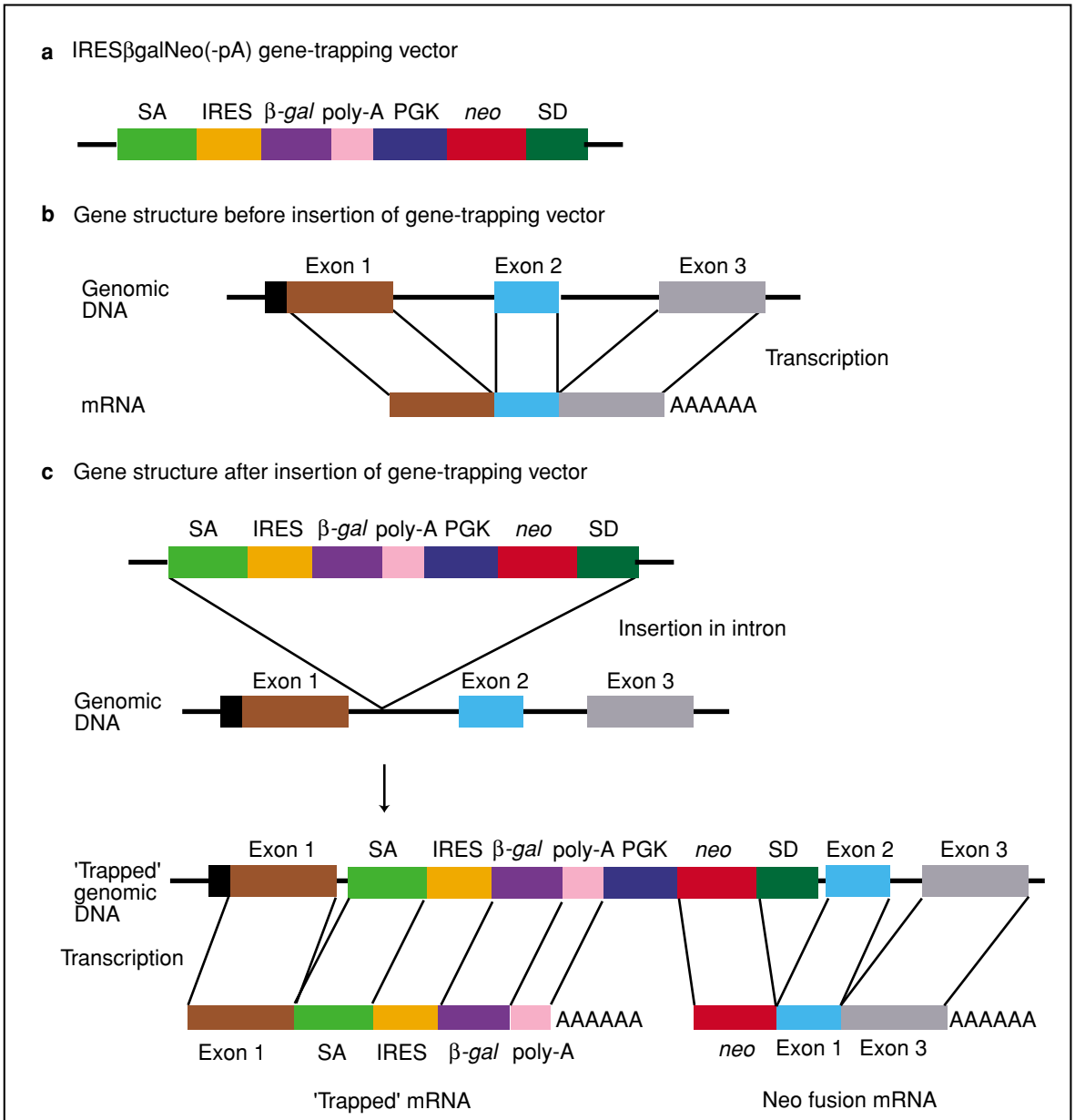
The identification of endogenous genes that are associated with gene-trap insertions has proved to be a major bottleneck, which is reflected by the fact that, to date, only a handful of genes have been isolated from gene traps (Refs 10, 13, 25). Most gene-trapping strategies have relied on PCR-based techniques to isolate the gene sequences that are fused at their 5′ or 3′ ends to the marker gene of the trap vector (Ref. 9). Although the amplification of specific cDNAs can be achieved rapidly, the subsequent cloning and screening of informative-sized cDNAs has proven to be time-consuming. This has recently been circumvented by using microdialysis and solid-phase sequencing of 5′ RACE products, which lend themselves to automation and fluorescent sequencing (Ref. 27). The high-throughput capacity of these techniques enables the identification of genes and selection of mutations before germline transmission. This is particularly important in view of the large percentage of aberrant events that can be associated with ineffective splicing (Refs 26, 27, 28).

