GENE-TRAP MUTAGENESIS: PAST, PRESENT AND BEYOND

William L. Stanford, Jason B. Cohn*‡ and Sabine P. Cordes*‡*

Although at least 35,000 human genes have been sequenced and mapped, adequate expression or functional information is available for only ~15% of them. Gene-trap mutagenesis is a technique that randomly generates loss-of-function mutations and reports the expression of many mouse genes. At present, several large-scale, gene-trap screens are being carried out with various new vectors, which aim to generate a public resource of mutagenized embryonic stem (ES) cells. This resource now includes more than 8,000 mutagenized ES-cell lines, which are freely available, making it an appropriate time to evaluate the recent advances in this area of genomic technology and the technical hurdles it has yet to overcome.

MOUSE GENOMIC TECHNOLOGIES

INTRACISTERNAL A PARTICLE (IAP). Endogenous, non-infectious retroviral element that can undergo transposition and act as a mutagen.

ALLELIC SERIES
A collection of discrete mutations that affect the same gene.

In Simone de Beauvoir's existentialist novel *All Men Are Mortal*, two creatures attain immortality: a man who intends to be an enlightened ruler and a circling mouse. Clearly, this was a mutant mouse — perhaps a descendant of the Chinese waltzing mice that were first described several thousand years ago. de Beauvoir's would-be king chose the mouse as his companion for eternity for the same reason that we are pursuing mouse mutants today: the mouse provides us with an effective model of ourselves, be it for testing potions of immortality or for understanding human disease and development.

Mouse fanciers have collected spontaneous mouse mutants for millennia. Regrettably, the spontaneity of their appearance has often been matched, if not surpassed, by their impromptu disappearance. As genetics became a formalized science, the collection and analysis of spontaneous mutants became more systematic. Today some large mouse colonies, such as those at The Jackson Laboratory, have regular, in-house meetings — so-called mouse circuses — in which the newest spontaneous mutants make their debut. By nature of their identification, these mutants usually have strikingly visible phenotypes. Over time, numerous spontaneous mutants, such as the white spotted (*W*) mutation, which was first identified by its white belly spot and which affects haematopoiesis, have lent credence to the mouse as a clinically relevant model organism[1,2]. However, the low frequency at which spontaneous mutations occur ($\sim 5 \times 10^{-6}$ per locus) makes dissecting an entire genetic pathway by awaiting the appearance of spontaneous mutations unfeasible in a normal human lifespan or with the resources of an averagely sized mouse house. In addition, the molecular lesions that are responsible for spontaneous mutations can range from small base changes in coding sequences to retroviral insertions, to regulatory mutations of unknown character. For example, several spontaneous mutant alleles of the agouti coat colour locus (*a*), such as a^{16H} and a^{e}, are caused by mutations in the coding region[3], whereas others (A^{hvy}, A^{iy} and A^{vy}) are caused by the integration of a retrovirus-like element — the INTRACISTERNAL A PARTICLE (IAP) — into different positions upstream of the transcriptional start site, which deregulates agouti expression[4,5]. So, although the unpredictable molecular nature of spontaneous mutations is valuable for generating ALLELIC SERIES, these spontaneous events do little to expedite gene identification, and they do not allow mouse models of human disorders to be specifically designed. Many mouse mutagenesis strategies have evolved to address these shortcomings, and each generates mutations of a different molecular nature and at varying frequencies, as shown in TABLE 1. Together, these strategies should allow us to design ideal mouse models for studying human disease and genetic pathways in vertebrates. In this

*Programme in Development and Fetal Health, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Room 983, Toronto, Ontario, Canada M5G 1X5.
‡Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario, Canada M5S 1A8. Correspondence to W.L.S. email: stanford@mshri.on.ca

© 2001 Macmillan Magazines Ltd

Table 1 | **A comparison of mutagenesis strategies**

Mutagenesis strategy (type)	Mutagenesis frequency	Type of mutation induced	Primary advantages	Primary disadvantages
Spontaneous	5×10^{-6} per locus	Point mutations, small deletions, chromosomal rearrangements, and insertions of endogenous retrovirus-like sequences.	Visible phenotypes; only requirement is observant mouse handlers.	Only visible phenotypes detected, at very low frequency.
X-ray	$13–50 \times 10^{-5}$ per locus	Chromosomal rearrangements: ranging from simple deletions, inversions and translocations, to complex rearrangements.	Rearrangements act as a molecular landmark for cloning.	Multiple genes affected, hard to dissect individual gene function.
Chlorambucil (chemical mutagen)	127×10^{-5} per locus	Chromosomal rearrangements, especially smaller deletions (100–500 kb) and translocations.	Same as X-ray, but higher mutagenesis frequency.	Multiple genes affected, hard to dissect individual gene function.
Ethylnitrosourea (chemical mutagen)	150×10^{-5} per locus	Primarily generates point mutations, occasionally very small deletions (20–50 bp).	Single-gene mutations, amenable to high throughput.	No molecular landmarks for cloning.
Transgene/ retroviral (insertional mutagen)	5–10% of transgenic animals	Disrupts endogenous gene expression or coding sequence. Sometimes causes chromosomal rearrangements.	Provides a molecular landmark for cloning.	Labour-intensive, not applicable to high-throughput approaches.
Gene targeting (insertional mutagen)	Almost 100% of transgenic animals*	Generates insertions or deletions, as designed.	Can design type of mutation as required.	Requires knowledge of gene and its structure, labour-intensive, unpredictable phenotypes.
Trapping (insertional mutagen)	Almost 100% of transgenic animals*	Disrupts endogenous coding sequence.	Forward-genetic strategy, easy to clone mutated gene, reports endogenous gene-expression pattern.	Unpredictable phenotypes.

* Requires pre-screening of embryonic stem cells *in vitro*.

review, we discuss the development of gene-trap mutagenesis and place it in context with other mouse functional genomic techniques. We also summarize the field's progress and impediments, and point to some potential directions that gene trapping might take after SATURATION MUTAGENESIS has been achieved.

Mutagenesis strategies

On considering the pitfalls of spontaneous mutations, mouse geneticists sought to identify high-efficiency mutagenesis strategies that could produce mutations of a relatively defined molecular nature. Towards this end, the first X-ray mutagenesis experiments on the mouse were carried out in the 1930s, and gained momentum as larger centres (such as Oakridge, Tennessee, USA, and Harwell, Oxford, UK) began to study the effects of radiation fall-out on genetic stability[6]. The frequency of X-ray-induced mutations is 20–100 times greater than that of spontaneously occurring mutations. Furthermore, X-ray mutagenesis causes chromosomal rearrangements, which can provide a molecular landmark for identifying the affected gene(s). However, several genes are often affected by these chromosomal rearrangements and, as a result, their type and complexity is difficult to control. Similar multigene deletions and chromosomal translocations can be generated at a greater frequency by mutagenizing mice with the chemical chlorambucil[7]. Although these mutagenesis strategies can be useful for generating deficiencies and translocations for genetic screens and mapping studies, they generally do not lead to rapid single-gene identification and so are not used in high-throughput approaches.

Chemical mutagenesis with ethylnitrosourea (ENU), which primarily introduces point mutations into spermatogonial stem cells[8], is a preferred approach to the above strategies as ENU-induced mutations usually affect only single genes. However, ENU mutagenesis provides no molecular landmarks with which to recover mutated genes. Because, in addition to the advantages listed in TABLE 1, ENU is easy to administer, and because ENU-treated males can be used to generate mutant progeny for many months, this mutagenesis strategy has gained considerable popularity[9].

In 1976, the introduction of exogenous retroviral DNA into the mouse germ line was first reported[10], and INSERTIONAL MUTAGENESIS began to be hotly pursued in the mouse. But because retroviral infection usually occurred after the one- or two-cell stage of embryonic development, multiplicity of infection was extremely variable and could lead to the generation of chimeric animals. However, a mutation could be recovered by additional breeding if an insertion was transmitted through the germ line[11,12]. Despite the problems associated with this approach, the insight that retroviral insertion could alter endogenous genes and their expression, leading to tumorigenesis and leukaemia[10] in mutagenized mice, spurred the cloning of retroviral insertion sites to recover the affected, tumorigenic genes, and validated the potential usefulness of insertional mutagenesis.

In 1981, microinjecting DNA into fertilized oocytes yielded the first transgenic mice[13–16]. Approximately 5–10% of transgenes that are introduced into a fertilized oocyte by PRONUCLEAR INJECTION insert into and affect the function of an endogenous gene. In addition, tandem

SATURATION MUTAGENESIS
This occurs when a mutagenesis screen recovers at least one mutation in every gene.

INSERTIONAL MUTAGENESIS
A strategy that uses the insertion of DNA to mutagenize genes at the insertion site. The inserted sequence acts as a tag from which to clone the mutated gene.

PRONUCLEAR INJECTION
DNA injected into a one-cell mouse embryo before a single nucleus forms, when a male pronucleus and a female pronucleus are still present. DNA is usually injected into the more-visible male pronucleus to generate a transgenic embryo.

