A modified gene trap approach to identify secretory molecules involved in mouse development

JANE BRENNAN

Thesis presented for the degree of Doctor of Philosophy
University of Edinburgh

1997

I declare that the work presented in this thesis is my own,
except where otherwise stated.

Jane Brennan

To Euan and my family

ACKNOWLEDGMENTS

I would like to thank my supervisor Bill Skarnes for inspiration, support and advice during the course of this PhD. I would also like to thank Val Wilson, Ian Chambers, Penny Rashbass and Julie Wallis for helpful suggestions on drafts of my thesis. Many thanks go to Andrew Jeske for looking after my animals so diligently. I would like to thank past members of the Skarnes group for creating a wonderful environment to learn and play in. In particular I would like to thank Brian Avery, Sue Monkley, Julie Moss and Barry Rosen for their science and their chat. I also want to thank Martin, Chris, Ben, Mog and Si for making Edinburgh a fun place to be. Finally I want to thank Euan for love and support and for being my best friend.

Abstract

Gene trapping in murine embryonic stem (ES) cells provides a method to create random insertional mutations in mice which are easily accessible to molecular characterisation. This approach involves electroporating a reporter gene construct into ES cells which is transcribed following integration into an endogenous transcription unit. In an effort to understand the molecular mechanisms that mediate cellular interactions during development I have used a modified gene trap vector, the secretory trap vector (pGT1.8TM), to enrich for insertions in genes encoding secreted and membrane-spanning molecules. The selective property of this vector depends on the presence of a 680 bp fragment containing the transmembrane domain of CD4 fused to the bgeo reporter gene.

Secretory trap lines were generated by electroporating pGT1.8TM into CGR8 ES cells. The expression patterns of trapped genes were determined by monitoring reporter gene activity in chimeric embryos between 8.5d and 10.5d of development. The endogenous genes carrying vector insertions were identified from the fusion transcripts using the 5' RACE. Two thirds of the insertions occurred in secretory molecules with the remainder in non-secretory molecules. A model is proposed to explain why the insertions in non-secreted genes are detected.

The requirement for pGT1.8TM to insert into introns is likely to favour insertion into genes composed of large intronic regions. To detect insertions in smaller transcription units composed of few or no introns I developed an exon trap version of this vector (pET1.8TM) by removing the splice acceptor sequence from pGT1.8TM. Previous reports indicated that the efficiency of gene trapping was reduced by 12 to 40 fold in the absence of a splice acceptor, however I obtained similar numbers of ßgal staining colonies following electroporation of pGT1.8TM and pET1.8TM in ES cells. In all eight pET1.8TM integrations characterised deletions had occurred at the 5' end of the vector prior to integration into the genome. Deletions which interfere with reporter activity may explain the poor frequency of events seen with previous attempts at exon trapping and suggest that the addition of buffer sequences upstream of the reporter increases the efficiency of pET1.8TM.

One line generated in this study, JST185, carries an insertion of pGT1.8TM and was selected for detailed characterisation on the basis of its spatially restricted expression pattern during development. The disrupted gene encodes a multimembrane spanning protein which contains a tandem array of tetratricopeptide repeats (TPRs), motifs shown to mediate protein-protein interactions. The expression pattern of JST185 during embryogenesis suggests it plays a role in mediating epithelial-mesenchymal interactions during development. Most animals homozygous for the insertion die shortly after birth due to respiratory failure. However, a few homozygous animals survive beyond weaning and display a striking lung pathology.

Abbreviations

Ac	*Activator*
ATG	Initiation of translation codon
βgal	*βgalactosidase*
βgeo	*βgalactosidase* and *neomycin* fusion
cM	centiMorgan
CHL	*chlorambucil*
CNS	central nervous system
dpc	days post coitum
DNA	deoxyribonucleic acid
Ds	*Dissociation*
En2	mouse *engrailed-2*
ENU	ethylnitrosourea
ER	endoplasmic reticulum
ES cells	Embryonic Stem cells
ET	exon trap
GalTase	*β1,4-galactosyltransferase*
gDNA	genomic DNA
GT	gene trap
hsp	*heat shock protein*
LTR	long terminal repeat
neo^R	G418 resistant
PCR	Polymerase Chain Reaction
RACE	Rapid Amplification of cDNA Ends
RFLP	restriction fragment length polymorphism
RPA	RNase protection assay
RNA	ribonucleic acid
sm-actin	smooth muscle actin
SA	splice acceptor
SD	splice donor
SS	signal sequence
T	*Brachyury*
Ti	tumour inducing plasmid
TPR	tetratricopeptide repeat domain
TM	transmembrane
UTR	untranslated region

Nomenclature

Gene symbols are always italicised. All letters in human genes are upper-case; all letters in mouse genes are lower case, excepting the first letter.

Protein products of genes are in Roman type; human proteins are upper-case.

TABLE OF CONTENTS

Chapter 6

MATERIALS AND METHODS

FIGURES

TABLES

Chapter 1

GENERAL INTRODUCTION

1.1 Introduction to mutational analysis of development

Mutation analysis is a powerful tool that has been used to unravel many of the complexities of embryonic development. By creating mutations, the biological function of a single gene product can be studied *in vivo*. Interactions among gene products can also be deciphered by breeding different mutations together. Probably the most informative mutagenic screens have been performed in the fruitfly *Drosophila melanogaster* where the systematic isolation and analysis of mutants have revealed many of the genes involved in embryonic pattern formation (Nüsslein-Volhard *et al.*, 1984). A similar large scale genetic screen has been performed in the nematode *Caenorhabditis elegans* (reviewed in Kemphues, 1988) and more recently in a vertebrate, the zebrafish *Danio rerio* (Driever *et al.*, 1996; Haffter *et al.*, 1996). During the latter screen over 400 genes essential for early vertebrate development were identified. Phenotypic descriptions of these mutations have already identified many of the morphogenetic processes which lead to vertebrate embryonic patterning.

However, in any genetic screen it is essential that the mutated gene be easily identifiable, and in zebrafish this is not straightforward at present. However, understanding vertebrate development, and in particular mammalian development, requires genetic analysis.

An obvious candidate for a vertebrate genetic screen is the mouse as its detailed genetic map should allow the identification of virtually any disrupted gene. However, the small litter size and the relatively long generation times of mice compared to invertebrates make large genetic screens very costly. Consequently other strategies have been used to identify genes involved in early mammalian development. One particularly fruitful strategy has been the creation of null mutations using homologous recombination in embryonic stem (ES) cells (Capecchi, 1989). Candidate genes for targeted loss of function analysis are selected for study on the basis of the biochemical properties of the encoded protein, their expression patterns, or by extrapolation from data on the function of their homologues in other systems such as *Drosophila* (reviewed in Kessel & Gruss, 1990). Although more than 700 genes have been disrupted in this manner (Brandon *et al.*, 1995), the isolation and mapping of genomic clones and construction of targeting constructs for each candidate gene is labour intensive and not amenable to a large scale screen. In addition, the 'homology cloning' approach, although very successful, will fail to identify genes which are specifically involved in mammalian development. Thus it is important to consider alternative methods to creating new mutations.

The purpose of this introduction is to review the mutagenic strategies being used to generate a resource of new mouse mutants. Particular emphasis is placed on comparing the ease with which disrupted loci can be identified using the different approaches. The first section describes commonly used mutagens; radiation, DNA-modifying chemicals and foreign DNA. The second section introduces the entrapment technique which involves the random integration of a vector into a transcriptional unit

such that a reporter gene is placed under the control of endogenous regulatory elements. Insertions simultaneously mutate the endogenous gene and provide a molecular tag to simplify its identification. Entrapment technology in the mouse derives in part from strategies used in other organisms, so these too are discussed. The advantages of applying such techniques in mouse ES cells are discussed in the third section. Briefly, the use of these cells allows pre-selection of mutations based on *in vitro* criteria and also allows the subsequent germline transmission of the mutation. The pre-selection for insertions into secreted and membrane bound proteins using a modified gene trap vector is discussed in the final section.

1.1.1 Classical mouse mutants

The first mouse mutants to be studied arose spontaneously or were the offspring of mice exposed to X-rays (Lyon *et al.*, 1996). They were identified on the basis of a visible mutant phenotype such as coat colour, hair morphology, gross behavioural defects and skeletal abnormalities. Advancing from descriptions of phenotype to identifying the responsible molecular lesion is the rate limiting step in these types of studies. In some cases the mutated genes have been identified serendipitously by colocalizing "candidate" genes on the physical genetic map. This approach identified the gene on chromosome 5 responsible for the various *Dominant white spotting (W)* phenotypes which display defects including lack of hair pigmentation, sterility and severe macrocytic anaemia (Chabot *et al.*, 1988). The *c-kit* receptor tyrosine kinase was selected as a candidate gene for the W locus on the basis of it mapping to the human region syntenic to mouse chromosome 5 and from the observation that it is expressed in haematopoietic tissues. Interspecific backcross analysis demonstrated that the W locus is tightly linked to *c-kit.* Later characterisation confirmed the presence of point mutations in the kinase domain of *c-kit* which accounted for deficient enzyme activity in several W alleles (Reith *et al.*, 1990). Another spontaneous mutant, Steel (Sl), shows similar pleiotropic defects to W and

has been found to encode the ligand for the *c-kit* receptor (Huang *et al.*, 1990). Analysis of the W and Sl mutants has revealed a role for Kit mediated signal transduction pathways in the development of haematopoietic, melanogenic and germ cell lineages.

In the absence of candidate genes, mutant loci can in principle be identified by positional cloning. This approach depends on finding linkage to the mutant phenotype which can then be used to create a high resolution genetic map followed by a physical map. Finally the region is examined for candidate genes. The complexity of the mammalian genome makes such an approach a formidable undertaking. The first mouse gene to be positionally cloned was *Brachyury (T)* located on chromosome 17 (Hermann, 1990). The adult heterozygous phenotype is a short tail. Embryos homozygous for the phenotype die during midgestation. The success of positionally cloning *T* was undoubtedly due to a 60 year accumulation and analysis of different alleles and locus rearrangements in the *T* region.

Cloning of another classic mutant, the short ear locus, was aided by the inclusion of this gene in large specific-locus mutagenesis experiments (Russell, 1951). In the specific locus test, wild type animals were treated with X-rays or chemicals and mated to a special marker strain which is homozygous for the short ear mutation and 6 distinct recessive coat colour markers. If a mutation is induced at one of the specific loci, the associated mutant phenotype is uncovered in offspring. Such experiments led to the identification of hundreds of new mutations that affect the short ear gene and closely linked coat-colour markers. A detailed linkage map was built around the short ear and dilute loci which was the basis of a chromosome walk (Kingsley *et al.*, 1992). The gene for bone morphogenetic protein 5 (BMP-5) was identified in the short ear region and found to be deleted or rearranged in several independent short ear mutants (Kingsley *et al.*, 1992).