In view of the current accumulation of ESTs in databases, it also seems logical to develop a 'tag-sequencing' procedure to generate only short tags of the sequences derived from the trapped genes. Such tags can then be entered in the nucleic acid databases to identify the trapped gene. When a sequence matches a known gene or EST, attention should be paid to the information already existing in the databases regarding its expression or the tissue it was isolated from; thus, only those genes of interest to a particular study can be selected before germline transmission. When there is no match, or when the sequence obtained from the gene trap is longer than the one reported in the database, the cDNA clone recovered from the trapped gene should be completely sequenced. Because many researchers still use SA vectors, it is likely that the generated sequences will correspond to 5′ UTRs, which might be unknown for many of the genes that are only represented in sequence databases by ESTs. The submission of these sequences to the databases will make an additional contribution towards the complete sequencing of the mouse genome.



**Accession information:** (00)00182-4a.pdf (short code: txt001nhd); 6 July 2000
ISSN 1462-3994 ©2000 Cambridge University Press

https://doi.org/10.1017/S1462399400001824 Published online by Cambridge University Press




The IRESβgalNeo(-pA) gene-trapping vector and gene-trapping strategy

Expert Reviews in Molecular Medicine © 2000 Cambridge University Press

**Figure 2. The IRESβgalNeo(-pA) gene-trapping vector and gene-trapping strategy.** (a) The IRESβgalNeo(-pA) gene-trapping vector, showing the splice acceptor (SA) site, the internal ribosomal entry site (IRES), the β-galactosidase gene (*β-gal*), the neomycin-resistance gene (*neo*), the phosphoglycerate kinase-1 (PGK) promoter, the splice donor (SD) site and the polyadenylation (poly-A) site. (b) The IRESβgalNeo(-pA) gene-trapping strategy. IRESβgalNeo(-pA) uncouples selection from marker expression, and therefore helps to trap more genes than the IRESβgeo gene-trapping strategy outlined in Figure 1. The neomycin-resistance gene (*neo*), which is used for clone selection is driven by the PGK promotor, a 'housekeeping' gene; this eliminates the reliance on endogenous gene expression to drive clone selection. However, expression of the marker gene (*β-gal*) remains reliant on the transcriptional machinery of the trapped gene **(fig002nhd).**



**Accession information:** (00)00182-4a.pdf (short code: txt001nhd); 6 July 2000
ISSN 1462-3994 ©2000 Cambridge University Press

https://doi.org/10.1017/S1462399400001824 Published online by Cambridge University Press