© 2001 Macmillan Magazines Ltd

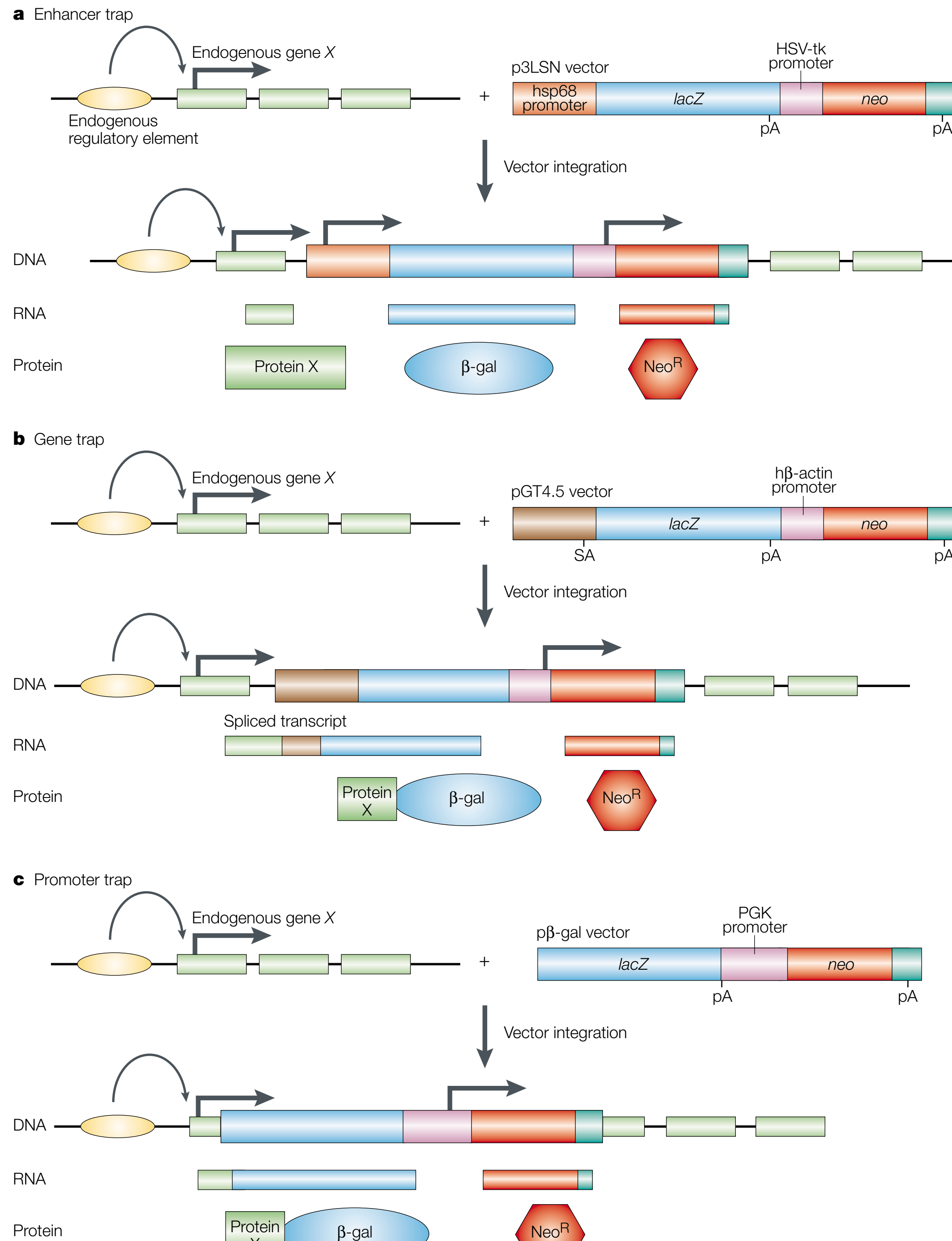

Figure 1 | **The basic trap vectors.** Enhancer-, gene- and promoter-trap vectors, which all contain a *lacZ* reporter gene and a NEOMYCIN RESISTANCE GENE (*neo*) that is driven by an autonomous promoter, are shown trapping an endogenous gene '*X*'. Integration of the trap vectors into the embryonic stem (ES)-cell genome will lead to neomycin selection whether the insertion has occurred intergenically or intragenically. **a** | The p3LSN enhancer-trap vector[25] contains a truncated heat-shock inducible minimum (hsp68) promoter immediately upstream of *lacZ*. Insertion of the enhancer-trap vector close to the enhancer of gene *X* will lead to the transcription and translation of the *lacZ* reporter when the enhancer of gene *X* is activated. This vector usually generates hypomorphic rather than null mutations. **b** | The pGT4.5 gene-trap vector contains a splice acceptor (SA) site immediately upstream of a promoterless *lacZ* gene. Its integration in an intron leads to a fusion transcript being generated from the upstream exon of gene *X* and *lacZ* upon transcriptional activation of gene *X*. **c** | The pβ-gal promoter-trap vector needs to be inserted into the coding sequence of gene *X* to activate transcription of *lacZ*. On activation of gene *X*, a fusion transcript and protein between the upstream gene *X* sequence and *lacZ* will be generated. (Adapted from REF. 103, with permission.) (β-gal, β-galactosidase; β-geo, β-galactosidase–Neo^R fusion; HSV-tk, herpes simplex virus thymidine kinase; hβ-actin, human β-actin; pA, polyadenylation; PGK, phosphoglycerate kinase 1.)

NEOMYCIN RESISTANCE GENE
The bacterial neomycin phosphotransferase gene provides resistance to the amino-glycoside analogue, G418.

© 2001 Macmillan Magazines Ltd

Box 1 | **Electroporation versus retroviral infection**

Trapping vectors can be introduced into the genome by either electroporation or retroviral infection. Whereas retroviruses have a propensity for inserting into the 5′ portion of a gene, including the 5′ untranslated region and first intron[89–91], plasmid-based vectors introduced by electroporation most often show random genomic integration[92]. Vector insertion immediately downstream of the initiation codon generally produces null mutations, which indicates that using retroviruses to introduce gene-trap vectors might give a higher percentage of null mutations, whereas electroporation might be better for generating allelic series. The main advantage of retroviral infection is that it ensures the integration of a single copy of the entire vector, although electroporation strategies can be optimized such that multiple insertions occur in less than 20% of cell lines. Tandem insertions into the same locus are problematic because they can cause ectopic reporter expression and aberrant splicing, and can interfere with the cloning of the trapped fusion transcript. Furthermore, electroporated plasmid DNA is often digested by exonuclease, making the cloning of insertion sites by INVERSE PCR problematic. By contrast, proviral DNA always retains the LONG-TERMINAL-REPEAT (LTR) sequences, which allows the high-throughput cloning of retroviral insertion sites. Retroviral vectors have three disadvantages: their packaging size is limiting, and they can induce retroviral-mediated gene silencing and ectopic reporter gene expression. However, the last two disadvantages have not been seen in vectors that lack LTR enhancer sequences or are in reverse orientation to the reporter gene. Direct comparisons of trapping by retroviral infection and electroporation have been made by the German Gene Trap Consortium, which found that each strategy has some biases for specific genes and chromosomes (W. Wurst, personal communication). Therefore, achieving saturation mutagenesis will probably require the use of both strategies.

INVERSE PCR
This technique allows the DNA that flanks a region of known sequence to be amplified. Cleavage products from restriction digests are circularized and then amplified by PCR using primers to the known sequence.

LONG TERMINAL REPEAT
(LTR). A sequence repeated at both ends of a retroviral DNA that is required for retroviral insertion into its target genomic DNA.

arrays of transgenes can also generate chromosomal rearrangements and deletions. Although generating transgenic animals by pronuclear injection has provided a means to identify and study many genes, the labour-intensive and technical aspects of this approach, as well as the difficulties of isolating genes affected by rearrangements or deletions, have precluded it from being developed for high-throughput screens.

In stark contrast to the hit-and-miss nature of these strategies is that of targeted mutagenesis by homologous recombination in embryonic stem (ES) cells[17,18]. Although the effort in generating these mutations is high, they usually affect a single gene and are of a molecularly well-defined nature. The main drawback of targeted mutagenesis is the inability to reliably predict the exact biological process that will be affected. Sometimes, genetic redundancy causes diminished phenotypes, which can hinder the functional analysis of the mutation. At other times, the targeted gene is required for early embryonic viability. The resulting failure of a mutant embryo to survive prevents the function of the gene in later developmental processes from being studied. In addition, the molecular lesion generated does not always have the predicted, and/or an easily interpretable, effect. The advent of conditional recombinase systems, which allow gene expression to be abrogated in a temporally and spatially controlled manner, has helped to overcome this limitation[19–21]. (See the accompanying review by Mark Lewandoski on p743 of this issue for more on this technique.) Combining recombinase systems with traditional homologous-recombination strategies has allowed chromosome-specific deletions and rearrangements to be made (see the review by Yuejin Yu and Allan Bradley on p780 of this issue), which can be used to screen for recessive mutations in whole-animal mutagenesis screens[9] or in cell culture.

Gene targeting has led to unprecedented insights into gene function. However, the null mutations that are generally created by investigators using homologous-recombination strategies often do not resemble the types of molecular lesions found in disease. So, despite the success of gene targeting, the interest in random mutagenesis screens has intensified. Each functional genomics approach has its strengths and weaknesses, and only by taking advantage of each one will the functions of the mammalian genome be understood.

However, one mutagenesis system — gene trapping — takes the middle path between random and molecularly defined mutations. Gene-trap vectors have evolved from enhancer-trap vectors, a molecular tool used to identify and characterize mammalian enhancer sequences from cell lines[22]. This technique was quickly modified in both *Drosophila* and the mouse to generate vectors that usurp the transcriptional regulatory apparatus of endogenous genes to report the endogenous expression of a gene that flanks the vector insertion site[23–25]. The introduction of enhancer-trap vectors, and of the related promoter- and gene-trap vectors, into the mouse germ line through ES cells has permitted random mutations to be generated and characterized *in vitro* before germ-line transmission[26]. So, from the outset, gene trapping has promised to be an efficient system for simultaneously characterizing gene function, sequence and expression. It is this approach, and its limitations, that we discuss further in this review.

Trapping vectors

Three types of trap vector — the enhancer-, promoter- and gene-trap vector (FIG. 1), which can be introduced into the genome by electroporation or by retroviral infection (BOX 1) — were simultaneously being developed for ES-cell mutagenesis while targeted mutagenesis mediated by homologous recombination was still in its infancy. The enhancer-trap vector contains a minimal promoter that requires the vector to insert near to a *cis*-acting enhancer element to produce expression of a *lacZ* reporter gene (FIG. 1a). Enhancer traps were first tested by introduction into fertilized oocytes by pronuclear injection[24,25]. More than 20% of the resulting transgenic lines that were generated showed restricted patterns of reporter gene expression during embryogenesis, with at least 5% of insertions being mutagenic, which indicates that this strategy can be used to trap various loci. When enhancer trapping was carried out in ES cells[26,27], this approach, on the basis of *lacZ* expression, was found to be more efficient at trapping genes than the use of pronuclear injection to introduce the enhancer-trap vectors (TABLE 2). Cloning insertion sites from ES-cell lines that showed reporter expression *in vivo* indicated that insertions occurred adjacent to coding sequences. The mutagenicity rate from enhancer trapping has not been reported, but the nature of the insertions indicates that loss-of-function mutations from enhancer traps might

© 2001 Macmillan Magazines Ltd

Table 2 | **A comparison of trapping vectors**

Vector (reporter)	Selection	Frequency*	Advantages	Disadvantages	Refs
Enhancer trap (β-gal)	Autonomous promoter drives selection.	0.2	High trapping frequency, including genes not expressed in undifferentiated ES cells.	Not expected to be very mutagenic, potentially harder to clone genes.	27
Promoter trap (β-gal)	Autonomous promoter drives selection.	10^{-3}	High mutagenicity rate.	Very low trapping efficiency.	29
Promoter trap (β-geo)	β-geo	NR‡	High mutagenicity rate, high trapping frequency.	Requires gene to be expressed in undifferentiated ES cells.	
Gene trap (β-gal)	Autonomous promoter drives selection.	0.04	Can trap genes not expressed in undifferentiated ES cells.	Low trapping efficiency.	34
Gene trap (β-geo)	β-geo	0.34	High trapping frequency.	Requires gene to be expressed in undifferentiated ES cells.	34
PolyA gene trap (β-gal)	Endogenous polyA drives selection.	0.13	Can trap genes not expressed in undifferentiated ES cells.	Mutagenicity still has not been adequately tested.	W.L.S.§

*Frequency of trapped events per number of clones selected for vector integration, as reported by *lacZ* expression in undifferentiated embryonic stem (ES) cells. ‡Not reported, to our knowledge, but should be similar to the frequency of trapped events for the gene-trap β-geo vector. §W.L.S., unpublished observations. (β-gal, β-galactosidase; β-geo, β-galactosidase–Neo^R fusion.)

be rare. Therefore, enhancer-trap vectors have not been widely exploited in the mouse.