1.1.2 "Saturation" mutagenesis

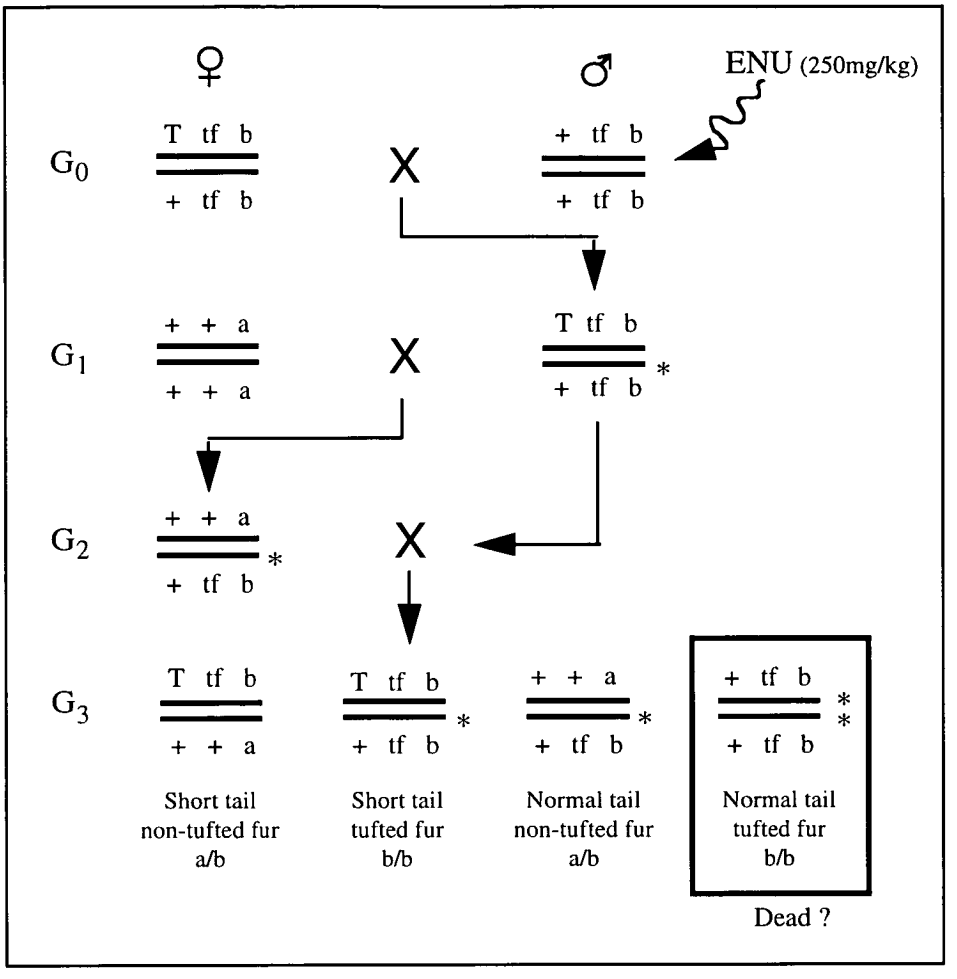
Due to its high mutational efficiency chemical mutagenesis was used to generate *Drosophila* mutants in the saturation screen of Nüsslein-Volhard and co-workers (Nüsslein-Volhard *et al.*, 1984). In the mouse, chemical mutagens such as ethylnitrosourea (ENU) and chlorambucil (CHL) can induce mutations at an average per locus frequency of greater than 1 in 1000 which is ten fold more efficient than using X-irradiation and up to a thousand fold more efficient than spontaneous rates (Rinchik, 1991). Rather than the large chromosomal deletions and rearrangements induced by X-rays, ENU generates point mutations in DNA and CHL induces short chromosomal deletions.

Although genetic screens for recessive phenotypes in mice are a huge undertaking, two regions of the mouse genome have been subjected to ENU mutagenesis (Shedlovsky *et al.*, 1988; Rinchik, 1991). In the first example alleles of new genes were isolated from the *t* region of chromosome 17. The breeding strategy is described in Figure 1.1A. The offspring of mutagenised males were screened directly for dominant mutations and a three generation breeding protocol was used to recover recessive mutations. From 423 mutagenised gametes 17 lethal mutations were recovered. This screen was a large undertaking and was aided by a number of linked markers and an inversion that prevents recombination yet it still failed to saturate an area which is equivalent to only 0.01% of the mouse genome. If a random ENU screen for recessive lethal mutations was to be performed without linked markers huge numbers of mice would be required. For example, if we presume an average litter size is 8 offspring, starting with only 100 mice, by the third generation 51,200 mice will have been generated. Such enormous numbers only make it feasible to mutagenise small well-defined areas containing selectable genetic markers.

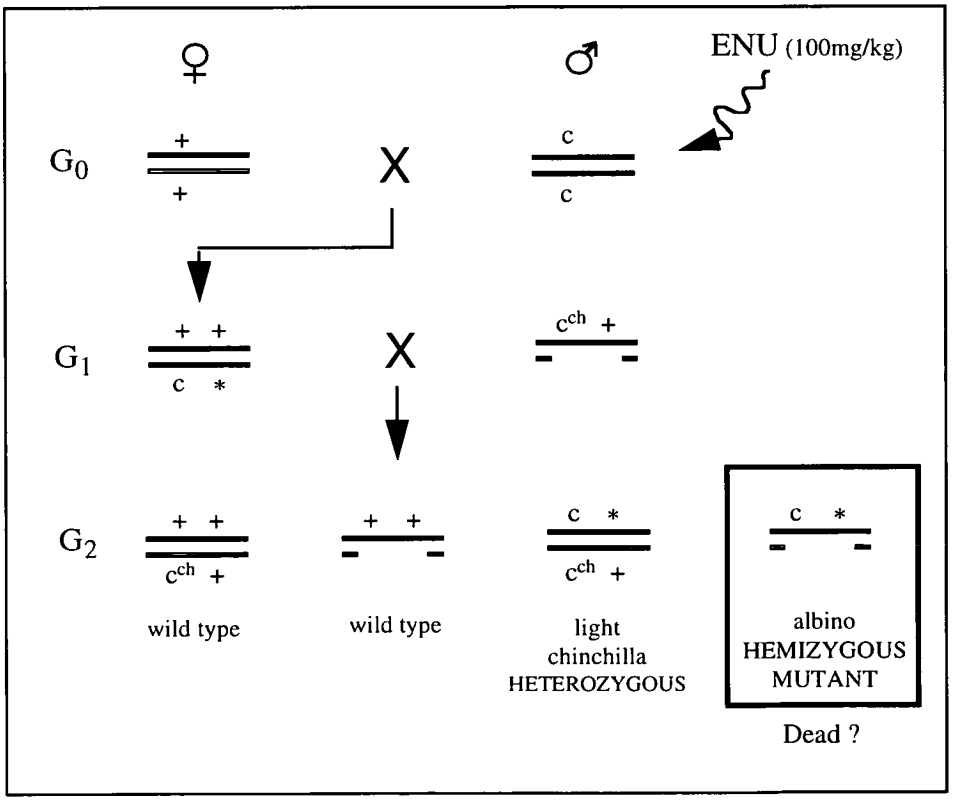
Figure 1.1. Breeding schemes to detect ENU-induced mutations

A. Recessive mutations were identified within a 15cM region of mouse chromsome 17 around the t region (taken from Shedlovsky *et al.*, 1988). Chromosomal markers (proximal to distal) are: T (*Brachyury*), tf (*tufted*), and haplotypes of the major histocompatibility complex locus a and b. The asterisks indicate the mutagenized chromosome. By the third generation new, recessive ENU-induced mutations can be detected by the absence of normal tailed, tufted $H\text{-}2^{b}/H\text{-}2^{b}$ animals.

B. Recessive mutations were recovered in the *albino* region by breeding mutagenized males to a mutant carrying a deletion around *albino*. Genetic markers include the c (albino) and c^{ch} (chinchilla) alleles of the albino locus. The heavily outlined box represents the G2 albino test class. The absence of albino progeny indicates a new recessive lethal. The G2 light chinchilla carrier can be used to retrieve the mutation.


♀
♂
ENU (250mg/kg)
G0
T tf b
+ tf b
X
+ tf b
+ tf b
G1
+ + a
+ + a
X
T tf b
+ tf b
*
G2
+ + a
+ tf b
*
X
G3
T tf b
+ + a
T tf b
+ tf b
*
+ + a
+ tf b
*
+ tf b
+ tf b
* *
Short tail
non-tufted fur
a/b
Short tail
tufted fur
b/b
Normal tail
non-tufted fur
a/b
Normal tail
tufted fur
b/b
Dead ?

A


♀
♂
ENU (100mg/kg)
G0
+
+
X
c
c
G1
+ +
c *
X
cch +
G2
+ +
cch +
+ +
c *
cch +
c *
albino
HEMIZYGOUS
MUTANT
wild type
wild type
light
chinchilla
HETEROZYGOUS
Dead ?

B

In the second example Rinchik and colleagues utilised mice carrying deletions around the *albino* locus in an ENU mutagenesis screen. To recover recessive lethal mutations a two generation screen was devised in which the progeny of mutagenised males were mated to the deletion stock. As the chromosomes are genetically marked by coat colour the mutagenised chromosome can be tracked through the crosses. (Figure 1.1B). 3000 gametes were screened and 6 new lethal loci were identified. Repeat mutations were discovered at each newly defined locus suggesting that saturation mutagenesis of the region was reached.

Breeding ENU mutagenised males to a deletion stock is a powerful way to create new mutants using fewer animals than the three generation crosses. However, it is dependent on deletion stocks covering the genome which need to be genetically characterised. The recent development of efficient methods to generate new chromosomal deletions in defined areas using site specific recombination (Ramirez-Solis *et al.*, 1995) or γ-irradiation (You *et al.*, 1997) should make this a feasible approach to take. In addition, the recovery of recessive mutations against a defined chromosome deletion pre-localises the mutation thus simplifying identification of the disrupted gene. However, less than 11 % of the genes responsible for mutant phenotypes have been cloned in the 900 mouse strains carrying spontaneous, radiation or chemical induced mutations (Dietrich *et al.*, 1995). Currently the genetic map in the mouse contains over 14, 000 markers (Dietrich *et al.*, 1995), which is equivalent to one approximately every 0.1 cM or 200 kb. However, as the resolution of this physical map increases it will undoubtedly become feasible to clone mutated genes on the basis of position alone. In summary, the difficulty in making the leap from mutant phenotype to a molecular analysis of the defect continues to pose a serious drawback to radiation and chemical induced mutagenic strategies.

1.1.3 Insertional mutagenesis

Insertional mutagenesis aims to alleviate problems associated with cloning mutant loci as the introduction of foreign DNA provides a molecular tag to simplify cloning. Although the mutation rate is less efficient than chemical mutagenesis the benefit of easy access to mutated loci makes it an attractive approach to take. Insertional mutations have been introduced into mice by microinjection of cloned DNA into zygotes to generate transgenic mice, by retroviral infection of embryos or embryonic stem cells and by electroporation of DNA into ES cells (reviewed in Rijkers *et al.*, 1994). Approximately 5-10 % of transgenic lines and retroviral insertions result in mutant phenotypes.