### In vitro pre-screening

Large-scale gene trapping lends itself to the generation of a large number of clones, all of which cannot be analysed at the same time; thus, some pre-selection guidelines should be set out in advance. The developmental expression pattern of the endogenous gene that is mimicked by the expression of the reporter gene could provide criteria for the selection of genes. However, it should be remembered that, according to recent evidence, the expression of β-galactosidase can sometimes be more restricted than that of the endogenous gene (Refs 26, 28). There are two possible explanations for this observed phenomenon. First, the trapped gene is normally subject to 'alternative splicing'. Eukaryotic genes are composed of exons and introns. Introns are removed by RNA splicing before transcribed mRNA leaves the nucleus. Often, a single gene can encode several different mRNA transcripts, caused by the cell- or tissue-specific combination of different exons. This is known as alternative splicing. Thus, splicing around the vector-containing intron in some tissues or cell types could generate wild-type transcripts and the loss of β-galactosidase activity (Ref. 28). Second, vector insertion during gene trapping may have disrupted the enhancer elements that are necessary for gene activity in some tissues. This second explanation is supported by increasing evidence that tissue-specific enhancers can reside in introns (Refs 29, 30).

Because there is no guarantee that the trapped gene will display an interesting expression pattern or that mutated ES cell clones will show an obvious phenotype, an in vitro pre-screen that limits the number of gene-trap events that are transmitted through the germline would be very useful. Because ES cells can differentiate into many different lineages in vitro, it should be possible to screen for gene-trap integrations that would affect the differentiation of these lineages. For example, the in vitro differentiation of ES cells into haematopoietic lineages can be enhanced by exposure to dimethyl sulphoxide or repressed by exposure to retinoic acid (Ref. 31). Therefore, it may be possible to use this type of direct screening approach to select for gene-trap integrations into genes that affect the haematopoietic lineage. This type of screen has now been adapted to screen for genes that are expressed in ES cells after they have been exposed in vitro to specific growth factors (e.g. nerve growth factor; Ref. 15) or differentiation factors (e.g. retinoic acid; Ref. 19). In addition, this procedure can be used to screen for genes that are expressed only after differentiation in suspension culture (see Fig. 3; Refs 14, 32). Such an in vitro pre-screen strategy has successfully identified haematopoietic genes (Refs 33, 34).

### Absence of phenotype

Even after pre-selection, some of the mouse lines that are created will not display an overt phenotype (Ref. 28). One likely explanation for this is an increase in splicing around the vector-containing intron, such that sufficient quantities of wild-type protein are generated to maintain the appropriate function. In other cases, the integration event may not disrupt the gene as expected, thereby creating hypomorphic or hypermorphic alleles. These alleles are particularly useful for studying the function of genes that lead to lethality if they are knocked out completely, impeding the analysis of their function at later stages of development. Alternatively, the mutated gene might be compensated for by related members of the same gene family, thus leading to a lack of overt phenotype in homozygous embryos.

### Repeat trapping

A further problem with gene trapping is that some genes may be more likely to be trapped than others. According to two reports, the same gene has been trapped twice using different trapping strategies (Refs 8, 32, 35). Research carried out by Chowdhury and colleagues clearly indicates that there are no preferred sites for integration into the mouse genome (Ref. 23). However, the use of one type of vector in each study could limit the number of genes that are likely to be targeted; therefore, the use of several vectors is recommended (Ref. 8). The construction of an international database reporting all identified gene traps and the analysis of their expression patterns, combined with the adoption of the 'sequence-first' approach, should help to prevent the same gene being trapped twice and increase the efficiency of the gene-trapping procedure.