In essence, promoter-trap vectors consist of a promoterless reporter gene and selectable marker (FIG. 1b), which are often the same[28–30]. Reporter expression and mutagenesis occur when the vector inserts into an exon to generate a fusion transcript that comprises upstream endogenous exonic sequence and the reporter gene. Because transcription of the reporter requires that the vector inserts into an exon, the mutagenicity rate of promoter-trap vectors should be very high, as indicated by there having been very few reports of hypomorphic mutations from promoter traps. In addition, because the insertion site is in transcribed DNA, cloning the insertion site will identify the disrupted gene. However, the frequency with which promoter-trap vectors insert into exons is exceedingly low[29,31], at least 200-fold lower than that of the enhancer trap (TABLE 2). Therefore, promoter-trap vectors generally contain a selectable marker, such as the neomycin resistance gene (*neo*) or the β-GALACTOSIDASE–Neo^R (β-GEO) fusion marker, as a reporter, so that only ES-cell clones that contain vector insertions can be selected. This approach, however, means that only insertions into those genes that are transcriptionally active in ES cells will be selected.

The gene-trap vector contains a SPLICE ACCEPTOR SITE immediately upstream of a promoterless reporter[26] (FIG. 1c). On transcriptional activation of the endogenous *cis*-acting promoter and enhancer elements of the trapped gene, a fusion transcript is generated from the upstream coding sequence and the reporter gene, simultaneously mutating the trapped gene and reporting its expression pattern. The fusion transcript also serves as the template for PCR-based cloning by a technique called 5′RACE[32,33]. The relative size of introns to exons increases the efficiency of gene traps per vector insertion by at least 50-fold in comparison to promoter traps[29,34] (TABLE 2). The efficiency of gene trapping is sufficiently high for some groups, including ours, to have developed strategies to trap genes that are not expressed in undifferentiated ES cells (as discussed below). The main disadvantage to using gene-trap vectors is that, because the insertion occurs in an intron, alternative splicing can sometimes take place, leading to lower levels of wild-type transcripts and often resulting in hypomorphic alleles[35].

Box 2 | **A public resource of trap insertions**

- **University of Manitoba Institute of Cell Biology, Winnipeg, Canada (Director: Geoff Hicks)**

  This project is, at present, depositing 300 clones per month into a searchable database at http://www.escells.ca, and is now being automated to increase clone production and the implementation of new vectors, with the aim of generating an allelic series of promoter-trap mutations for genes that are expressed in undifferentiated embryonic stem (ES) cells.

- **The Gene Trap Project of the German Human Genome Project (Director: Wolfgang Wurst)**

  The German Human Genome Project is carrying out both a phenotypic-driven ethylnitrosourea (ENU)[93] and a genotype-driven gene-trap mutagenesis programme[94]. They are delivering four different β-geo gene-trap and promoter-trap vectors into ES cells by electroporation and retroviral infection. As of May 2001, 6,000 sequenced clones had yielded 4,587 trapped genes. Clones are freely available and can be searched by gene name or sequence at http://tikus.gsf.de/.

- **The BayGenomics Gene Trap Project, California, USA (Director: William C. Skarnes)**

  From the combined screens by the Skarnes and Tessier-Lavigne laboratories, 774 properly spliced, sequence-identified insertions have been obtained that represent 341 genes. Insertion sequence, expression and phenotypic data can be searched, and clones can be requested, at http://baygenomics.ucsf.edu/. A second genotypic screen has been initiated to generate an insertion resource for all classes of gene.

- **The CMHD Gene Trap Project, Toronto, Canada (Director: William L. Stanford)**

  The Centre for Modeling Human Disease (CMHD) is carrying out ENU-based phenotypic screens, and gene-trap-based expression and genotypic screens, to generate new models for studying human disease and mouse development. The complementary approaches are designed to isolate different types of mutations in overlapping developmental and functional pathways. Using a polyA-trap vector with recombination sites for post-insertional manipulations, insertions are generated and screened using *in vitro* differentiation and induction assays. Expression profiles have been generated for more than 4,000 clones, and sequence tags for all insertions that show restricted expression patterns (~20%) are now being generated. All clones will be publicly available and can be requested at http://www.cmhd.ca.

© 2001 Macmillan Magazines Ltd

β-GALACTOSIDASE
This enzyme, encoded by *lacZ*, is a commonly used reporter, as it can act on various substrates to produce colour or fluorescent reactions to indicate the expression pattern of a gene.

β-GEO
A fusion protein encoded by the *lacZ* gene, which encodes the marker β-galactosidase, and the *neo* gene, which confers neomycin resistance.

SPLICE ACCEPTOR SITE
Sequences from exon–intron boundaries that mediate mRNA splicing.

5′RACE
(5′ rapid amplification of cDNA ends). RACE is a PCR-based method for amplifying unknown cDNA sequences by using primers that correspond to known sequence.

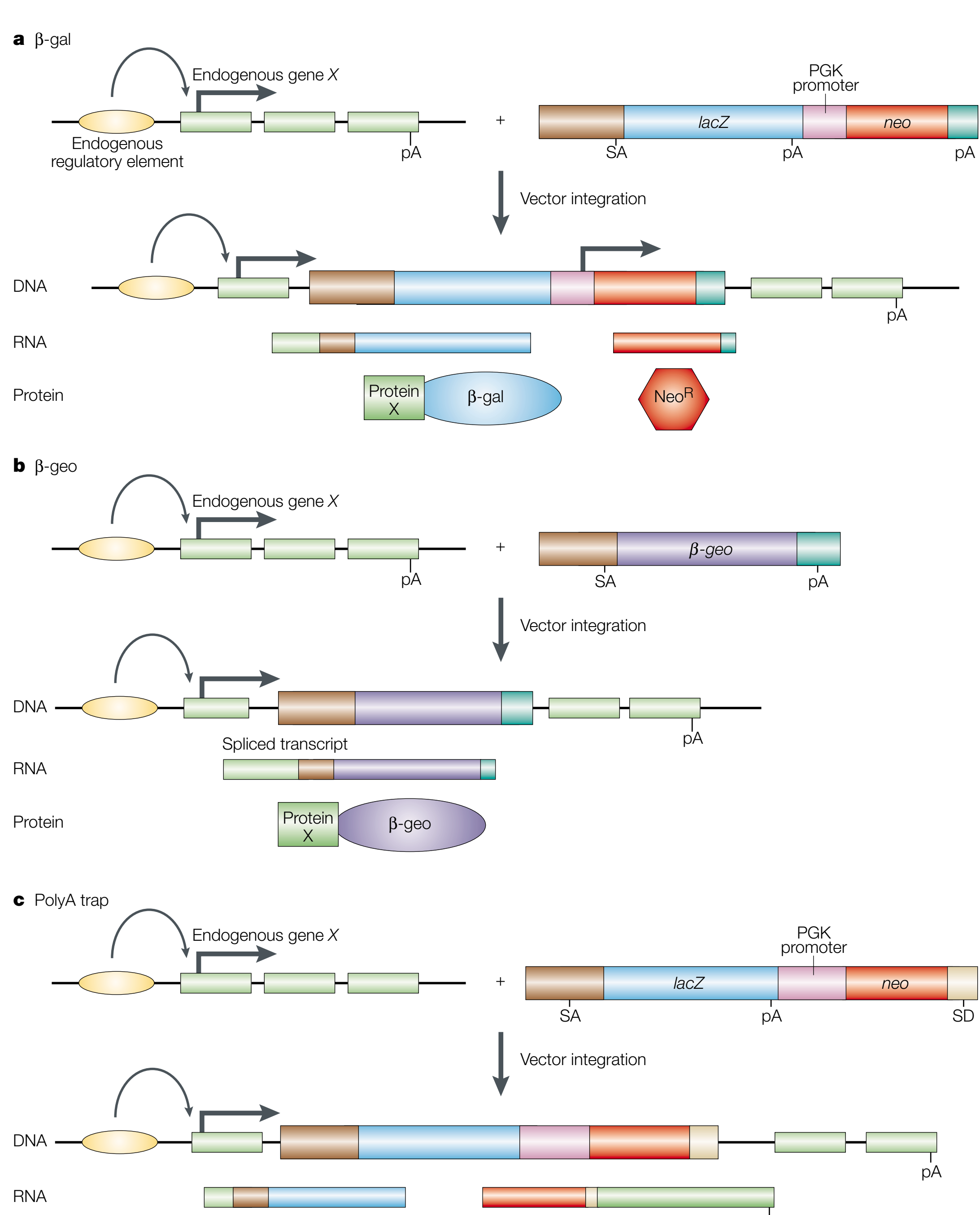

Figure 2 | **A comparison of β-gal, β-geo and polyA gene-trap vectors.** Each vector is shown inserting into intron 1 of gene *X*. **a** | The β-gal vector contains a splice acceptor (SA) site immediately upstream of the *lacZ* reporter gene followed by a neomycin resistance (*neo*) selectable marker that is driven by an autologous promoter (PGK). All insertions, regardless of whether the insertion occurs in an intron (as shown) or in intergenic regions, lead to neomycin resistance and selection. If the insertion occurs in an intron, a fusion transcript is generated between the *lacZ* reporter and the upstream exon of gene *X* on transcriptional activation of the locus. **b** | The insertion of the β-geo vector generates a fusion transcript and protein from the *β-geo* reporter gene and the upstream exon of gene *X*, providing that gene *X* is transcriptionally active in undifferentiated embryonic stem (ES) cells. **c** | In this polyA-trap insertion, a fusion transcript and protein are generated from *lacZ* and the first exon of gene *X*; however, the polyA-trap also leads to a fusion transcript between *neo* and the downstream exons of gene *X*, providing a polyA site with which to stabilize the *neo* transcript. Several stop codons that follow *neo* prevent the 3′ exons of gene *X* from being translated. The β-gal and polyA-trap vectors will trap genes that are not transcriptionally active when trapping is carried out. The advantage of the polyA-trap vector over the β-gal vector is that neomycin selection should only occur when the polyA-trap vector inserts upstream of a splice acceptor and a polyA site of an endogenous gene, thereby eliminating selection of intergenic insertions. (β-gal, β-galactosidase; β-geo, β-galactosidase–Neo^{R} fusion; pA, polyadenylation; PGK, phosphoglycerate kinase 1; SD, splice donor.)