Although in theory transgene insertions should simplify the identification of the mutated gene, in reality cloning is often complicated by complex deletions, duplications and rearrangements which often accompany transgene insertion events. Nevertheless, a number of classical mutations have been identified using transgene insertions as a molecular tag such as *limb deformity* (Woychik *et al.*, 1990), *microphthalmia* (Hodgkinson *et al.*, 1993) and *dystonia* (Brown *et al.*, 1995). In contrast to transgenic insertions, cloning genes interrupted by retroviral infection is often more straightforward. This is because the viral endonuclease which catalyses integration does not generate gross DNA rearrangements. Integration simply results in a 4-6bp duplication of cellular DNA at the integration site.

Curiously over half the transgenic insertional mutations (13/23) have been found to be allelic with classical mutations (Meisler, 1992). A number of factors may contribute to such a high frequency of re-isolation. A major consideration is that transgenic insertional mutations are not made with the sole purpose of generating mutant phenotypes. As it is easier to characterise a phenotype that has been previously isolated it is likely that viable insertional mutants that exhibit visible abnormalities, like

most of the classical mutants, will be frequently re-isolated. Another contributing factor is the evidence for particular sites in the genome being more susceptible to insertion than others. Evidence for preferred sites comes from two sets of data which show that independent transgene insertion events occurred in the same genes (Hodgkinson *et al.*, 1993; Krakowsky *et al.*, 1993; Vogt *et al.*, 1992). Interestingly 3 out of 13 of the re-isolated classical mutations occurred on chromosome 10. In another report of twenty five mapped transgene insertions five occurred on chromosome 10 and four occurred on chromosome 6 (Rijkers *et al.*, 1994). There is also strong evidence that retroviral integration is not random (discussed in Section 1.3.2 b). Such biases would make saturation mutagenesis hard to achieve.

1.2 Entrapment technology

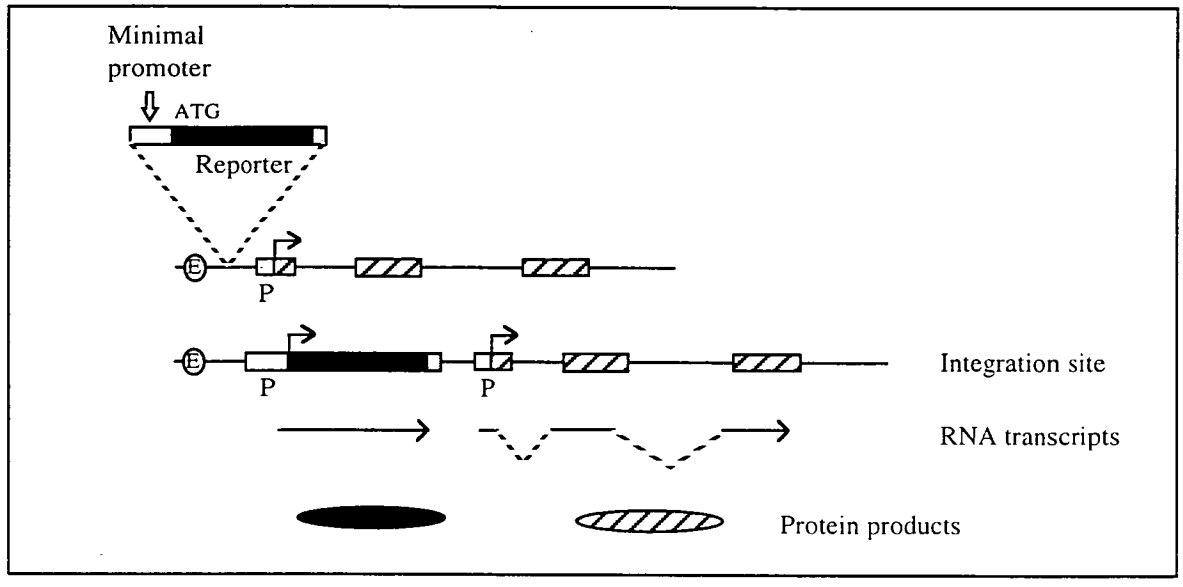
Although the methods for insertional mutagenesis described in the previous section allow the mutated locus to be cloned from genomic DNA, the identification of genes from flanking sequences remains time consuming. To allow rapid identification of mutated genes, entrapment vectors have been developed (reviewed in Skarnes, 1990). These vectors contain a reporter gene which can only be expressed if the vector integrates into a gene or close to cis-acting elements in the genome. Such vectors can create reporter fusion transcripts with endogenous genes allowing direct identification of the mutated locus thus alleviating the problem of searching for exons in large flanking genomic DNA.

Expression of some genes which are known to be important for specific developmental processes exquisitely reflect their function in those tissues. For example, at the onset of gastrulation *Brachyury (T)* is expressed in the primitive streak, mesoderm emerging from the streak, head process and notochord (Beddington, 1992). Mouse embryos that are homozygous for a *T* deletion die at mid-gestation with prominent defects in the notochord, the allantois and the primitive streak. These

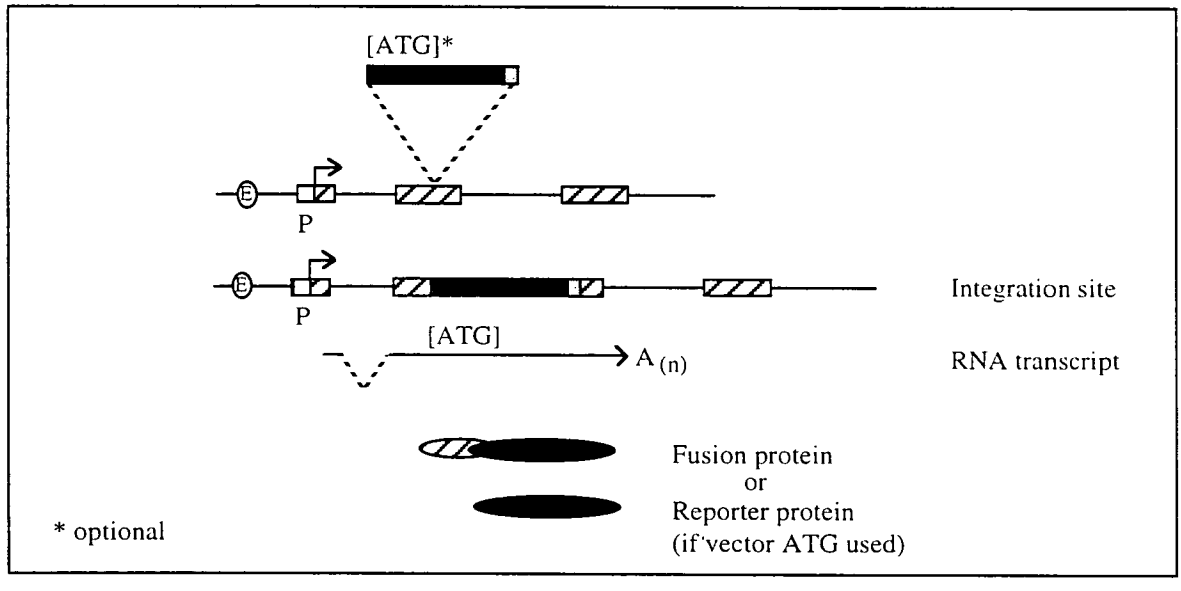
defects are thought to be due to defective cell mesoderm movements (Beddington *et al*, 1992). The presence of a reporter gene in entrapment vectors allows the selection of mutations in genes showing interesting spatial or temporal expression patterns during embryogenesis. Moreover, such screening is anticipated to lead to the identification of genes which would be missed in conventional mutagenesis screens. For example, if a gene is functionally redundant because of the presence of a second locus that can substitute for the same function, inactivation of the gene will not result in any phenotype (Cooke *et al.*, 1997). Such genes cannot be identified in screens for mutant phenotypes but may be detected by screening for restricted expression patterns.

Two categories of entrapment construct have been developed which differ in the elements required from the endogenous locus to produce reporter activity. Enhancer traps contain a minimal promoter driving a reporter gene (Figure 1.2A). The minimal promoter is silent unless the vector inserts near an endogenous enhancer element. Promoter traps contain no promoter elements of their own and rely on endogenous promoter elements for expression. This category can be sub-divided into three groups, exontraps, genetraps and polyadenylation A (poly A) traps (Figure 1.2B). Firstly, exon trap vectors simply contain a reporter gene and require integration into an exon in the correct orientation and reading frame to produce reporter activity (von Melchner, 1989; Friedrich and Soriano, 1991; von Melchner, 1989; Friedrich, 1991; Macleod *et al.*, 1991). Insertions can also be detected in non-translated exons if the reporter gene has its own ATG. Secondly, gene trap vectors contain a splice acceptor site upstream of a reporter so that integration into an intron creates a spliced fusion transcript which encodes a fusion protein with reporter activity when the reading frame is maintained (Gossler *et al.*, 1989; Friedrich and Soriano, 1991). Once again if the reporter includes an ATG codon insertions can be detected in untranslated regions. Theoretically gene trap insertions can also be detected in exons. Finally, poly A trap vectors are designed such that they need to trap an endogenous