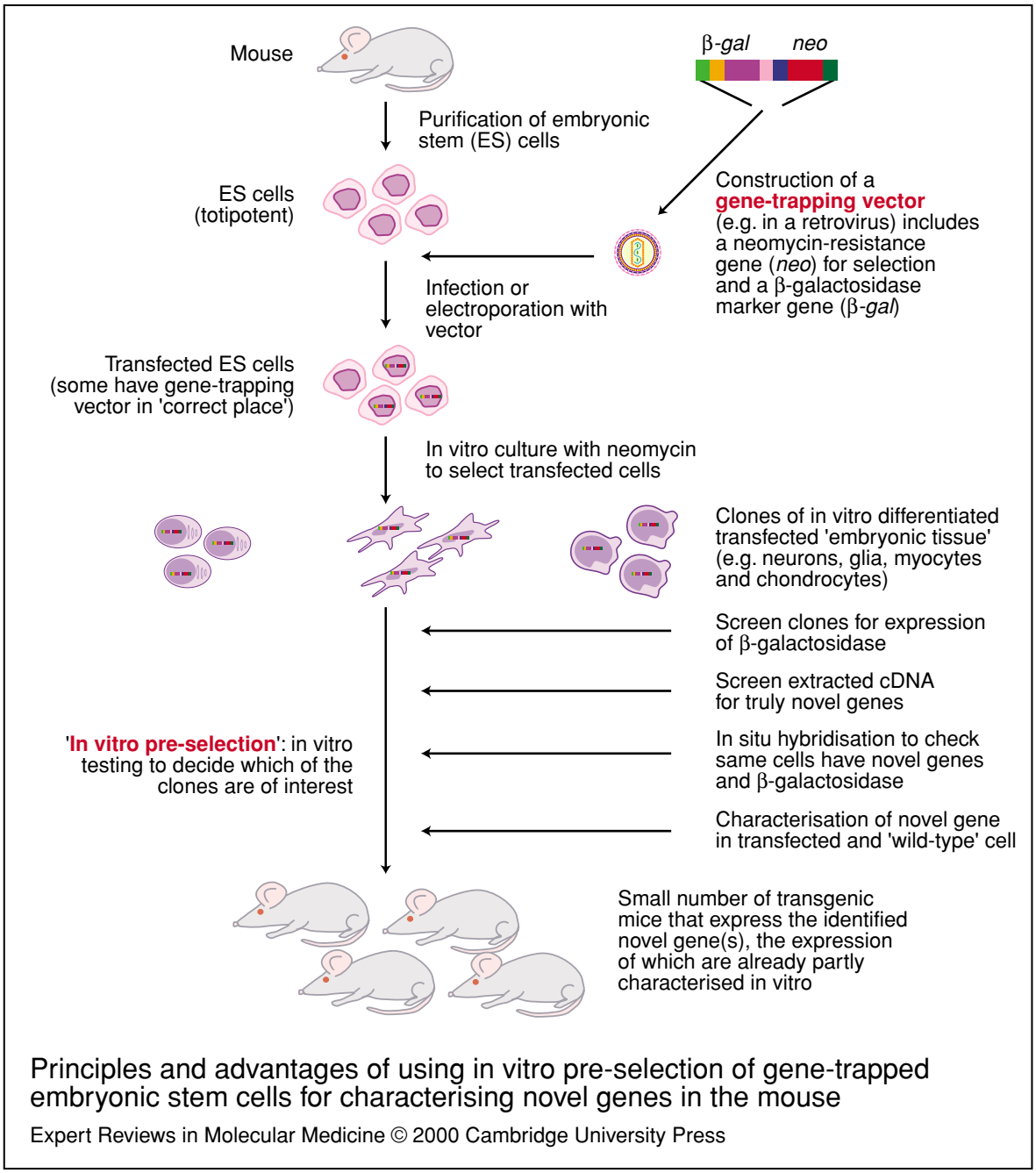
### Flexibility of mutation

Conventional gene trapping is good at producing null (or 'loss of function') mutations in trapped genes. One possible shortcoming is that this is a sledgehammer approach; thus, it might be



**Accession information:** (00)00182-4a.pdf (short code: txt001nhd); 6 July 2000
ISSN 1462-3994 ©2000 Cambridge University Press

https://doi.org/10.1017/S1462399400001824 Published online by Cambridge University Press




Principles and advantages of using in vitro pre-selection of gene-trapped embryonic stem cells for characterising novel genes in the mouse