© 2001 Macmillan Magazines Ltd

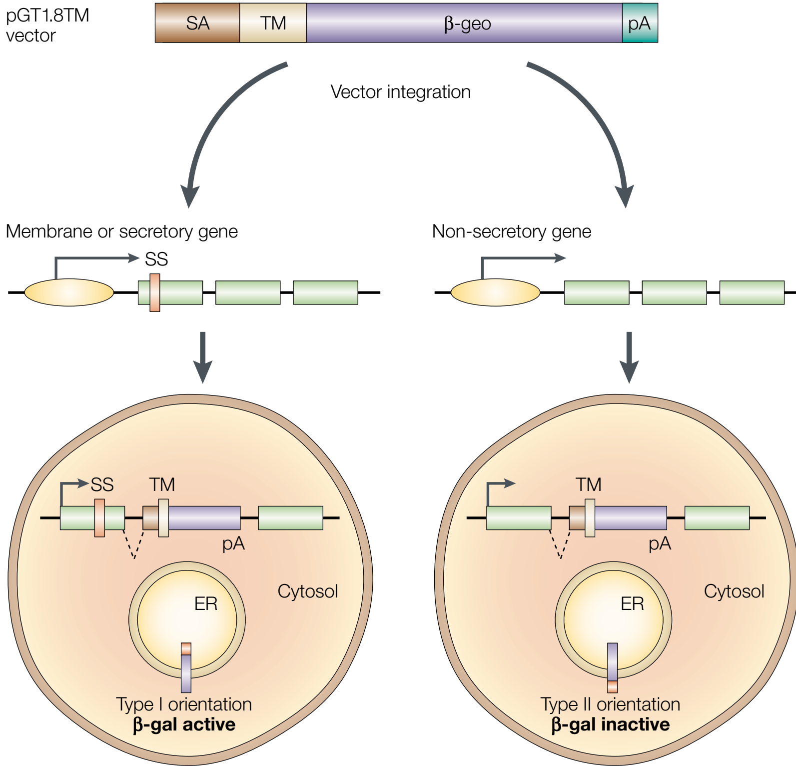

Figure 3 | **The secretory-trap vector.** The secretory-trap vector uses protein sorting and the fact that β-galactosidase (β-gal) activity is abolished in the endoplasmic reticulum (ER) specifically to trap genes that encode secreted and transmembrane proteins that are expressed in embryonic stem (ES) cells[41]. The pGT1.8TM secretory-trap vector contains a transmembrane (TM) domain immediately downstream of a splice acceptor (SA) site, followed by the *β-geo* reporter with its own polyA site. The TM domain of the secretory-trap vector sequesters gene-trap fusion proteins that do not have a signal sequence into the ER lumen, thereby extinguishing β-gal activity. When a fusion protein contains a secretory signal (SS) sequence, it is translocated into the cytosol where β-gal activity can be assayed. So, the secretory-gene-trap vector enriches for insertions into genes that encode secreted or transmembrane proteins by using a modification of blue–white selection. (pA, polyadenylation.)

New vector designs

Gene trapping is an exceptional tool for gene discovery because genes are trapped regardless of their transcriptional activity. However, as has been found in homologous-recombination experiments, there are 'hot' and 'cold' genomic spots for gene-trap vector insertions. But, computer modelling, albeit with a limited data set, has indicated that on the basis of the percentage of unique clones that have been trapped, and given enough rounds of mutagenesis and enough different types of vector to negate biases, virtually the entire ES-cell genome could be saturated with insertions[36]. This modelling exercise needs to be repeated using the new data sets from the large genotype-driven screens; however, data from the German Gene Trap Consortium, which has used four different vectors to isolate ~6,000 sequences that represent ~4,500 genes (BOX 2) (W. Wurst, personal communication), are consistent with this model. The question that remains is whether current trap vectors can report insertions into all classes of genes and their products. To tackle this, a new generation of vectors has been developed (FIG. 2) to increase trapping efficiency[37–40], to enhance the recovery of trapped sequence[37–41], to trap specific classes of protein[42], and/or to allow post-insertional modification of the trapped loci[43,44].

The primary reporter used in gene- and promoter-trap vectors is β-geo (FIG. 2). Because the selection of trapped clones requires gene expression, insertions into only those genes that are expressed in undifferentiated ES cells are selected for, eliminating the selection of intergenic insertions but also eliminating the selection of insertions into genes that are not expressed in undifferentiated ES cells[29]. As with other types of stem cell[45], ES cells transcribe an abnormally high number of genes; however, to achieve genome-wide mutagenesis using this approach requires that researchers should be able to select independently for gene-trap insertions of gene expression in undifferentiated ES cells. The first generation of gene-trap vectors, which generate β-galactosidase (β-gal) fusion proteins and do not require gene expression for antibiotic selection, generated a high percentage of intergenic insertions[26]. So, to trap genes that are not expressed in undifferentiated cells, some groups, including ours, have developed polyadenylation (polyA)-trap vectors in which a constitutive promoter drives the expression of a selectable marker (such as *neo*) that lacks a polyA signal but contains a splice donor signal[37–40] (FIG. 2). A spliced polyA signal from an endogenous gene is therefore required to generate stable *neo* mRNA and, in turn, neomycin-resistant (Neo^R) clones. So, only insertions in genes should generate Neo^R clones, and background intergenic insertions should be lost. Several termination codons following the selectable marker prevent the translation of the 3′ trapped exons. Because the stability of many transcripts is linked to their translation and their endogenous 3′ untranslated regions, polyA-trap vectors are expected to work best when inserted near the 3′ end of genes. So, this strategy might bias selection towards 3′ insertion events, resulting in fewer null mutations. Unfortunately, because these vectors are new, there is not enough data to evaluate this potential bias. In our own limited analysis of polyA traps generated in our laboratory, we have not noticed any biases (M. Yu and W.L.S., unpublished data). However, further modifications of this strategy might be required. PolyA-trap vectors also circumvent a second technical hurdle of gene-trap technology — that of obtaining enough sequence from 5′RACE to identify the gene into which a gene-trap vector has inserted at its 5′ end. 3′RACE can also be used to identify the untranslated 3′ coding sequence of the trapped gene. Because these are new vectors, only one polyA gene-trap strain has so far been reported[46] (the insertion created a null mutant), so the mutagenicity rate of this type of vector is, at present, unknown.

New vectors have also been designed to trap specific classes of protein. For example, Bill Skarnes and Rosa Beddington have used protein sorting and the fact that β-gal activity is abolished in the endoplasmic reticulum to design a vector that specifically traps genes that encode secreted and transmembrane proteins, and are expressed in ES cells[42] (FIG. 3).

© 2001 Macmillan Magazines Ltd

DIPHTHERIA TOXIN
A toxin secreted by the bacteria *Corynebacterium diphtheriae*. A cell lineage can be ablated using diphtheria toxin, by driving its cDNA from a lineage-specific promoter.

SV40 T ANTIGEN
A viral oncogene from the polyoma virus that can bind the tumour suppressors p53 and retinoblastoma, which blocks their functions and leads to the immortalization of the cell.

To increase the versatility of trapping, several groups have modified vectors to include recombination sites, which allows recombinase-mediated, post-insertional modifications of the gene-trap locus[43,44]. This 'knock-in' strategy permits additional modifications to be made to a trapped locus, such as co-opting the promoter elements of the trapped gene to drive the expression of a knocked-in transgene for use in rescue or cell-labelling experiments. In addition, this approach can also be used to generate an allelic series by recombining in different mutant cDNAs of the trapped gene. Furthermore, DIPHTHERIA TOXIN or SV40 T ANTIGEN can be inserted into a locus to target the lineages that express the trapped gene for ablation or immortalization[47,48], respectively. So, by combining homologous recombination and gene-trapping strategies, gene trapping can be used for more than just expression, sequence and simple functional analysis.

Finally, promoter traps must be used to trap single-exon genes, such as odorant receptors, which cannot be trapped by gene-trap vectors, which require the splicing of upstream exons. Second-generation promoter-trap vectors have recently been developed that allow the high-throughput sequence identification of the trapped locus by the use of a vector that employs plasmid rescue to clone genomic insertion sites[41] (FIG. 4). This strategy is much more efficient than the RNA-isolation step and the multiple rounds of reverse transcription (RT)-PCR that are required for RACE.

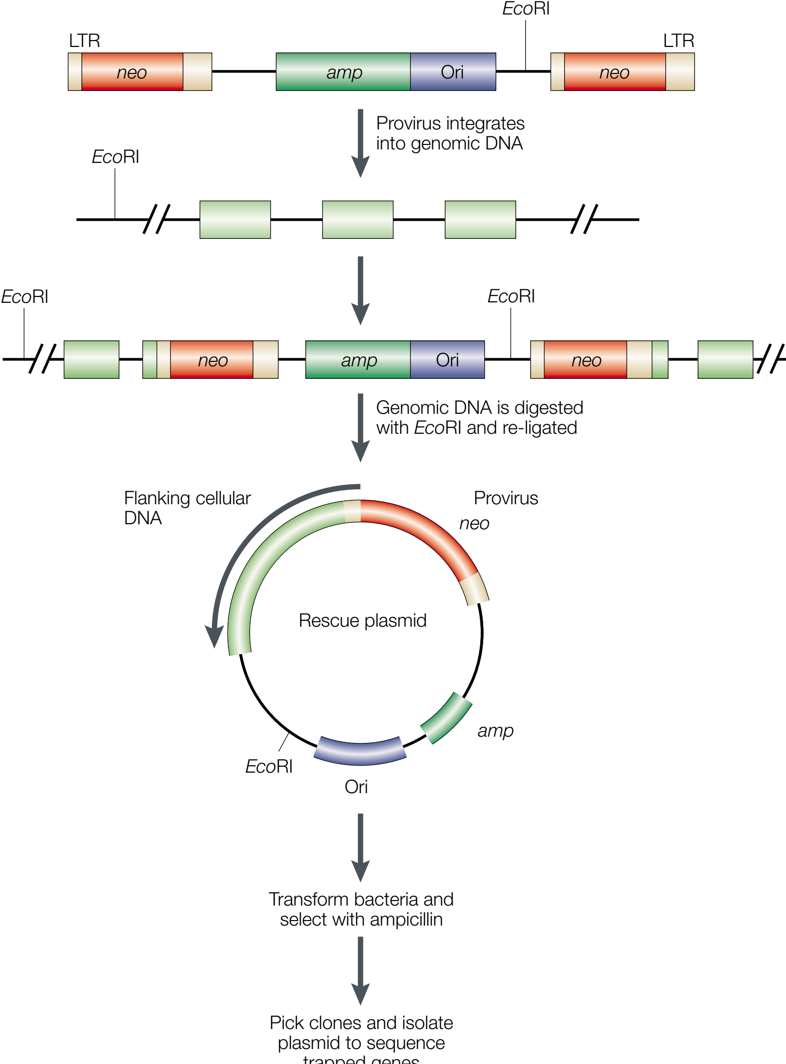

Figure 4 | **The plasmid rescue vector.** The U3NeoSV1 promoter-trap provirus contains the ampicillin resistance (*amp*) gene and a plasmid origin of replication (Ori) flanked by the neomycin resistance (*neo*) gene in each long terminal repeat (LTR). Selecting for Neo resistant (Neo^R) clones identifies those cells in which an endogenous gene has been disrupted as a result of the proviral insertion. Genomic DNA is isolated from mutagenized clones, digested with *Eco*RI, and then ligated and used to transform bacteria. Only bacteria that contain the vector will grow in the presence of ampicillin. Plasmid DNA is then prepared and sequenced to identify the gene mutated by the insertion of the promoter-trap vector.