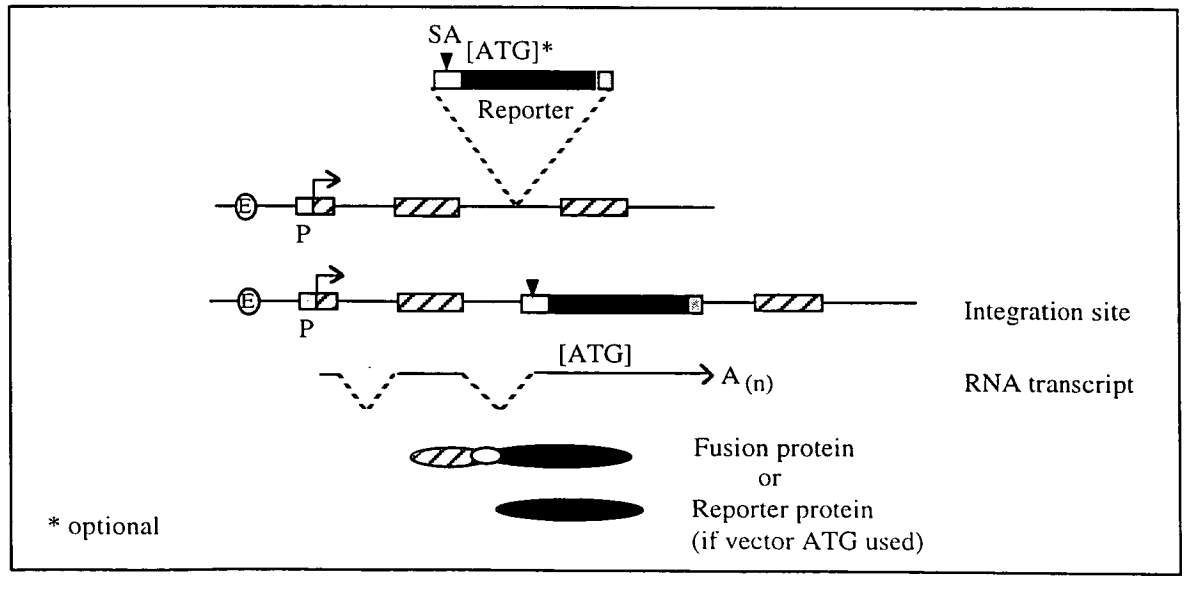
A. Enhancer trap



B (I). Exon trap vector



B (II). Gene trap vector



B (III). Poly A trap vector

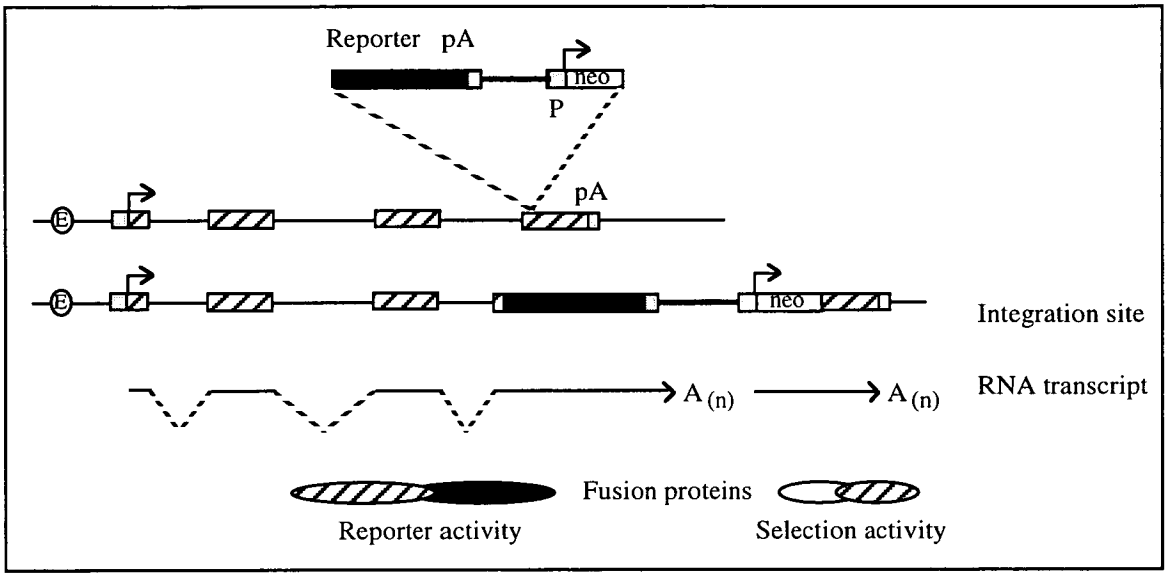


Figure 1.2 Structure and mode of action of different entrapment vectors.

There are two categories of entrapment vector which differ with respect to the modes of activation of the reporter gene. **(A)** Enhancer trap: contains a minimal promoter driving the reporter gene which has its own translation initiation start site. The promoter is silent unless the vector inserts near an endogenous enhancer element. **(B)** Promoter trap: relies on endogenous promoter elements for expression. There are three types. **I** Exontrap: inserts into an exon of an active transcription unit in the correct orientation and reading frame. If the reporter has its own translation initiation site insertions can be detected in coding exons in any reading frame and in non-translated exons. **II** Genetrap: a splice acceptor site is placed upstream of the reporter. Integration into an intron in the correct orientation and reading frame results in a primary reporter fusion transcript which is further processed. If the reporter has its own translation initiation site, insertions can be detected in non-translated regions. Insertions can also be detected in exons. On these occasions there will be competition between the endogenous splice acceptor and the introduced splice acceptor. **III** PolyA trap: vector needs to acquire a poly A addition signal to allow selectable drug resistance.

poly A signal for expression of the drug resistance selection gene (Niwa *et al.*, 1993; Yoshida *et al.*, 1995). The next three sections review the development and use of entrapment vectors in lower organisms before focusing on their use in mammals.

1.2.1 Bacteria

The prototypic promoter trap vectors were developed in bacteria to identify temporally controlled genes (Casabadan & Cohen, 1979; Bellofatto *et al.*, 1984). The vectors utilise the integration properties of the bacteriophage Mu genome. The coding regions of either an assayable enzyme such as βgalactosidase (βgal) or an antibiotic resistance gene were incorporated at the end of the Mu genome. After transposition to random sites in the genome the reporter gene came under the control of regulatory elements from cellular promoters close to the insertion site. Vectors were made in which the initiation codon of βgal was deleted thus forcing the expression of reporter activity to be driven via gene fusion (Casabadan *et al.*, 1980). The authors found that βgal could tolerate large N-terminal fusions whilst retaining enzyme activity.

1.2.2 Plants

Transposon-mediated mutagenesis has been adapted to trap cellular promoters and enhancers in tobacco and *Arabidopsis*. The first promoter trap screen utilised the tumour inducing (Ti) plasmid of *Agrobacterium* which is transferred and inserted into the host genome on infection (Koncz *et al.*, 1989). In both Arabidopsis and tobacco 25 % of transformants expressed the reporter gene. The similarity is unexpected as the density and distribution of transcribed sequences of both plants is very different (Koncz *et al.*, 1989). It suggests that Ti plasmid, rather than inserting into the genome in a random manner, preferentially integrates into sequences that can be potentially transcribed. Only one out of 450 transformants produced a morphologically visible mutant phenotype when bred to homozygosity.

In another more efficient gene trap approach the maize transposable elements *Activator (Ac)* and *Dissociation (Ds)* were used (Sundaresan *et al.*, 1995). A two-element transposon system was devised in which transgenic *Arabidopsis* plants carrying an immobilised *Ac* element were crossed to plants carrying a *Ds* gene trap element. The *Ds* element carries a βglucuronidase (GUS) reporter gene preceded by a splice acceptor site. As introns in plants are small the gene trap vector was designed to detect insertions in exons as well as introns. This was achieved by the presence of multiple splice donor sites at the end of the *Ds* element. When the Ds element inserts into an exon, the sequence between the splice donor site at the end of the element and the splice acceptor in front of the GUS gene, can be spliced out from the transcript resulting in a fusion transcript that encodes GUS activity. Of the transformed lines, 14% showed restricted expression patterns and 4 % had a visible mutant phenotype. One of insertions occurred in a novel gene, PRL, which is required for embryo development (Springer *et al.*, 1995). It was possible to confirm that insertion of the *Ds* element into PRL was responsible for the mutant phenotype by generating *Ds* revertants by crossing in *Ac* which effectively excises the Ds gene trap element (Springer *et al.*, 1995).

1.2.3 *Drosophila melanogaster*

Enhancer trapping has been particularly productive at detecting developmental control genes in *Drosophila* using the P-element system (reviewed in Spradling *et al.*, 1995). P-elements are transposons with terminal inverted repeats which transpose via DNA intermediates (reviewed in Engels, 1989). The first enhancer trap used a βgal reporter gene driven by a weak promoter placed within the P-element to detect cis-acting regulatory sequences (O'Kane and Gehring, 1987). Of 49 transformed fly lines, 70 % showed restricted reporter expression patterns. Further refinement to the method has allowed the generation of tens of thousands of lines with about 65 % showing spatially and temporally restricted reporter expression (O'Kane and Gehring,

1987; Bellen *et al.*, 1989; Bier *et al.*, 1989; Wilson *et al.*, 1989). As well as identifying valuable markers for tissues and organs, between 10 and 17 % of lines carrying P-elements insertions exhibit recessive lethal phenotypes (Cooley *et al.*, 1988; Bier *et al.*, 1989). Cloning the P-element insertion sites confirms that the reporter picks up genuine regulatory elements from neighbouring genes and importantly, insertions near known developmentally regulated genes display patterns of reporter expression similar or identical to those of the known genes (Wilson *et al.*, 1989).

The high frequency of reporter activation in P-element mutagenesis may reflect preferential integration into specific sites in the genome similar to that seen with retrovirus insertion in animal genomes (Jaenisch, 1988). In both cases insertions are generally found outside of heterochromatin, often in 5' untranslated regions. The more open chromatin confirmation of these areas is thought to make them more accessible to insertion. Sequence analysis of P-element integration sites has not revealed any biases, yet some genes represent hot spots for P-element insertions, whereas others are relatively refractory to insertion (Smith *et al.*, 1993) From 68 lethal lines which show an embryonic phenotype in the nervous system 5 of the P-element insertions occurred in *neuralized* and 4 in *master mind* (Bier *et al.*, 1989). Extensive analysis of a 1.8 megabase genomic region indicates that only one third of lethal loci which were identified by chemical mutagenesis can be functionally disrupted by P-elements. Another transposable element, *hobo*, has a different distribution of insertion hot spots (Smith *et al.*, 1993) . This suggests that a *hobo* enhancer trap strategy may allow access to genes resistant to P-element insertion .

1.3 Entrapment in mouse ES cells

Early experiments in vertebrates revealed that the position a transgene integrates in the genome can influence its temporal and spatial expression pattern (Jaenisch, 1981). This observation, along with studies in *Drosophila* (O'Kane and

Gehring, 1987), laid the foundations for the first enhancer trap experiments in the mouse. Position effects were explored using truncated versions of HSV thymidine kinase (Allen *et al.*, 1988) or the mouse hsp68 gene (Kothary *et al.*, 1988) which acted as weak promoters driving lacZ expression. Constructs randomly integrated into the genome following injection into the male pronucleus of fertilized mouse oocytes. One out of 7 hsp68lacZ transgenic lines and 5 out of 20 HSVtk lacZ transgenic lines exhibited restricted patterns of reporter activity during development. These results indicated that enhancer trap constructs could be used in mouse to identify expression patterns of endogenous genes and to isolate the genes. However, as mentioned previously, transgenic mice generated by zygote injection often suffer large chromosomal abnormalities associated with insertions which make identification of the interrupted gene difficult (Section 1.1.3). Such was the case for the hsp68lacZ transgenic line where a 45 kb deletion accompanied the integration. Seven years passed between identifying the insertion as an allele of the classical mouse mutant *dystonia musculorum (dt)* and cloning the disrupted gene (Brown *et al.*, 1995). As well as difficulties in cloning, another drawback making the transgenic approach unsuitable for large scale screening is that the expression of trapped genes cannot be assayed quickly, as this would require sacrificing valuable founder animals.

A more efficient entrapment strategy utilizes ES cell technology and allows preselection for integration events and reporter activity *in vitro* before producing mouse lines (Gossler *et al.*, 1989; Friedrich & Soriano, 1991). Mouse ES cells are derived by culturing the inner cell mass of pre-implantation mouse embryos under conditions which actively inhibit differentiation and/or promote stem cell self renewal (Evans & Kaufman, 1981; Martin, 1981). These cells, upon injection into host blastocysts, can colonize all embryonic lineages, including the germ line (Bradley *et al.*, 1984). These properties, along with their ability to be maintained indefinitely *in vitro,* make ES cells one of the most powerful tools in mammalian genetics. DNA can

be introduced into ES cells in a number of ways, including electroporation, lipofection and retroviral infection, without compromising their ability to contribute to all lineages of the developing mouse embryo (Robertson *et al.*, 1986; Thompson *et al.*, 1989). In this way large numbers of mutations can be introduced into ES cells and then transmitted through the germ line of mice so that phenotypic consequences of mutations can be assessed *in vivo*.