Expert Reviews in Molecular Medicine © 2000 Cambridge University Press

**Figure 3. Principles and advantages of using in vitro pre-selection of gene-trapped embryonic cells for characterising novel genes in the mouse.** Totipotent embryonic stem (ES) cells from a mouse are infected with a gene-trap vector; here, the retroviral construct includes a selectable *lacZ–neo* fusion gene, which is expressed only if the gene trap inserts within an active transcription unit. Transfected ES cells expressing the *neo* gene are selected for in vitro with neomycin. Clones of ES cells are induced to differentiate in vitro, and different cell types are screened for β-galactosidase activity. cDNAs of the gene-trap transcripts are obtained by 5' rapid amplification of cDNA ends and sequenced. In situ hybridisation analyses show that trapped genes are expressed in vivo within the cell types that express β-galactosidase in vitro. Gene traps and their wild-type alleles are then further characterised. This approach is more economical than standard in vivo gene-trap screening because it is possible to determine if likely knock-out alleles are expressed in one or more specific embryonic tissues before generating chimaeric or transgenic mice **(fig003nhd)**.



**Accession information:** (00)00182-4a.pdf (short code: txt001nhd); 6 July 2000
ISSN 1462-3994 ©2000 Cambridge University Press

https://doi.org/10.1017/S1462399400001824 Published online by Cambridge University Press

more useful in some cases to be able to produce less severe mutations. Recent research has generated gene-trap vectors that incorporate lox sites for the cre recombinase (Ref. 36), which should allow greater control of the nature of the mutations generated. In this process, cre recombinase will specifically excise and/or insert sequences that are flanked by lox sites. This allows the conditional replacement of vectors in the trapped genome with more-subtle mutations than are produced normally; examples could include point mutations and/or alternative phenotypic markers.

## Concluding remarks

The study of genomics is rapidly accelerating. New strategies for comparing the expression of thousands, or even tens of thousands, of genes are developing (for example, see Ref. 37). However the study of functional, rather than descriptive, genomics has not kept pace. Strategies such as gene trapping have begun to expand the horizons of researchers who are studying functional genomics, in as much as whole classes of genes can be studied and their mutations examined. However, as discussed earlier, gene trapping is still an emerging technology. It is anticipated that further advances will result in an even more powerful technique for the study of functional genomics, which will help investigators dissect the genetic control of biological processes. These advances could include the development of new and more-subtle targeting vectors and, in particular, the compilation of better in vitro pre-screened and defined libraries (such as those recently described by Wiles and colleagues; Ref. 38).

## Clinical implications

To date, gene trapping has been most useful for researching developmental systems, typically embryogenesis. As more-sophisticated gene-trapping vectors are developed, it is likely that this technique will be applicable to a wider range of systems, typically those involving differentiation or pharmacological response. Thus, it should prove increasingly useful as a tool for identifying targets for therapeutic intervention.

## Acknowledgements

Work contributing to this article was supported by project grants from the Wellcome Trust. We thank Dr Lesley Forrester (John Hughes Bennett Laboratories, University of Edinburgh, UK) and Dr Colin Jahoda (Biological Sciences, University of Durham, UK) for helpful criticism of this article.

## References

1 Schuler, G.D. et al. (1996) A gene map of the human genome. Science 274, 540-546, PubMed ID: 97002443
2 Nusslein-Volhard, C. and Wieschaus, E. (1980) Mutations affecting segment number and polarity in Drosophila. Nature 287, 795-801, PubMed ID: 81052383
3 Brenner, S. (1974) The genetics of Caenorhabditis elegans. Genetics 77, 71-94, PubMed ID: 74271161
4 Mayer, U. et al. (1991) Mutations affecting body organization in the Arabidopsis embryo. Nature 353, 402-407
5 Brown, S.D. (1998) Mouse models of genetic disease: new approaches, new paradigms. J Inherit Metab Dis 21, 532-539, PubMed ID: 98397579
6 Evans, M.J. and Kaufman, M.H. (1981) Establishment in culture of pluripotential cells from mouse embryos. Nature 292, 154-156, PubMed ID: 81220990
7 Thomas, K.R. and Capecchi, M.R. (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. Cell 51, 503-512, PubMed ID: 88027032
8 Evans, M.J., Carlton, M.B. and Russ, A.P. (1997) Gene trapping and functional genomics. Trends Genet 13, 370-374, PubMed ID: 97433599
9 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. Proc Natl Acad Sci U S A 85, 8998-9002, PubMed ID: 89057866
10 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 Dev 6, 903-918, PubMed ID: 92275355
11 Zambrowicz, B.P. et al. (1998) Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. Nature 392, 608-611, PubMed ID: 98219085
12 Skarnes, W.C. et al. (1995) Capturing genes encoding membrane and secreted proteins important for mouse development. Proc Natl