Mutagenicity

Several types of trapping screen have been carried out to test the limitations of the technology, to develop improvements and to generate interesting mutations to study. The three types of data that trap insertions can provide — gene function, expression and sequence — have been used as the basis of mutagenesis screens. The most important test for any mutagenesis strategy is the mutagenicity rate. All insertional mutagens, including those introduced by gene targeting, generate silent or hypomorphic mutations at various rates[43]; however, phenotype-driven approaches, such as ENU, do not detect silent mutations, whereas GENOTYPIC or EXPRESSION SCREENS will detect non-mutagenic insertions.

Homologous-recombination experiments often generate several correctly targeted cell lines, thereby increasing the likelihood of transmitting a mutation through the germ line, whereas trap insertions are one-time events. As a result, not all gene-trapped ES-cell lines will generate mouse lines. In an early screen, Glenn Friedrich and Philippe Soriano attempted to increase the efficiency of germ-line transmission by pooling together ES cells that carried different proviral insertions for blastocyst injection[29]. Thirty-eight out of 45 chimaeras tested showed contribution of the ES cells to the germ line, transmitting between one and three independent proviral genomes to their offspring. Twenty-four strains were derived from outbreeding chimaeras. Heterozygous intercrosses showed that nine out of the 24 strains carried embryonic-lethal mutations (38%). Although the pooling strategy increased germ-line-transmission efficiencies, it proved difficult to clone the mutated genes without the original ES-cell lines. Out of the 75 strains generated by the Soriano lab, 11 have been cloned and functionally characterized in depth. Each of the characterized ROSA strains has been a null mutation.

Recently, the Bill Skarnes and Marc Tessier-Lavigne labs described the phenotypic analysis of 60 mouse lines with secretory gene-trap insertions generated by electroporation[49]. One-third of the mouse strains generated from these ES-cell lines had recessive lethal phenotypes, including nine that were embryonic, ten that were neonatal and one that was postnatal. Five additional strains showed visible adult phenotypes. A detailed analysis of the other 35 strains has not yet been done. Gene-targeting alleles had previously been reported for

© 2001 Macmillan Magazines Ltd

11 of the 25 gene traps that had observable phenotypes. Ten (91%) of these strains exactly phenocopied the gene-targeted mutations. The remaining strain, an insertion in the 3′ coding sequence of the exostoses (multiple) 1 (*Ext1*) gene, caused a less-severe phenotype than the gene-targeted allele but still led to embryonic lethality, indicating that the insertion created a partial loss-of-function fusion protein.

In addition to the two large-scale phenotypic screens described above, nearly 100 additional gene-trap insertions have been described in the literature. Sixty per cent of the insertions show obvious phenotypes, and 40% are recessive lethal mutations. So, in total, the frequency of recessive lethal mutations and obvious phenotypes generated by gene-trap mutagenesis resembles that generated by gene-targeted mutagenesis. In addition, nearly all of the insertions that are predicted to result in a null mutation phenocopy the targeted null alleles of their respective genes, indicating that insertions result in null alleles. In the future, therefore, researchers might be able to check libraries of trapped ES-cell lines for an insertion into their gene of interest before generating targeted mutations themselves.

Expression screens

Unlike phenotypic screens of ENU mutagenized mice, in which mutagenized animals carry high mutation loads to allow genetic defects to be identified in relatively small numbers of pedigrees, phenotypic screens of gene-trap strains are generally prohibitively inefficient. Therefore, many groups have used reporter expression as a way to assess whether a specific developmental pathway has been disrupted. For example, the laboratories of Janet Rossant and Alexandra Joyner generated almost 300 AGGREGATION CHIMAERAS using ES-cell lines that contain gene-trap insertions into genes expressed by ES cells; they then analysed *lacZ* expression in mouse embryos generated from these clones[50]. Two-thirds of the chimeric embryos expressed *lacZ*; most showed widespread *lacZ* expression, especially in the central nervous system (CNS). In fact, only 15% of the embryos generated from these clones did not show *lacZ* expression in the CNS; in this sample of insertions, *lacZ* expression was found in all cell lineages. Although *lacZ* was expressed in a temporally and spatially restricted manner in many of these embryos, only 4% showed highly restricted *lacZ* expression throughout embryogenesis. These data indicate that phenotypic screens for genes required for CNS development or early embryogenesis can be carried out by trapping genes that are expressed in undifferentiated ES cells. The laboratory of Peter Gruss has used two different β-geo vectors in a similar approach. Seventy-five per cent of the 64 mouse lines they analysed showed embryonic expression of *lacZ*, and nearly 50% of the lines expressed it predominantly, or in an upregulated manner, in the developing nervous system. Half of the 16 lines that were bred to homozygosity were found to carry recessive embryonic-lethal mutations[51].

Recently, the Tessier-Lavigne and Skarnes laboratories used a variation of the *in vivo* EXPRESSION-TRAPPING SCREEN to identify and mutate the receptors and ligands that control axon guidance[52]. The secretory-trap vector was modified to generate a BICISTRONIC message that encoded the trapped β-geo product and human placental alkaline phosphatase (PLAP), which stains axonal projections. So, β-galactosidase and PLAP stain both the cell body and axon, respectively, which provides markers to study axon guidance and growth. In a current screen, sequence-verified secretory-trap ES clones have been used to generate mice to identify neuronally expressed genes, to map their axonal expression patterns and to determine their function in axon guidance.

The apparently high number of neural genes accessible to trapping in undifferentiated ES cells probably reflects the diversity of neuronal cells that require a panoply of gene products for their differentiation, as well as the fact that by default ES cells will differentiate into neural stem cells[53]. However, more sophisticated screens are required to identify gene-trap insertions into genes that function in non-neural lineages and developmental pathways. In addition to developing into all cell lineages and tissues *in vivo* (except trophoblasts), ES cells can also differentiate into various cell types *in vitro* and can respond to many physiological and molecular signals[53–57]. The developmental programmes, gene-expression profiles and cell-signalling pathways of *in vitro* differentiated ES cells approximate certain aspects of early embryonic development, including early organogenesis[58–61]. Furthermore, ES cells that have differentiated *in vitro* provide a tool with which to observe and manipulate transient populations of cells that cannot be studied easily *in utero*, such as haemangioblasts[56]. So, the *in vitro* differentiation of ES cells has been used to study the effects of targeted mutations on the haematopoietic, vascular, myoblast and other early lineages[62–65]. Expression trapping and INDUCTION TRAPPING exploit this use of ES cells to identify and mutate genes that are expressed in specific cell lineages or those that respond to specific cues. These screens are amenable to small laboratories that want to isolate a few mutants for their laboratory to work on for the next few years, because they are directed screens that are much like phenotype-driven ENU screens. Expression-trap screens have been done to identify and mutate genes that are expressed in haematopoietic[34,66,67] and endothelial lineages[34], cardiomyocytes[68], chondrocytes[68] and neurons[69], both *in vitro* and *in vivo*. Induction screens have identified and mutated genes regulated *in vitro* and *in vivo* by retinoic acid[35,70–73], engrailed homeobox proteins[74], γ-irradiation[75] and Bmp2 (bone morphogenetic protein 2) (J. Wrana, personal communication).

Expression and induction trapping can be carried out very efficiently. Therefore, in our laboratory we have compared the expression patterns of hundreds of genes trapped by β-gal and β-geo gene-trap fusion vectors (FIG. 2), and determined that the β-gal vector traps a much higher percentage of genes that have restricted expression patterns both *in vitro* and *in vivo*[34,66] (W.L.S., unpublished data). This indicates that genes with restricted expression patterns are less likely to be expressed in undifferentiated ES cells and so are not accessible to β-geo vectors, which require at least

GENOTYPIC SCREEN
A genetic screen of mutants based solely on the sequence of the mutated gene.

EXPRESSION SCREEN
A genetic screen of mutants based solely on the expression pattern of the mutated gene.

ROSA
(Reverse orientation splice acceptor). In the ROSA series of gene-trap vectors, the long terminal repeat (LTR) is in the reverse orientation to the trapping vector to ensure that the LTR does not interfere with the endogenous expression pattern of the trapped gene.

AGGREGATION CHIMAERAS
A simple and inexpensive technique to generate mouse strains derived from embryonic stem (ES) cells, in which clumps of about eight ES cells are fused with eight-cell-stage embryos, developing into morphologically normal embryos that contain cells derived from both the ES cells and the donor embryo.

EXPRESSION-TRAPPING SCREEN
A gene-trap screen that identifies clones with trapped events in genes expressed in specific lineages.

BICISTRONIC
A bicistronic message allows two different proteins to be translated from the same mRNA strand, usually from a promoter and an internal ribosomal entry site.

INDUCTION TRAPPING
A gene-trap screen that identifies clones with trapped events in specific signalling pathways.

© 2001 Macmillan Magazines Ltd

Box 3 | Trapping in other species

Enhancer trapping in *Drosophila* has provided an invaluable resource of spatially and temporally regulated cell and tissue markers. Transposon-mediated enhancer traps allow flanking genomic sequence to be rapidly identified, and 10–15% of the insertions are mutagenic[95–97]. At present, the Berkeley *Drosophila* Genome Project is using enhancer traps to disrupt the estimated 3,600 genes that are required for adult viability. On the basis of current results, the project predicts that at least 85% of open reading frames can be mutated using this strategy[98]. Transposon-mediated promoter trapping in *Arabidopsis* and other plant species has also been a fruitful endeavour[99]. Current screens in plants have moved away from the traditional β-glucuronidase reporter to green fluorescent protein and luciferase reporters, which allow developmental processes to be tracked in real time[100]. Finally, the Nancy Hopkins' lab has undertaken a *tour de force* zebrafish screen, in which 250,000 embryos have been injected with a high-titre gene-trap retrovirus to establish 36,000 founders and 6,800 families. After screening ~50% of the families, ~400 recessive mutants have been identified, and at present 2–3 insertions are being sequenced each week[101,102] (N. Hopkins, personal communication).

low-level gene expression of a trapped locus in undifferentiated ES cells for neomycin selection. Our laboratory is now carrying out a large-scale screen with polyA-trap vectors (BOX 2) and has found that, based on more than 3,000 insertions, these vectors trap the same types of gene as β-gal vectors, but at a higher efficiency. In addition, induction-trap strategies have been applied to non-ES-cell lines to identify genes that are activated or repressed in specific signalling pathways, including LIPOPOLYSACCHARIDE-responsive genes in B cells[76], apoptotic factors in haematopoietic cells[77], inhibitors of oncogenesis in fibroblasts[78], TGF-β-responsive genes in lung carcinoma cells[79] and germ-cell signalling in Sertoli cells[80].