1.3.1 Enhancer traps

To overcome problems inherent with transgenic approaches, an enhancer trap vector was introduced into ES cells by electroporation and reporter activity was analysed in chimaeric embryos (Gossler *et al.*, 1989; Korn *et al.*, 1992). The construct consisted of a minimal hsp68 promoter driving the lacZ gene upstream of a neomycin gene under the control of the herpes simplex virus thymidine kinase promoter. Between 14 and 22 % of neomycin resistant lines exhibited lacZ expression in undifferentiated ES cells, and of these 40 % [6/15] showed distinct reporter gene expression patterns during early embryonic development (Gossler *et al.*, 1989; Korn *et al.*, 1992). Of the lines which showed no expression in undifferentiated ES cells 31% [16/51] gave rise to patterns during embryonic development. Many of the staining patterns were exclusively, or included in, the central nervous system. A similar skew towards CNS staining was seen in transgenic lines carrying a different design of enhancer trap (Allen *et al.*, 1988) and is also reminiscent of the enhancer trap patterns obtained in Drosophila (Bellen *et al.*, 1989). Together this data may reflect the large number of expressed genes in the developing nervous system.

Genomic sequences flanking the insertion sites have been cloned for two enhancer trap lines and cDNA clones representing novel sequences were isolated from them (Soininen *et al.*, 1992; Neuhaus *et al.*, 1994). In both cases expression of the endogenous genes was more widespread than the corresponding βgal activity,

suggesting that the reporter gene responded to only a subset of the regulatory elements of the endogenous gene. Both enhancer trap integrations were transmitted through the germline but neither showed an obvious phenotype when homozygous. The rate at which enhancer trap insertions yields mutant phenotypes is likely to be fairly low as the insertion is expected to leave the endogenous promoter and coding sequence intact (Table 1.1).

1.3.2 Promoter traps

Numerous experiments have confirmed the feasibility of using the promoter trap approach to identify and mutate novel genes involved in development. Firstly promoter traps have been found to mutate the endogenous gene at the site of integration leading to a lethal phenotype in 37 % [23/63] of lines analysed (Friedrich & Soriano, 1991; Skarnes *et al.*, 1992; Camus *et al.*, 1996; Forrester *et al.*, 1996; Hicks *et al.*, 1997). This percentage of lethal loci is in keeping with the 1 in 3 loci essential for viability in yeast, worms and flies (Miklos and Rubin, 1996). Secondly the endogenous genes carrying insertions have been easily cloned from fusion transcripts or flanking genomic DNA (Table 1.2). Finally the expression patterns of reporter genes have been found to be similar to endogenous genes so that screens based on reporter activity can be used to pre-select for insertions which may play a role in development (Skarnes *et al.*, 1992; Skarnes *et al.*, 1995; Chen *et al.*, 1996; Torres *et al.*, 1997). The various promoter trap designs which have been used in ES cells are summarised in Appendix I and are discussed in the following sections.

(a) Vector design

In mice, exon trap vectors contain a promoterless reporter gene and are equivalent to the prototypical bacterial promoter traps (Casabadan & Cohen, 1979; Bellofatto *et al.*, 1984). Reporter activity results from insertion of the vector into an exon in the correct orientation and reading frame (Friedrich & Soriano, 1991). Rather

Table 1.1 Comparison of entrapment vector properties

		Promoter trap		
	Enhancer trap	Exon trap	Gene trap	Poly A trap
Expected target in genome	Close to cis-acting elements able to activate minimal promoter	Coding exon (if reporter has ATG then UTRs also targets)	Intron (or exon) of coding region (if reporter has ATG then UTRs also tragets)	Intron (or exon) upstream of poly A addition signal
Transcript	Defined start at vector promoter	Fusion transcript between endogenous gene and reporter	Fusion transcript between endogenous gene and reporter	Fusion transcripts between 5' endogenous gene and reporter and between 3'endogenous and selector gene
General route of gene identification	Often difficult	Simple	Simple	Simple
	Clone gDNA flanking insertion, screen genomic library with product and try to identify transcribed sequences	Plasmid rescue of gDNA should contain coding sequences or 5'RACE	5'RACE	5'RACE or 3'RACE
Mutagenic	Possible	Very probable	Very probable	Probable
Requirement for endogenous gene expression	No	Yes	Yes	No

Table 1.2 Table of characterised mutations identified by entrapment

Vector	Insertion site	Gene	Gene function	βgal activity[a]	Homozygous phenotype	Reference
Retroviral promoter trap						
ROSAβgeo	5'UTR	TEF-1	transcription factor	widespread	embryonic lethal	(Chen *et al.*, 1994)
"	intron	BTF-3	transcription factor	restricted*	postimplantation lethal	(Deng and Behringer, 1995)
U3βgeo	5'UTR	Eck	receptor tyrosine kinase	restricted	none	(Chen *et al.*, 1996)
"	?	ArMT	arginine methyltransferase	N.D.	embryonic lethal	(Scherer *et al.*, 1996)
U3neo	5'UTR	fug1	unknown	N.A.	embryoinc lethal	(Degregori *et al.*, 1994)
Plasmid promoter trap						
pGT4.5	intron	Gt4-1	unknown	widespread	perinatal lethal	(Skarnes *et al.*, 1992)
"	intron	Gt4-2	zinc finger trancsription factor	restricted	lethal[b]	"
"	intron	Gt10	unknown (proline-rich)	restricted	N.D.	"
"	?	Gt2	N.D.	widespread	none	"
"	intron	Cordonbleu	unknown	restricted	none	(Gasca *et al.*, 1995)
pT1-ATG	intron	I.23	N.D.	restricted	none	(Forrester *et al.*, 1996)
"	intron	I.114	unknown	restricted	none	"
"	intron	I.163	unknown	restricted	none	"
"	intron	I.193	unknown	restricted	none	"
"	intron	R.140	unknown	restricted	embryonic lethal	"
pGT1.8TM	intron	Netrin	secreted axon guidance cue	restricted	postnatal lethal	(Skarnes *et al.*, 1995)
"	intron	PTPκ	protein tyrosine phosphatase	restricted	none	"
"	intron	LAR	protein tyrosine phosphatase	restricted	none	"
pGT1.8βgeo	intron	E-catenin	cell-cell adhesion molecule	restricted	preimplantation lethal	(Torres *et al.*, 1997)
TV2	intron	Jumonji	unknown	restricted	embryonic lethal	(Takeuchi *et al.*, 1995)
Plasmid Poly A trap						
pPAT	intron	PAT-12	unknown	N.D.	none	(Yoshida *et al.*, 1995)

* endogenous gene expression is not restricted
RBP-2 retinoblastoma-binding protein
N.D. not determined
N.A. not applicable as vector does not contain a reporter gene
a βgal activity during development
b variable penetrance, some homozygotes viable with growth defect

confusingly this vector design is sometimes referred to simply as a 'promoter trap', this nomenclature is based on the observation that retroviral exon traps tend to insert within 500 nucleotides of cellular promoters. Gene trap vectors contain a splice acceptor site upstream of the reporter gene (Gossler *et al.*, 1989; Friedrich & Soriano, 1991). Insertion into an intron of an active transcription unit leads to a fusion transcript between the endogenous gene and lacZ (Skarnes *et al.*, 1992).

Promoter trap insertions into the same gene have been found on several occasions. For example *L29* and *α-NAC* were both disrupted three times in an analysis of 400 insertions (Chowdhury *et al.*, 1997) and the protein tyrosine phosphatase *Lar* was disrupted twice in a sample of only six insertions (Skarnes *et al.*, 1995). There are a number of factors which could potentially influence the position a vector inserts into the genome. These include the presence of integration hot spots, the intron/exon structure of a gene and homology between the vector and areas of the genome which would increase the likelihood of homologous recombination. Genomic hotspots have been discussed in Section 1.2 and include regions with open chromatin structure which may offer easy access to vector insertion. The intron/exon structure of a gene will also influence its ability to be detected. Transcription units with large intronic targets are more likely to be targets for gene trap insertion.

The reading frame of the endogenous splice donor must be compatible with the vector splice acceptor for reporter expression to be detected. There are two strategies to overcome reading frame constraints. The first uses the MoMuLV splice acceptor from the *env* gene which splices in all three reading frames (Schuster-Gossler *et al.*, 1994). Alternatively, including an internal ribosomal entry site (IRES) between splice acceptor and the reporter allows the reporter to be independently translated from the IRES irrespective of the reading frame of the transcript (Chowdhury *et al.*, 1997). The number of $G418^R$ colonies increased three fold when an IRES sequence was added to a SAβgeo gene trap vector (Chowdhury *et al.*, 1997).

The final issue of vector homology is contentious as there is no hard evidence to suggest it exists. The repeat insertions into the same genes have been found using different vector designs. For example *jumonji* was originally identified using pTV2 (Takeuchi *et al.*, 1995) and has since been disrupted by pGT1.8TM (W.C. Skarnes, unpublished observations) and ROSAβgeo (Baker *et al.*, 1997). It is possible that there is some degree of homology between the neo or β-gal sequences and the *jumonji* locus, particularly as the two plasmid vectors inserted into the same *jumonji* intron. In other repeat events the insertions occurred at different places in the gene (Skarnes *et al*, 1995; Chowdhury *et al*, 1997; W.C. Skarnes, personal communication).

I Selectable markers

To allow direct selection of promoter trap events a reporter gene was developed in which the ORFs of lacZ and neomycin were fused in the same reading frame (Friedrich & Soriano, 1991). The encoded protein, βgeo, exhibits both βgal reporter activity and neo selection activity. Therefore all $G418^R$ colonies should express the βgal protein. In the original promoter trap experiment using βgeo 95 % of G418 resistant pSAβgeo colonies showed βgal activity (Friedrich & Soriano, 1991). In a subsequent experiment (using pGT1.8βgeo) only 60 % of G418 resistant colonies showed βgal activity (Skarnes *et al.*, 1995). These colonies showed a variety of βgal staining intensities in contrast to the uniform high levels of staining seen for pSAβgeo colonies. These anomalies were explained by the presence of a point mutation in the *neo* of pSAβgeo which is known to reduce its enzyme activity (Yenofsky *et al.*, 1990). Consequently, only pSAβgeo colonies expressing high levels of βgeo can survive the selective pressure of G418. The different levels of pGT1.8βgeo staining patterns reflect the highly sensitive nature of its *neo,* such that low levels of expression are sufficient to confer drug resistance, yet much higher levels of *lacZ* expression are required to detect βgal activity. The requirement for high levels of expression may also explain why far fewer G418 colonies were recovered with pSAβgeo compared to

pGT1.8βgeo (Skarnes *et al.*, 1995). This increased sensitivity allows pGT1.8βgeo to detect insertions into genes expressed at very low levels in ES cells which may not be possible with pSAβgeo.