**Accession information:** (00)00182-4a.pdf (short code: txt001nhd); 6 July 2000
ISSN 1462-3994 ©2000 Cambridge University Press

https://doi.org/10.1017/S1462399400001824 Published online by Cambridge University Press

Acad Sci U S A 92, 6592-6596, PubMed ID: 95327693
13 Forrester, L.M. et al. (1996) An induction gene trap screen in embryonic stem cells: Identification of genes that respond to retinoic acid in vitro. Proc Natl Acad Sci U S A 93, 1677-1682, PubMed ID: 96202336
14 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. Dev Dyn 212, 326-333, PubMed ID: 98289809
15 Stoykova, A. et al. (1998) Gene trap expression and mutational analysis for genes involved in the development of the mammalian nervous system. Dev Dyn 212, 198-213, PubMed ID: 98289797
16 Bellen, H.J. et al. (1989) P-element-mediated enhancer detection: a versatile method to study development in Drosophila. Genes Dev 3, 1288-1300, PubMed ID: 90108679
17 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. Proc Natl Acad Sci U S A 86, 5517-5521, PubMed ID: 89315828
18 von Melchner, H. et al. (1992) Selective disruption of genes expressed in totipotent embryonal stem cells. Genes Dev 6, 919-927, PubMed ID: 92275356
19 Wurst, W. et al. (1995) A large-scale gene-trap screen for insertional mutations in developmentally regulated genes in mice. Genetics 139, 889-899, PubMed ID: 95229068
20 Xiong, J.W. et al. (1998) Large-scale screening for developmental genes in embryonic stem cells and embryoid bodies using retroviral entrapment vectors. Dev Dyn 212, 181-197, PubMed ID: 98289796
21 Schuster-Gossler, K. et al. (1998) The mouse Gtl2 gene is differentially expressed during embryonic development, encodes multiple alternatively spliced transcripts, and may act as an RNA. Dev Dyn 212, 214-228, PubMed ID: 98289798
22 Friedrich, G. and Soriano, P. (1991) Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. Genes Dev 5, 1513-1523, PubMed ID: 91357468
23 Chowdhury, K. et al. (1997) Evidence for the stochastic integration of gene trap vectors into the mouse germline. Nucleic Acids Res 25, 1531-1536, PubMed ID: 97248607
24 Yoshida, M. et al. (1995) A new strategy of gene trapping in ES cells using 3'RACE. Transgenic Res 4, 277-287, PubMed ID: 95384208
25 Hicks, G.G. et al. (1997) Functional genomics in mice by tagged sequence mutagenesis. Nat Genet 16, 338-344, PubMed ID: 97385236
26 Voss, A.K., Thomas, T. and Gruss, P. (1998) Efficiency assessment of the gene trap approach. Dev Dyn 212, 171-180, PubMed ID: 98289795
27 Townley, D.J. et al. (1997) Rapid sequence analysis of gene trap integrations to generate a resource of insertional mutations in mice. Genome Res 7, 293-298, PubMed ID: 97228906
28 Voss, A.K., Thomas, T. and Gruss, P. (1998) Compensation for a gene trap mutation in the murine microtubule- associated protein 4 locus by alternative polyadenylation and alternative splicing. Dev Dyn 212, 258-266, PubMed ID: 98289802
29 Lothian, C. and Lendahl, U. (1997) An evolutionarily conserved region in the second intron of the human nestin gene directs gene expression to CNS progenitor cells and to early neural crest cells. Eur J Neurosci 9, 452-462, PubMed ID: 97258113
30 Haerry, T.E. and Gehring, W.J. (1997) A conserved cluster of homeodomain binding sites in the mouse Hoxa-4 intron functions in Drosophila embryos as an enhancer that is directly regulated by Ultrabithorax. Dev Biol 186, 1-15, PubMed ID: 97332514
31 Hole, N. et al. (1996) A limited temporal window for the derivation of multilineage repopulating hematopoietic progenitors during embryonal stem cell differentiation in vitro. Blood 88, 1266-1276, PubMed ID: 96329559
32 Baker, R.K. et al. (1997) In vitro preselection of gene-trapped embryonic stem cell clones for characterizing novel developmentally regulated genes in the mouse. Dev Biol 185, 201-214, PubMed ID: 97330624
33 Stanford, W.L. et al. (1998) Expression trapping: identification of novel genes expressed in hematopoietic and endothelial lineages by gene trapping in ES cells. Blood 92, 4622-4631, PubMed ID: 99061779
34 Hidaka, M. et al. (2000) Gene trapping of two novel genes, Hzf and Hhl, expressed in hematopoietic cells. Mech Dev 90, 3-15, PubMed