Identifying the mutated gene sequence by 5′RACE analysis is used as a secondary screen in expression- and induction-trap screens to assign priority to which mutagenized cell lines to analyse *in vivo*. Single RACE products are also indicative of productive (or correct) splicing between the vector and the upstream endogenous sequence[81]. Non-productive splicing (splicing around the vector) usually results in hypomorphic mutations that are often informative about the protein or protein-domain function, but rarely recapitulate null-mutant phenotypes.

LIPOPOLYSACCHARIDE
Component of the outer membrane of Gram-negative bacteria. LPS is commonly used to elicit B-cell proliferation and differentiation.

Genotype-driven screens

Sequence-based (genotypic) screens have been the most recent type of gene-trap screen to emerge, largely because 5′RACE has only recently become amenable to high-throughput analysis[82]. Improvements in 5′RACE tagging and direct sequencing, as well as the emergence of plasmid rescue vectors and polyA traps, which permit 3′RACE, have driven genotype-based screens. In 1997, Geoff Hicks and Earl Ruley carried out the first high-throughput genotypic screen using a modified promoter-trap vector that included a plasmid origin of replication and ampicillin resistance gene, which allowed the DNA that flanks the insertion site to be recovered by plasmid rescue[41] (FIG. 4). A frozen library of 400 trapped ES-cell lines was generated and annotated by the insertion site sequence (or 'PST' for 'promoter-proximal sequence tag'), and by gene or expressed sequence tag identity. Every important class of proteins was represented in the identified trapped genes, and this screen acted as a proof-of-principle study for a current project to generate a freely available public resource of trapped ES-cell lines (BOX 2).

Sequence-based initiatives are well suited to the biotech world and therefore, predictably, Lexicon Genetics, Inc., was founded on the basis of using polyA gene trapping. More than 100,000 trap insertions in ES cells have been deposited into Lexicon Genetics' 'OmniBank'[40] (see links box). By searching the OmniBank database, clones can be identified that harbour an insertion in a particular gene, and mice derived from the trapped cell lines can be purchased at a minimum cost of US $25,000, plus additional compensation if patents are generated from work with trapped strains.

Current and future directions

The current direction of gene-trap mutagenesis is similar to that of chemical mutagenesis: a combination of large-scale mutagenesis centres carrying out high-throughput screens to generate a worldwide mutant resource and smaller, investigator-driven focused screens. The four continuing large-scale screens that will establish frozen libraries of mutagenized ES cells, which will be freely accessible to the scientific community, are discussed in BOX 2. In addition to these continuing screens, Ken-Ichi Yamamura in Kumamoto, Japan, has initiated a genotype-driven screen based on a β-geo vector that contains recombination sites for the recombinase-mediated cassette exchange discussed above[43] (K.-I. Yamamura, personal communication). Also, Earl Ruley in Nashville, Tennessee, USA, is setting up a large-scale, polyA trap screen in which 3′RACE products will be arrayed onto microarrays. These microarrays will be made available to investigators so that they can identify gene-trap clones that lie in a given pathway of interest (E. Ruley, personal communication).

Unless many strategies and vectors are used, it will be impossible to approach saturation mutagenesis of the mouse genome using the trapping approach. For example, the German Gene Trap Consortium (BOX 2) has used four different vectors, and both electroporation and retroviral infection, to deliver the vectors. Analysis of more than 6,000 sequences obtained from these vectors has shown that each vector and mode of delivery has preferential insertion sites (W. Wurst, personal communication). Their results, however, indicate that using multiple vectors, as well as electroporation and retroviral delivery, will minimize any collective biases.

Each of the large-scale, academic trapping laboratories is allowing full access to the libraries of frozen ES-cell clones. In a recent international workshop on gene-trap mutagenesis, a consortium of insertional mutagenesis laboratories was formed in association with the International Mutant Mouse Consortium[83] to promote the accessibility of gene-trap resources by working towards common distribution policies for ES cells and mice, to establish a common web site to centralize searches, and to work together with public databases to form links to the Consortium's insertional mutation database. (A report on the meeting and the development of the Consortium is now in preparation.)

© 2001 Macmillan Magazines Ltd

Gene-trap strategies have evolved divergently and convergently, and the community as a whole is in the process of executing these screens. So, at this stage, we wonder whether we — the gene-trap mutagenesis community — have developed elaborate mousetraps but failed to trap many mice. Has gene trapping failed to live up to its immense potential? Unfortunately, at this point the answer to all of these questions to a certain extent is 'maybe'. However, the same could be said for all mutagenesis systems. Once an interesting mutation has been recovered, it might stimulate the work of a student, postdoctoral fellow or entire laboratory for years to come; therefore, not enough mutations have been characterized to truly measure the potential of gene-trap mutagenesis. Has gene targeting failed to reach its potential because targeted null mutants often have no apparent defect or die *in utero*, therefore preventing a full analysis of the function of a gene? The answer is no — gene targeting evolved to incorporate conditional mutagenesis, as reviewed in this issue, and gene trapping should continue to evolve as well. The fact that there are different vectors and gene-trap strategies has helped to resolve the problems that have so far been encountered in this developing technology. However, not all gene-trap insertions are mutagenic; if many trapped alleles are freely available, then the likelihood increases that one of them will generate a null mutation, which can be predicted from the insertion site. It is true that much of the work in gene-trap mutagenesis has focused on technique and tool development. However, the large-scale screens now underway indicate that gene-trap mutagenesis might provide considerable functional data to aid with annotating the mammalian genome. That said, the gene-trap mutagenesis consortium will fall short of its goals if its combined efforts are unable to at least approach saturation mutagenesis of the mouse genome. To meet this goal, more centres must adopt or develop new vectors and strategies to isolate mutations in genes not expressed in undifferentiated ES cells.

Despite these concerns, an increasing number of success stories are emerging that show that gene trapping is a fruitful approach. For example, novel genes identified by trapping can be immediately analysed phenotypically, which can often point to shared molecular pathways. For example, homozygous embryos that contain an insertion in *Lrp6* — a novel low-density lipoprotein receptor family gene — had a truncated body axis, loss of distal limb structures, microphthalmia and cranial facial defects that resembled phenotypes caused by mutations in the Wnt signalling pathway[84]. Biochemical analysis confirmed that Lrp6 signals in the Wnt pathway, shedding new light on an already well-studied pathway[84]. In other cases, sequence can indicate molecular pathways, as in the case of the gene-trap insertion into a novel mouse gene called shroom, because the neural folds mushroom outward and do not converge at the dorsal midline in homozygous mutant embryos. Sequence analysis showed shroom to be a PDZ-domain-containing, actin-binding protein, and biochemical analysis showed that shroom was required for proper subcellular localization of F-actin in the neural tube[85]. Finally, trapping might have its most profound effect on the model organisms in which mutagenesis mediated by homologous recombination cannot be done (BOX 3).

It is hoped that every mouse gene, whether the number is 35,000 or closer to 50,000, will eventually contain at least one trap insertion. In fact, gene trapping should help to define the total number of genes. If and when this happens, the main trapping initiatives will continue to distribute clones and to make use of new tools, including recombinases, integrases, reporters, conditional expression systems and selectable markers. Enzo Medico and Philippe Soriano, for example, have recently developed a novel reporter gene, constructed from fusing ENHANCED GFP (eGFP) to bacterial nitroreductase (designated GFNR). The ROSA–GFNR vector, which contains the *neo* gene driven by an autologous promoter to select for insertions in all genes, can be used for positive and negative selection. This is done by combining flow cytometry for eGFP expression and/or drug selection against nitroreductase expression, to recover genes that are induced or repressed by growth factors and other stimuli, regardless of gene expression in infected cells[86]. New vector designs have also succeeded in inducing gene activation in cell lines[87]; trapping, therefore, might move towards gain-of-function mutations or *in vivo* conditional expression systems using tetracycline-controlled transactivators. *In vitro* functional screens might be achieved either by generating loss of heterozygosity in trapped ES-cell lines[88] using site-specific recombinases, or by combining trapping with chemical mutagenesis or with chromosome engineering. Finally, the mice generated from gene trapping might be the true lasting legacy of the approach. These mice will be used for functional analyses and will act as reporters for sensitized phenotypic screens. With all of these advances, it is clear that gene trapping — unlike de Beauvoir's mouse that was eternally chasing its tail — will continue to evolve, and will ultimately be considered a classic technique that should make many valuable contributions to mouse genetics.

ENHANCED GFP
(eGFP). An autofluorescent 27-kDa protein, originally identified in the jellyfish *Aequorea victoria*, that has been mutated to enhance fluorescence in mammalian cells.

1. Bedell, M. A., Jenkins, N. A. & Copeland, N. G. Mouse models of human disease. I. Techniques and resources for genetic analysis of mice. *Genes Dev.* **11**, 1–10 (1997).
2. Bedell, M. A., Largaespada, D. A., Jenkins, N. A. & Copeland, N. G. Mouse models fo human disease. II. Recent progress and future directions. *Genes Dev.* **11**, 11–43 (1997).
3. Hustad, C. M. *et al.* Molecular genetic characterization of six recessive viable alleles of the mouse agouti locus. *Genetics* **140**, 255–265 (1995).
4. Siracusa, L. D. *et al.* Hypervariable yellow (Ahvy), a new murine agouti mutation: Ahvy displays the largest variation in coat color phenotypes of all known agouti alleles. *J. Hered.* **86**, 121–128 (1995).
5. Duhl, D. M., Vrieling, H., Miller, K. A., Wolff, G. L. & Barsh, G. S. Neomorphic agouti mutations in obese yellow mice. *Nature Genet.* **8**, 59–65 (1994).
6. Silver, L. M. *Mouse Genetics: Concepts and Applications* (Oxford Univ. Press, New York, 1995).
7. Russell, L. B. *et al.* Chlorambucil effectively induces deletion mutations in mouse germ cells. *Proc. Natl Acad. Sci. USA* **86**, 3704–3708 (1989).
8. Russell, W. L. *et al.* Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. *Proc. Natl Acad. Sci. USA* **76**, 5818–5819 (1979).
9. Justice, M. J. Capitalizing on large-scale mouse mutagenesis screens. *Nature Rev. Genet.* **1**, 109–115 (2000).
10. Jaenisch, R. Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus.