II Poly A traps

Novel promoter trap vectors called poly A traps have been designed to allow the identification of genes which are not expressed in ES cells by virtue of the selective gene within the vector being driven by its own promoter (Niwa *et al.*, 1993; Yoshida *et al.*, 1995). It is desirable to access insertions in this class of molecule as it is clear that some developmentally essential genes such as *Wnt-1* and *En-1* are not expressed in ES cells (McMahon & Bradley, 1990; Jeanotte *et al.*, 1991). The first poly A trap vector consisted of a promoterless lacZ gene preceded by a splice acceptor with *neo* driven by its own promoter but lacking a poly A addition signal (Niwa *et al.*, 1993) and Appendix I [26 & 25]. Insertions were required to trap an endogenous poly A signal for expression of *neo* selection. As neo is driven by the β-actin promoter which is active in ES cells, endogenous genes can be tagged with lacZ even if they are not active in undifferentiated ES cells. Unfortunately, problems arose due to deletions from the ends of the electroporated vector which removed the splice acceptor site so that in 4 out of 5 cases the vector had inserted into the promoter region of the endogenous genes. The 3' end of neo fusion transcripts were amplified by 3' RACE and in three out of five lines the transcript had incorporated a poly A tail but curiously there was no poly A addition signal found in the sequences between neo and the poly A stretch. Products could not be amplified using 3' RACE in the other two lines. The authors postulated that on these occasions *neo* may be fused to the exon of a gene leading to the production of long mRNAs which could not be amplified.

Another group further refined the poly A trap design by including a splice donor site downstream of the stop codon of *neo* (Yoshida *et al.*, 1995) and Appendix

I [29]. This modification allows splicing of neo to downstream endogenous exons. Around 25 % of lines were successfully amplified with 3' RACE and in 5 out of 6 fusion transcripts poly A sites were identified within novel endogenous sequences. The full length cDNA was isolated for a line called PAT-12 and was confirmed by southern and northern blot analysis to represent the endogenous gene carrying the insertion. Homozygous mice were generated but no overt phenotype was apparent. The poly A trap insertion occurred at the 3' end of the gene which raised the question whether the gene had been successfully disrupted. No transcripts were found in homozygotes so the authors postulate that the vector insertion possibly inhibits transcription or destabilises mRNA. Unexpectedly the endogenous PAT-12 locus is expressed in undifferentiated ES cells yet the reporter gene in PAT-12 ES cells is inactive. Although there is only minimal data regarding poly A trap vectors, they may be particularly useful for *in vitro* induction screens where genes are selected on the basis of their upregulation in differentiated cells in response to chemical or growth factor stimulation (Section 1.3.4 d).

III Conditional promoter trapping

Another novel entrapment strategy has been the incorporation of site specific recombination technology into a promoter trap screen (Russ *et al.*, 1996). The authors developed a strategy to isolate genes that are transcriptionally activated during programmed cell death (summarised in Figure 1.3). A retroviral trap vector is used to bring the Cre recombinase under the control of cellular promoters. The vector was introduced into interleukin-3 (IL-3) dependent haematopoietic cells expressing a reporter plasmid that carries the PGK promoter driving expression of herpes simplex virus thymidine kinase, *neo* and IL-3. The *neo* gene is flanked by loxP sites and cells are *neo* resistant and require the addition of exogenous IL-3 for survival. Activation of Cre expression from promoter trap integrations within active genes results in Cre mediated removal of neo activity and brings IL-3 under the control of the PGK

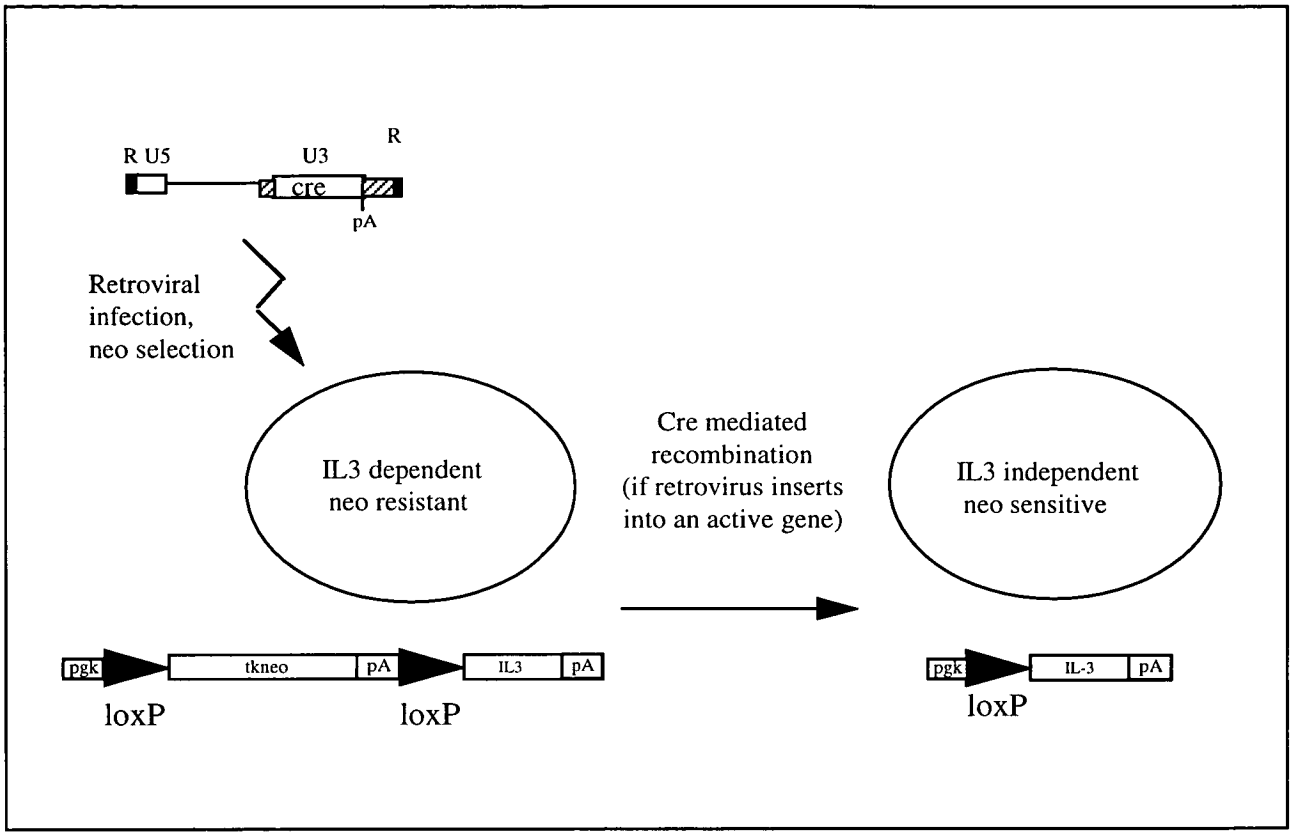


Figure 1.3 Recovery of apoptosis genes using gene-trap mutagenesis and site-specific recombination

Strategy used to identify genes that are transcriptionally activated during programmed cell death. Interleukin-3 (IL3) dependent haematopoietic cells (FDCP1) expressing a reporter plasmid that encodes herpes simplex virus thymidine kinase, *neo* and murine IL-3 are transduced with a retroviral gene trap vector carrying coding sequences for cre recombinase (Cre). Activation of Cre expression from promoter trap integrations within active genes results in Cre mediated removal of neo activity and production of IL-3 which converts FDCP1 cells to factor independence. Because the expression of the marker genes is independent of the trapped cellular promoter, insertions into genes can be identified that are transiently induced by IL-3 withdrawal.

promoter, thus converting cells to ganciclovir resistant and IL-3 independent state. Because the expression of the marker genes is independent of the trapped cellular promoter, insertions into genes could be identified that were transiently induced by IL-3 withdrawal. The authors found that 1.7 x 10^{-4} insertions detected a promoter induced by IL-3 withdrawal. 50 % of the analysed fusion transcripts were also inducible by serum starvation, suggesting that the disrupted genes are associated with an apoptotic program or with growth arrest. This type of strategy could be used to identify transcriptional targets in many different signaling pathways. An alternative use for the site-specific recombination system would be to add loxP sites to a gene trap vector. This will allow the endogenous locus carrying the insertion to be manipulated. In this way it might be possible to produce revertants which would be useful in confirming that insertions were responsible for particular phenotypes.

(b) Vector delivery

Promoter traps have been delivered into ES cells by electroporation and retroviral infection (Gossler *et al.*, 1989; Friedrich & Soriano, 1991; von Melchner, 1992). Whilst retroviruses clearly provide an efficient method for introducing mutations in the genome without gross DNA rearrangement there are potential drawbacks to their use in promoter trapping. As alluded to earlier a concern is whether retroviral insertion is random. Three fold more G418 resistant colonies showed βgal staining when a retroviral promoter trap was compared to the same ROSAβgal vector design electroporated (Friedrich & Soriano, 1991). This increase is thought to reflect specificity of proviral integration sites. Other studies have suggested that DNase I hypersensitive sites (Vijaya *et al.*, 1986; Rohdewohld *et al.*, 1987) and actively transcribed regions of chromatin (Mooslehner *et al.*, 1990; Scherdin *et al.*, 1990) are preferred sites within the genome for retroviral integration. It is clear that in some of these experimental systems a selection bias could have been introduced. Proviral insertions in tumours, and cell lines derived from them, may be selected for relatively

subtle advantages to cell growth (Vijaya *et al.*, 1986; Rohdewohld *et al.*, 1987). A more recent report indicates that most regions of the genome are accessible to avian retroviral integration although a few sites are used up to 280 times the random frequency (Withers-Ward *et al.*, 1994). However, the same group previously showed that approximately 20% of all integrations involve highly preferred sites (Shih *et al.*, 1988). The authors suggest that some unknown variable in cell culture conditions or in the target cells used could account for the contradictory results. Although controversial, the presence of integration hot spots would pose a problem when undertaking saturation scale screens.

Another concern with retroviruses is the ability of the proviral reporter gene to accurately reflect the expression pattern of the endogenous gene carrying the insertion. Repression of transcription due to *de novo* methylation of the provirus which extends into flanking promoter sequences has been shown to occur in several retroviral transgenic strains (Jahner & Jaenisch, 1985). One of three published characterisations of retroviral gene trap lines describes an insertion of ROSAβgeo into the BTF3 transcription factor (Deng & Behringer, 1995). The endogenous BTF3 gene is ubiquitously expressed but the expression of the βgal reporter in BTF3 heterozygotes is restricted.

A second report characterised an integration in the Eck receptor tyrosine kinase by the U3βgeo retrovirus (Chen *et al.*, 1996). Although the retroviral vector is designed as an exon trap, the insertion occurred in an intron 8kb upstream of the 5' end of the published cDNA sequence, and 1.8 kb downstream of an alternatively spliced 5' exon. As the construct does not have a splice acceptor it is expected to be spliced from the primary fusion transcript along with the intron into which it has inserted. This does not appear to happen as adult mutant mice were shown to be severely deficient in Eck protein. As the transcripts were not analysed it remains to be

seen whether the protein deficiency is due to the insertion truncating the endogenous message or whether the provirus has silenced the Eck promoter.

The final retroviral gene trap report describes an insertion of ROSAβgeo in the transcriptional enhancer factor 1 (TEF-1) (Chen *et al.*, 1994). On this occasion βgal expression appears to accurately reflect the widespread expression seen with the human homologue of TEF-1. There are no reports of discrepancies in reporter expression with electroporated gene trap constructs suggesting that screens based on reporter expression may be more reliable using electroporated constructs than their retroviral counterparts.