**Accession information:** (00)00182-4a.pdf (short code: txt001nhd); 6 July 2000
ISSN 1462-3994 ©2000 Cambridge University Press

https://doi.org/10.1017/S1462399400001824 Published online by Cambridge University Press

ID: 20054088
35 Takeuchi, T. et al. (1995) Gene trap capture of a novel mouse gene, jumonji, required for neural tube formation. Genes Dev 9, 1211-1222, PubMed ID: 95278734
36 Araki, K. et al. (1999) Exchangeable gene trap using the Cre/mutated lox system. Cell Mol Biol 45, 737-750, PubMed ID: 99440754
37 Alizadeh, A.A. et al. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 403, 503-511, PubMed ID: 20140121
38 Wiles, M.V. et al. (2000) Establishment of a gene-trap sequence tag library to generate mutant mice from embryonic stem cells. Nat Genet 24, 13-14, PubMed ID: 20082798

## Further reading, resources and contacts

The world leader in mouse genetics has for many years been the Jackson Laboratory, Bar Harbor, Maine, USA. In collaboration with several different institutions, they have established first-rank databases concerned with mouse genomics that are relevant to gene trapping.

The Transgenic/Targeted Mutation Database (TMD) is a thorough database of genomic mutations in animal models. It is largely (but not exclusively) based on mouse models.
http://tbase.jax.org/

Mouse Genome Informatics. This website houses the Gene Expression Database (GXD), which is a searchable database that compiles reports on gene expression in mice. This URL is the UK mirror site of the Jackson Laboratory site.
http://www.informatics.jax.org/mgihome/GXD/aboutGXD.shtml

The Whole Mouse Catalog is a slightly more esoteric website that compiles data from many different sources, which are not just restricted to genomic mutation.
http://www.rodentia.com/wmc/

The Mouse Atlas and Gene Expression Database Project website provides developing resource of gene expression information.
http://genex.hgu.mrc.ac.uk/

## Features associated with this article

### Figures

Figure 1. The IRESβgeo gene-trapping vector and gene-trapping strategy (fig001nhd).
Figure 2. The IRESβgalNeo(-pA) gene-trapping vector and gene-trapping strategy (fig002nhd).
Figure 3. Principles and advantages of using in vitro pre-selection of gene-trapped embryonic cells for characterising novel genes in the mouse (fig003nhd).





**Accession information:** (00)00182-4a.pdf (short code: txt001nhd); 6 July 2000
ISSN 1462-3994 ©2000 Cambridge University Press

https://doi.org/10.1017/S1462399400001824 Published online by Cambridge University Press