© 2001 Macmillan Magazines Ltd

Proc. Natl Acad. Sci. USA **73**, 1260–1264 (1976).
11. Spence, S. E., Gilbert, D. J., Swing, D. A., Copeland, N. G. & Jenkins, N. A. Spontaneous germ line virus infection and retroviral insertional mutagenesis in eighteen transgenic Srev lines of mice. *Mol. Cell. Biol.* **9**, 177–184 (1989).
12. Jenkins, N. A. & Copeland, N. G. High frequency germline acquisition of ecotropic MuLV proviruses in SWR/J-RF/J hybrid mice. *Cell* **43**, 811–819 (1985).
13. Costantini, F. & Lacy, E. Introduction of a rabbit β-globin gene into the mouse germ line. *Nature* **294**, 92–94 (1981).
14. Gordon, J. W. & Ruddle, F. H. Integration and stable germ line transmission of genes injected into mouse pronuclei. *Science* **214**, 1244–1246 (1981).
15. Harbers, K., Jahner, D. & Jaenisch, R. Microinjection of cloned retroviral genomes into mouse zygotes: integration and expression in the animal. *Nature* **293**, 540–542 (1981).
16. Wagner, E. F., Stewart, T. A. & Mintz, B. The human β-globin gene and a functional viral thymidine kinase gene in developing mice. *Proc. Natl Acad. Sci. USA* **78**, 5016–5020 (1981).
17. Doetschman, T. *et al.* Targeted correction of a mutant *HPRT* gene in mouse embryonic stem cells. *Nature* **330**, 576–578 (1987).
18. Thomas, K. R. & Capecchi, M. R. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**, 503–512 (1987).
19. Lakso, M. *et al.* Targeted oncogene activation by site-specific recombination in transgenic mice. *Proc. Natl Acad. Sci. USA* **89**, 6232–6236 (1992).
20. Orban, P. C., Chui, D. & Marth, J. D. Tissue- and site-specific DNA recombination in transgenic mice. *Proc. Natl Acad. Sci. USA* **89**, 6861–6865 (1992).
21. Dymecki, S. M. Flp recombinase promotes site-specific DNA recombination in embryonic stem cells and transgenic mice. *Proc. Natl Acad. Sci. USA* **93**, 6191–6196 (1996).
22. Weber, F., de Villiers, J. & Schaffner, W. An SV40 ‘enhancer trap’ incorporates exogenous enhancers or generates enhancers from its own sequences. *Cell* **36**, 983–992 (1984).
23. O’Kane, C. J. & Gehring, W. J. Detection *in situ* of genomic regulatory elements in *Drosophila*. *Proc. Natl Acad. Sci. USA* **84**, 9123–9127 (1987).
24. Allen, N. D. *et al.* Transgenes as probes for active chromosomal domains in mouse development. *Nature* **333**, 852–855 (1988).
25. Kothary, R. *et al.* A transgene containing *lacZ* inserted into the dystonia locus is expressed in the neural tube. *Nature* **335**, 435–437 (1988).
26. Gossler, A., Joyner, A. L., Rossant, J. & Skarnes, W. C. Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Science* **244**, 463–465 (1989).
The first example of enhancer-trap and gene-trap mutagenesis in mouse ES cells, and of ES-cell-derived mice. It was published the same year as the first targeted mutagenesis of non-selectable genes.
27. Korn, R. *et al.* Enhancer trap integrations in mouse embryonic stem cells give rise to staining patterns in chimaeric embryos with a high frequency and detect endogenous genes. *Mech. Dev.* **39**, 95–109 (1992).
28. von Melchner, H. & Ruley, H. E. Identification of cellular promoters by using a retrovirus promoter trap. *J. Virol.* **63**, 3227–3233 (1989).
29. Friedrich, G. & Soriano, P. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev.* **5**, 1513–1523 (1991).
The development of the ROSA β-geo vectors and the first application of gene trapping to isolate lethal mutations.
30. Reddy, S., DeGregori, J. V., von Melchner, H. & Ruley, H. E. Retrovirus promoter-trap vector to induce *lacZ* gene fusions in mammalian cells. *J. Virol.* **19**, 17–23 (1991).
31. von Melchner, H. *et al.* Selective disruption of genes expressed in totipotent embryonal stem cells. *Genes Dev.* **6**, 919–927 (1992).
This paper provided the foundation for the current promoter-trapping efforts.
32. Frohman, M. A., Dush, M. K. & Martin, G. R. Rapid production of full-length cDNAs from rare transcripts: amplification using a single-gene specific oligonucleotide primer. *Proc. Natl Acad. Sci. USA* **85**, 8998–9002 (1988).
33. Skarnes, W. C., Auerbach, B. A. & Joyner, A. L. A gene trap approach in mouse embryonic stem cells: the *lacZ* reporter is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice. *Genes Dev.* **6**, 903–918 (1992).
34. Stanford, W. L. *et al.* Expression trapping: identification of novel genes expressed in hematopoietic and endothelial lineages by gene trapping in ES cells. *Blood* **92**, 4622–4631 (1998).
35. McClive, P. *et al.* Gene trap integrations expressed in the developing heart: insertion site affects splicing of the PT1-ATG vector. *Dev. Dyn.* **212**, 267–276 (1998).
36. Evans, M. J. Gene trapping — a preface. *Dev. Dyn.* **212**, 167–169 (1998).
37. Niwa, H. *et al.* An efficient gene-trap method using poly A trap vectors and characterization of gene-trap events. *J. Biochem.* **113**, 343–349 (1993).
38. Yoshida, M. *et al.* A new strategy of gene trapping in ES cells using 3′RACE. *Trans. Res.* **4**, 277–287 (1995).
39. Salminen, M., Meyer, B. I. & Gruss, P. 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 (1998).
40. Zambrowicz, B. P. *et al.* Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. *Nature* **392**, 608–611 (1998).
41. Hicks, G. G. *et al.* Functional genomics in mice by tagged sequence mutagenesis. *Nature Genet.* **16**, 338–344 (1997).
The first large-scale, genotype-driven gene-trap screen, which also introduces an elegant method of cloning trapped sequences. This project has also evolved into one of the gene-trap resource centres.
42. Skarnes, W. C., Moss, J. E., Hurtley, S. M. & Beddington, R. S. Capturing genes encoding membrane and secreted proteins important for mouse development. *Proc. Natl Acad. Sci. USA* **92**, 6592–6596 (1995).
An elegant method to trap a specific class of gene. This project has evolved into one of the gene-trap resource centres.
43. Araki, K. *et al.* Exchangeable gene trap using the Cre/mutated lox system. *Cell. Mol. Biol.* **45**, 737–750 (1999).
44. Hardouin, N. & Nagy, A. Gene-trap-based target site for Cre-mediated transgenic insertion. *Genesis* **26**, 245–252 (2000).
45. Enver, T. & Greaves, M. Loops, lineage, and leukemia. *Cell* **94**, 9–12 (1998).
46. Donoviel, D. B. *et al.* Proteinuria and perinatal lethality in mice lacking NEPH1, a novel protein with homology to NEPHRIN. *Mol. Cell. Biol.* **21**, 4829–4836 (2001).
47. Breitman, M. L. *et al.* Genetic ablation: targeted expression of a toxin gene causes microphthalmia in transgenic mice. *Science* **238**, 1563–1565 (1987).
48. Westerman, K. A. & Leboulch, P. Reversible immortalization of mammalian cells mediated by retroviral transfer and site-specific recombination. *Proc. Natl Acad. Sci. USA* **93**, 8971–8976 (1996).
49. Mitchell, K. J. *et al.* Functional analysis of secreted and transmembrane proteins critical to mouse development. *Nature Genet.* **28**, 241–249 (2001).
50. Wurst, W. *et al.* A large-scale gene-trap screen for insertional mutations in developmentally regulated genes in mice. *Genetics* **139**, 889–899 (1995).
This work served as proof of principle for high-throughput gene-trap screens.
51. Stoykova, A., Chowdhury, K., Bonaldo, P., Torres, M. & Gruss, P. Gene trap expression and mutational analysis for genes involved in the development of the mammalian nervous system. *Dev. Dyn.* **212**, 198–213 (1998).
52. Leighton, P. A. *et al.* Defining brain wiring patterns and mechanisms through gene trapping in mice. *Nature* **410**, 174–179 (2001).
53. Tropepe, V. *et al.* Direct neural fate specification from embryonic stem cells: a primitive mammalian neural stem cell stage acquired through a default mechanism. *Neuron* **30**, 65–78 (2001).
54. Doetschman, T. C., Eistetter, H., Katz, M., Schmidt, W. & Kemler, R. The *in vitro* development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood islands and myocardium. *J. Embryol. Exp. Morphol.* **87**, 27–45 (1985).
55. Wang, R., Clark, R. & Bautch, V. L. Embryonic stem cell-derived cystic embryoid bodies form vascular channels: an *in vitro* model of blood vessel development. *Development* **114**, 303–316 (1992).
56. Nakano, T., Kodama, H. & Honjo, T. Generation of lymphohematopoietic cells from embryonic stem cells in culture. *Science* **265**, 1098–1101 (1994).
57. Choi, K., Kennedy, M., Kazarov, A., Papadimitriou, J. C. & Keller, G. A common precursor for hematopoietic and endothelial cells. *Development* **125**, 725–732 (1998).
58. Schmitt, R. M., Bruyns, E. & Snodgrass, H. R. Hematopoietic development of embryonic stem cells *in vitro*: cytokine and receptor gene expression. *Genes Dev.* **5**, 728–740 (1991).
59. Bautch, V. L. *et al.* Blood island formation in attached cultures of murine embryonic stem cells. *Dev. Dyn.* **205**, 1–12 (1996).
60. Abe, K. *et al.* Endoderm-specific gene expression in embryonic stem cells differentiated to embryoid bodies. *Exp. Cell Res.* **229**, 27–34 (1996).
61. Schuldiner, M., Yanuka, O., Itskovitz-Eldor, J., Melton, D. A. & Benvenisty, N. Effects of eight growth factors on the differentiation of cells derived from human embryonic stem cells. *Proc. Natl Acad. Sci. USA* **97**, 11307–11312 (2000).
62. Weiss, M. J., Keller, G. & Orkin, S. H. Novel insights into erythroid development revealed through *in vitro* differentiation of *GATA-1*$^{-}$ embryonic stem cells. *Genes Dev.* **8**, 1184–1197 (1994).
63. Shalaby, F. *et al.* A requirement for Flk-1 in primitive and definitive hematopoiesis, and vasculogenesis. *Cell* **89**, 981–990 (1997).
64. Narita, N., Bielinska, M. & Wilson, D. B. Cardiomyocyte differentiation by GATA-4-deficient embryonic stem cells. *Development* **122**, 3755–3764 (1996).
65. Joza, N. *et al.* Essential role of the mitochondrial apoptosis inducing factor in programmed cell death. *Nature* **410**, 549–554 (2001).
66. Muth, K., Bruyns, R., Thorey, I. S. & von Melchner, H. Disruption of genes regulated during hematopoietic differentiation of mouse embryonic stem cells. *Dev. Dyn.* **212**, 277–283 (1998).
67. Hidaka, M. *et al.* Gene trapping of two novel genes, *Hzf* and *Hhl*, expressed in hematopoietic cells. *Mech. Dev.* **90**, 3–15 (2000).
68. Baker, R. K. *et al.* *In vitro* preselection of gene-trapped embryonic stem cell clones for characterizing novel developmentally regulated genes in the mouse. *Dev. Biol.* **185**, 201–214 (1997).
69. Shirai, M. *et al.* A gene trap strategy for identifying the gene expressed in the embryonic nervous system. *Zool. Sci.* **13**, 277–283 (1996).
70. Forrester, L. *et al.* Induction gene trapping in ES cells: identification of developmentally regulated genes that respond to retinoic acid. *Proc. Natl Acad. Sci. USA* **93**, 1677–1682 (1996).
The first high-throughput induction-trap screen.
71. Sam, M. *et al.* Aquarius, a novel gene isolated by gene trapping with an RNA-dependent polymerase motif. *Dev. Dyn.* **212**, 304–317 (1998).
72. Gajovic, S., Chowdhury, K. & Gruss, P. Genes expressed after retinoic acid-mediated differentiation of embryoid bodies are likely to be expressed during embryo development. *Exp. Cell Res.* **242**, 138–143 (1998).
73. Komada, M., McLean, D. J., Griswold, M. D., Russell, L. D. & Soriano, P. *E-MAP-115*, encoding a microtubule-associated protein, is a retinoic acid-inducible gene required for spermatogenesis. *Genes Dev.* **14**, 1332–1342 (2000).
74. Mainguy, G. *et al.* An induction gene trap for identifying a homeoprotein-regulated locus. *Nature Biotechnol.* **18**, 746–749 (2000).
A powerful use of gene trapping to isolate genes downstream of a transcription factor.
75. Vallis, K. A., Chen, Z., Stanford, W. L., Hill, R. P. & Bernstein, A. Idenitification of ionizing radiation-responsive genes: a large-scale, induction gene trap strategy in embryonic stem cells. *Radiat. Res.* (in the press).
76. Kerr, W. G., Heller, M. & Herzenberg, L. A. Analysis of lipopolysaccharide-response genes in B-lineage cells demonstrates that they can have differentiation stage-restricted expression and contain SH2 domains. *Proc. Natl Acad. Sci. USA* **93**, 3947–3952 (1996).
77. Russ, A. P. *et al.* Identification of genes induced by factor deprivation in hematopoietic cells undergoing apoptosis using gene-trap mutagenesis and site-specific recombination. *Proc. Natl Acad. Sci. USA* **93**, 15279–15284 (1996).
This paper shows a use of gene trapping other than for mutagenesis.
78. Andreu, T., Beckers, T., Thoenes, E., Hilgard, P. & von Melchner, H. Gene trapping identifies inhibitors of oncogenic transformation. The tissue inhibitor of metalloproteinases-3 (TIMP3) and collagen type I α2 (COL1A2) are epidermal growth factor-regulated growth repressors. *J. Biol. Chem.* **273**, 13848–13854 (1998).
79. Akiyama, N., Matsuo, Y., Sai, H., Noda, M. & Kizaka-Kondoh, S. Identification of a series of transforming growth factor β-responsive genes by retrovirus-mediated gene trap screening. *Mol. Cell. Biol.* **20**, 3266–3273 (2000).
80. Vidal, F. *et al.* Gene trap analysis of germ cell signaling to Sertoli cells: NGF-TrkA mediated induction of *Fra1* and *Fos* by post-meiotic germ cells. *J. Cell. Sci.* **114**, 435–443 (2001).
81. Skarnes, W. C. Gene trapping methods for the identification and functional analysis of cell surface proteins. *Methods Enzymol.* **328**, 592–615 (2000).
82. Townley, D. J., Avery, B. J., Rosen, B. & Skarnes, W. C. Rapid sequence analysis of gene trap integrations to generate a resource of insertional mutations in mice.