(c) Mutagenicity

To work as an effective mutagen it is important that expression of the endogenous transcript is ablated at the locus carrying the promoter trap insertion. In six lines homozygous for gene trap insertions endogenous expression was absent or negligible as detected by Northern or RNase protection (Skarnes *et al.*, 1992; Takeuchi *et al.*, 1995; Skarnes *et al.*, 1995). Northern analysis of a different promoter trap line showed that endogenous transcript was present in homozygotes (Gasca *et al.*, 1995). The levels indicated that the gene trap insertion was spliced out of about 10 % of primary transcripts. The interrupted gene in this line was *cordon-bleu* and although it exhibited an exquisite spatially restricted expression pattern there was no obvious homozygous phenotype. The authors suggested that the low levels of wild type transcript could rescue any potential phenotype. In another gene trap line low levels of wild type transcript were seen in homozygotes carrying an insertion into the axon guidance molecule *netrin* (Serafini *et al.*, 1996). However these homozygotes die at birth and exhibit defects in spinal commisural axon projections.

1.3.4 Screens for genes of interest

The major benefit of creating insertional mutations in ES cells by entrapment techniques is the ability to pre-select for insertions into genes of interest before the costly and time consuming production of germ line transmitting chimeras and subsequent breeding programs to look for phenotypes. ES cells lines can be selected on the basis of criteria which are discussed in detail in the next sections: sequence of disrupted gene, reporter gene expression during development and reporter gene expression in response to ES cell differentiation or other stimuli *in vitro*.

(a) Sequence

Genes disrupted by gene and poly A trap vectors can be identified from fusion transcripts by rapid amplification of cDNA ends (RACE). Genes carrying exon trap insertions can be identified using plasmid rescue of genomic DNA flanking the insertion site. To further facilitate the process of identification a solid-phase method to directly sequence 5'RACE products has been developed which eliminates the need for cloning (Townley *et al.*, 1997). In the last year three groups have assessed the feasibility of large scale identification of endogenous genes carrying promoter trap insertions (Chowdhury *et al.*, 1997; Hicks *et al.*, 1997; Townley *et al.*, 1997). Their results are summarised in Table 1.3.

The identities of the previously cloned gene traps indicate that many classes of molecule are amenable to promoter trapping. Over 50 % of the trapped sequences represent insertions in novel genes and so provide a resource of novel mutations which are available for *in vivo* functional analysis. The sequence information allows lines to be selected on the basis of structural homology to known or novel gene families.

Table 1.3 Summary of genes disrupted by large scale promoter trap screens

Vector	pSAβgeo & pSAIRESβgeo (Chowdhury *et al*, 1997)	pGT1.8TM (Townley *et al*, 1997)	U3neoSV1† (Hicks *et al*, 1997)
Known genes	17 (30%)	29 (51%)	42 (11%)
Homology to ests	11 (20%)	11 (19%)	21 (5%)
Novel	28 (50%)	17 (30%)	337 (84%)*

* The figures for the retroviral U3neoSV1 originate from genomic sequences flanking the insertion site compared to the database. As a consequence it is likely that database matches will be missed because flanking DNA often lacks sufficient exon sequences to generate statistically significant matches particularly if insertions occur near promoters or splice acceptor sites or further within an intron.

† Vector does not contain a reporter gene

(b) Reporter gene expression pattern in embryos

Since lacZ activity has been shown to reflect endogenous gene activity in all electroporated vector insertions (Section 1.3.2), screening for reporter activity in chimaeric embryos can be a useful approach to identify candidate genes which function at particular developmental stages or during particular developmental processes. A large scale screen for insertional mutations into genes which are developmentally regulated during mouse development has recently been undertaken using pPT-1 (Wurst *et al.*, 1995). Reporter expression was analysed in 8.5 dpc chimeras and approximately 33% showed widespread expression whilst 13 % [35/279] exhibited spatially restricted expression patterns. Of the lines which showed no expression at 8.5d more than 33% were found to express lacZ at 12.5 dpc. Therefore in total, roughly 66% of genes expressed in ES cells display developmentally regulated patterns of expression in embryos. The high percentage of restricted expression patterns clearly makes this type of screen an attractive approach to take.

Although the pattern-hunting approach has identified genes important for development (Table 1.2) not all such genes can be identified this way. Genes which are widely expressed but play important developmental roles in specific regions or cell types would be missed in such a screen. For example *notch* plays an important role in the *Drosophila* embryo by allowing cells in the ventrolateral region of the embryo to make the choice between assuming epidermal or neuronal fate yet its RNA does not have a restricted expression pattern in the prospective neurectoderm (Hartenstein *et al.*, 1994).

(c) Reporter gene expression pattern *in vitro*

In vitro strategies have been developed to pre-select for insertions which express the reporter gene in selected cell types or in response to inducers of differentiation. These approaches utilise the ability of aggregates of ES cells grown in suspension, called embryoid bodies, to give rise to a number of embryonic cell types (Keller, 1995). Co-localisation of cell-specific antibodies with βgal expression has been used to pre-select for insertions in genes expressed in neurons, glia, myocytes and chondrocytes (Baker *et al.*, 1997). To confirm that the regulation of gene expression in cultured embryoid bodies reflects expression during early development, changes in reporter gene expression *in vitro* were analysed in 191 lines carrying U3βgeo insertions (Scherer *et al.*, 1996). 16 % exhibited alterations in βgal expression upon *in vitro* differentiation. Seven of these lines were passed into the germline and the expression during development was consistent with that seen *in vitro.*

Another screen was designed to select integrations into genes which respond to retinoic acid (RA) (Forrester *et al.*, 1996). Intriguingly, this selection was found to significantly enrich for integrations in genes which show restricted expression patterns during development. From 3600 integrations 20 lines showed a repression or induction of βgal activity in response to retinoic acid. Of the 9 induced lines analysed 56% displayed a restricted pattern of expression at 8.5d. In a previous screen where lines were selected solely on the basis of lacZ expression in undifferentiated ES cells over 4 fold less lines showed restricted expression at 8.5d (Wurst *et al.*, 1995). In addition, eight of the RA-repressed integrations were expressed in the developing heart. This may reflect the proposed role of RA responsive genes in heart development, inferred from cardiac defects in RA-treated and RA-deprived embryos (Kirby, 1993). Similar types of pre-screens using growth factors, morphogens and

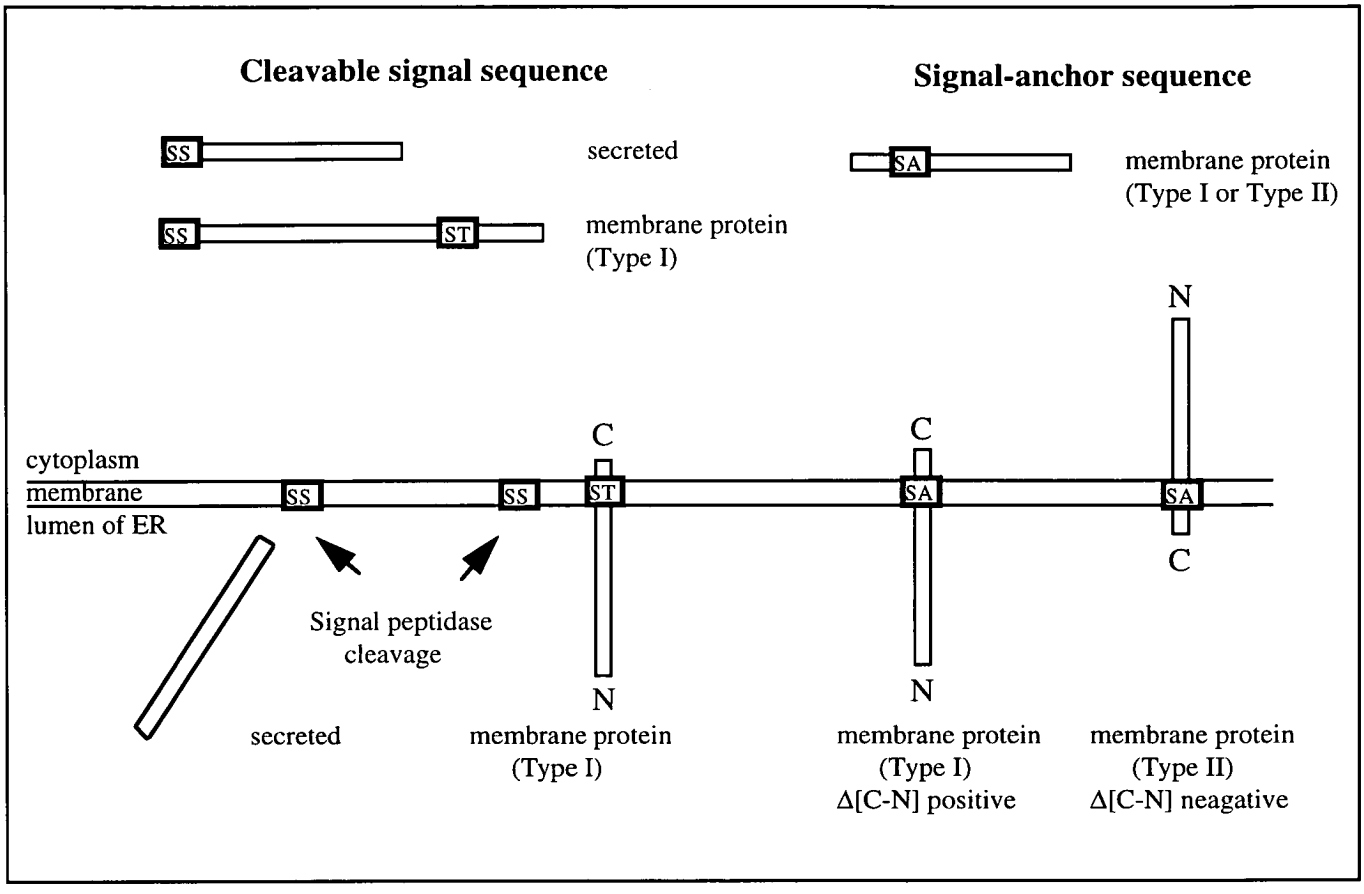


Figure 1.4 Two types of signal sequence target secretory proteins to the endoplasmic reticulum.

The two types of signal sequence that target secretory proteins to the endoplasmic reticulum (ER) and determine their orientation within the membrane are the cleavable signal sequence (SS) and signal anchor (SA) sequence. All proteins encoding a SS are either secreted or if they have a stop-transfer (ST) sequence are placed in the membrane in a type I orientation. Such type I proteins have their amino-termini in the lumen of the ER while type II proteins have their amino-termini in the cytoplasm. The membrane orientation of proteins encoding SA sequences depends on the difference in charge between the 15 amino acids flanking the carboxy-terminal end of the signal anchor and the 15 amino acids flanking the amino-terminal end (Δ[C-N]). If Δ[C-N] is positive the membrane protein is Type I, if Δ[C-N] is negative then the membrane protein is Type II. Figure adapted from High, 1992 and Hartmann *et al*, 1989.

growth substrates could be used to identify genes involved in other developmental programs.

1.3.5 Secretory trapping

Cell-cell interactions play a crucial role in the determination of cell fate within the early embryo, and in the differentiation of specific cell types during embryogenesis and adult life. Many families of signalling molecules have been shown to mediate these interactions including the BMPs (Hogan, 1996), the Wnts (Moon, 1997), homologues of the notch receptor and its ligands (Robey, 1997) and the hedgehogs (Hammerschmidt *et al.*, 1997). This section introduces the methods that have been used to identify novel receptors and ligands which mediate cell-cell interactions during development.