© 2001 Macmillan Magazines Ltd

Genome Res. **7**, 293–298 (1997).

83. Nadeau, J. H. *et al.* Sequence interpretation. Functional annotation of mouse genome sequences. *Science* **291**, 1251–1255 (2001).
84. Pinson, K. I., Brennan, J., Monkley, S., Avery, B. J. & Skarnes, W. C. An LDL receptor-related protein mediates Wnt signaling in mice. *Nature* **407**, 535–538 (2000).
85. Hildebrand, J. D. & Soriano, P. Shroom, a PDZ domain-containing actin-binding protein, is required for neural tube morphogenesis in mice. *Cell* **99**, 485–497 (1999).
86. Medico, E., Gambarotta, G., Gentile, A., Comoglio, P. M. & Soriano, P. A gene trap vector system for identifying transcriptionally responsive genes. *Nature Biotechnol.* **19**, 579–582 (2001).
87. Harrington, J. J. *et al.* Creation of genome-wide protein expression libraries using random activation of gene expression. *Nature Biotechnol.* **19**, 440–445 (2001).
88. Lefebvre, L., Dionne, N., Karaskova, J., Squire, J. & Nagy, A. Selection for transgene homozygosity in embryonic stem cells results in extensive loss of heterozygosity. *Nature Genet.* **27**, 257–258 (2001).
89. Harbers, K., Kuehn, M., Delius, H. & Jaenisch, R. Insertion of retrovirus into the first intron of α1(I) collagen gene leads to embryonic lethal mutation in mice. *Proc. Natl Acad. Sci. USA* **81**, 1504–1508 (1984).
90. Vijaya, S., Steffen, D. L. & Robinson, H. L. Acceptor sites for retroviral integrations map near DNaseI-hypersensitive sites in chromatin. *J. Virol.* **60**, 683–692 (1986).
91. Soriano, P., Gridley, T. & Jaenisch, R. Retroviruses and insertional mutagenesis in mice: proviral integration at the Mov 34 locus leads to early embryonic death. *Genes Dev.* **1**, 366–375 (1987).
92. Chowdhury, K., Bonaldo, P., Torres, M., Stoykova, A. & Gruss, P. Evidence for the stochastic integration of gene trap vectors into the mouse germline. *Nucleic Acids Res.* **25**, 1531–1536 (1997).
93. Hrabe de Angelis, M. H. *et al.* Genome-wide, large-scale production of mutant mice by ENU mutagenesis. *Nature Genet.* **25**, 444–447 (2000).
94. Wiles, M. V. *et al.* Establishment of a gene-trap sequence tag library to generate mutant mice from embryonic stem cells. *Nature Genet.* **24**, 13–14 (2000).
This paper describes the first set of clones from the largest academic gene-trap library.
95. Bellen, H. J. *et al.* P-element-mediated enhancer detection: a versatile method to study development in *Drosophila*. *Genes Dev.* **3**, 1288–1300 (1989).
96. Wilson, C. *et al.* P-element-mediated enhancer detection: an efficient method for isolating and characterizing developmentally regulated genes in *Drosophila*. *Genes Dev.* **3**, 1301–1313 (1989).
97. Bier, E. *et al.* Searching for pattern and mutation in the *Drosophila* genome with a P-*lacZ* vector. *Genes Dev.* **3**, 1273–1287 (1989).
98. Spradling, A. C. *et al.* The Berkeley *Drosophila* Genome Project Gene Disruption Project: single P-element insertions mutating 25% of vital *Drosophila* genes. *Genetics* **153**, 135–177 (1999).
99. Claes, B., Smalle, J., Dekeyser, R., Van Montagu, M. & Caplan, A. Organ-dependent regulation of a plant promoter isolated from rice by 'promoter-trapping' in tobacco. *Plant J.* **1**, 15–26 (1991).
100. Meissner, R. *et al.* A high throughput system for transposon tagging and promoter trapping in tomato. *Plant J.* **22**, 265–274 (2000).
101. Amsterdam, A. *et al.* A large-scale insertional mutagenesis screen in zebrafish. *Genes Dev.* **13**, 2713–2724 (1999).
102. Talbot, W. S. & Hopkins, N. Zebrafish mutations and functional analysis of the vertebrate genome. *Genes Dev.* **14**, 755–762 (2000).
103. Zachgo, J. & Gossler, A. in *Mammalian Development* (ed. Lonai, P.) 321–337 (Harwood Academic Publishers, Amsterdam, 1996).

Acknowledgements

We thank the following investigators who have shared unpublished data and preprints to help make this review as current as possible: G. Hicks, E. Ruley, P. Soriano, W. C. Skarnes, J. Wrana, W. Wurst and K.-I. Yamamura. We thank the Canadian Institutes of Health Research and the Leukemia & Lymphoma Society for generous funding of gene-trap research to W.L.S. We apologize to the many colleagues who have provided valuable insights and large amounts of data in the gene-trap field whose manuscripts were not cited owing to strict space limitations.

Online links

DATABASES

MGI: http://www.informatics.jax.org/
a | a^{16H} | a^{e} | A^{hvy} | A^{iy} | A^{vy} | Bmp2 | *Ext1* | *Lrp6* | shroom | *W*

FURTHER INFORMATION

Alexandra Joyner's lab: http://saturn.med.nyu.edu/groups/JoynerLab/
Berkeley Drosophila Genome Project: http://www.fruitfly.org/
Earl Ruley's lab: http://db.system-x.com/microbio/faculty/pi.cfm?ID=149
Harwell: http://www.mgu.har.mrc.ac.uk/
Janet Rossant's lab: http://www.mshri.on.ca/develop/rossant/homepage.html
Lexicon Genetics' 'OmniBank': http://www.lexgen.com/omnibank/omnibank.htm
Nancy Hopkins' lab: http://web.mit.edu/biology/www/Ar/hopkins.html
Oakridge: http://bio.lsd.ornl.gov/mgd/
Program in Developmental Biology and Fetal Health: http://www.mshri.on.ca/stanford
The BayGenomics Gene Trap Project: http://baygenomics.ucsf.edu/
The CMHD Gene Trap Project: http://www.cmhd.ca/genetrap
The Gene Trap Project of the German Human Genome Project: http://tikus.gsf.de/
The Jackson Laboratory: http://jaxmice.jax.org/index.shtml
University of Manitoba Institute of Cell Biology: http://www.escells.ca
William Stanford's lab: http://www.mshri.on.ca/stanford

© 2001 Macmillan Magazines Ltd