(a) Secretory pathway

To perform their function, secreted and transmembrane bound molecules must be trafficked to the extracellular environment through the secretory pathway (reviewed in Teasdale and Jackson, 1996). This pathway defines the route from the endoplasmic reticulum (ER) to the plasma membrane via the Golgi apparatus. The initial targeting of newly synthesized proteins to the ER occurs via a linear stretch of 7-20 apolar residues called the signal sequence. There are two types of signal sequence: cleavable signal sequences and uncleaved signal-anchor sequences (Figure 1.4). Cleavable sequences are found on both membrane and secreted proteins. They are present at the amino terminus of the nascent chain and their function is to target the protein to the ER and then initiate its translocation across the membrane. During the translocation process these signals are cleaved from the protein by the signal peptidase complex present at the luminel side of the ER membrane. A stop-transfer sequence on the carboxy terminal side of the signal sequence aborts translocation of the polypeptide so the carboxy terminus remains in the cytoplasm (type I orientation). In the absence of

the stop-transfer sequences the entire protein is translocated across the membrane into the lumen of the ER.

Membrane proteins with an uncleaved signal-anchor can expose either the carboxy (type I) or amino (type II) terminal to the cytoplasm. The difference in charges of the first 15 amino acids flanking the first signal anchor determines its orientation, with the more positive portion facing the cytosol (Figure 1.4; Hartmann, 1989). In proteins that span the membrane more than once, the orientation of all subsequent transmembrane segments are determined by that of the most amino terminal one. Membrane proteins that have a cleavable signal sequence always have a type I orientation. Throughout this thesis the term secretory molecule will refer to any protein which contains either a cleavable signal sequence or an uncleaved signal-anchor sequence.

(b) Cloning secretory molecules

Two expression cloning strategies have been devised to identify secreted and type I membrane proteins by exploiting the property of the signal sequence to direct secretory molecules to the ER (Tashiro *et al.*, 1993; Klein *et al.*, 1996). In the first example an expression vector containing the extracellular and transmembrane domains of interleukin-2 (IL-2) allowed cell-surface expression of fusion proteins when cDNAs encoding signal sequences are cloned into it. Fusion proteins expressed on the plasma membrane of cells can then be detected by antibodies to IL-2. Screening of 600 colonies from a bone marrow stromal cell line cDNA library identified six positive colonies each of which was found to encode a signal sequence. However, only one was novel. This approach is both technically demanding and labour intensive and not amenable to large scale analysis.

A more efficient method to identify genes containing signal sequences allows the simultaneous identification of many individual cDNAs in pools of a million clones

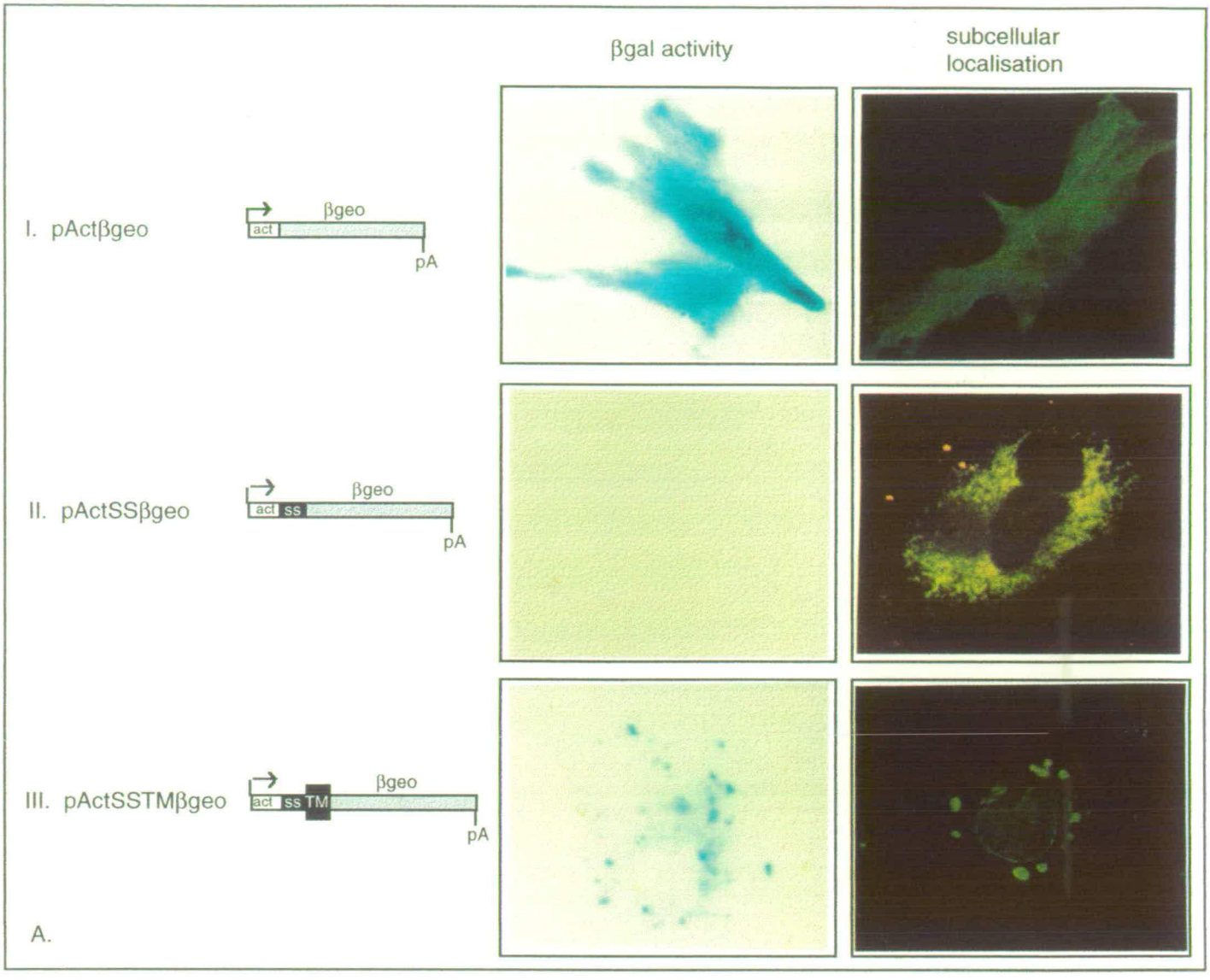
or more. The strategy is based on the ability of cloned cDNAs to provide a truncated secreted enzyme with a functional signal peptide (Klein *et al.*, 1996). The system utilises yeast invertase as secretion of this enzyme is essential for survival on a sucrose medium. cDNAs are cloned upstream of the invertase gene without a signal sequence. Only transformants which have a signal sequence are able to grow on a sucrose medium. cDNA libraries from a variety of embryonic mRNAs were transformed into yeast and from about 2 X 10^7 transformants 583 survived on sucrose. 161 cDNAs represent unique genes and at least 86 % are secretory molecules of which around 50% are novel. Although this is obviously a highly efficient strategy to identify novel secretory molecules there are drawbacks. It is potentially biased in favour of abundantly transcribed mRNAs and it provides no functional information on the identified genes. More importantly, it will not identify secretory molecules with uncleaved signal sequences.

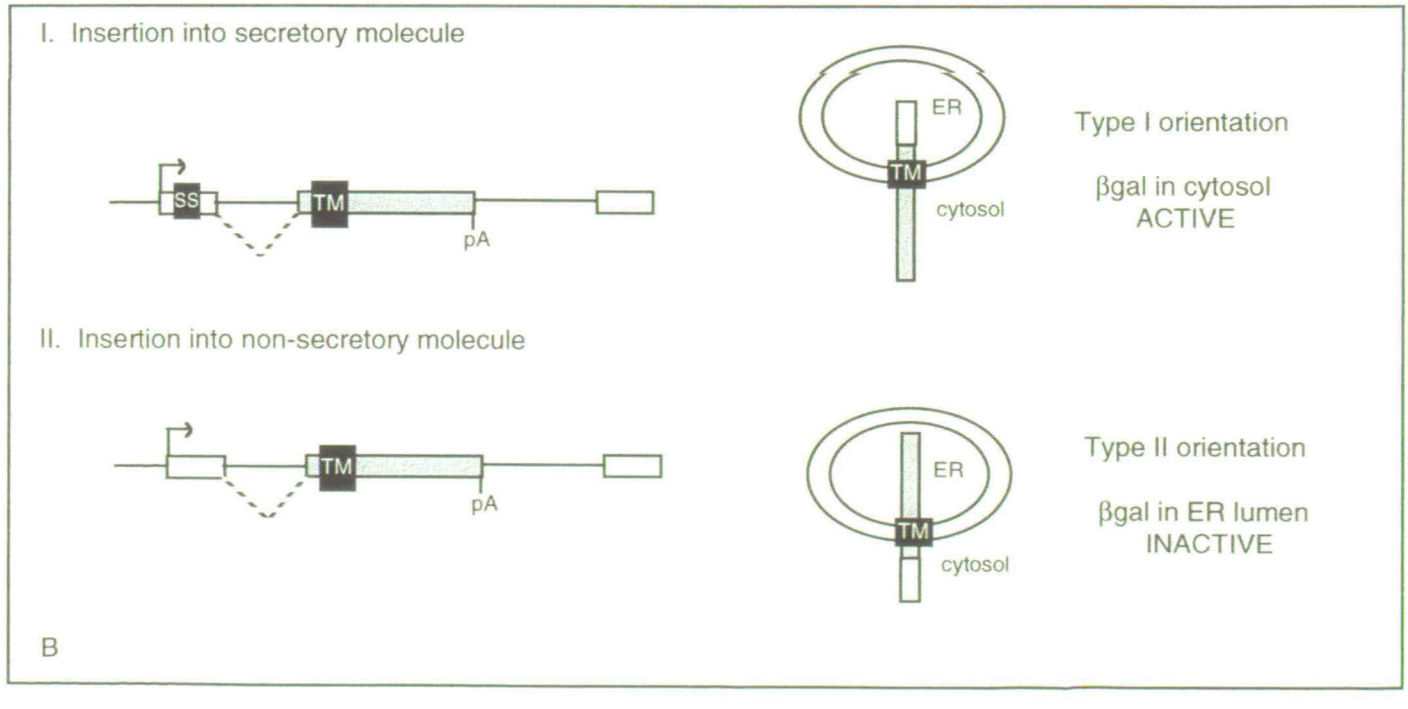
(c) Secretory trap vector

The protein sorting principle of the signal sequence has also been exploited in a gene trap approach to select for insertions into secretory molecules (Skarnes *et al.*, 1995). Unlike the expression cloning approaches, this strategy has the added advantage of allowing a functional analysis of the trapped genes. The rationale behind the approach is based on observations that βgal activity can be detected in a variety of subcellular locations in promoter trap lines (Gossler *et al.*, 1989; Wurst *et al.*, 1995). As βgal is normally distributed throughout the cytoplasm, the localisation to different subcellular structures presumably reflects the acquisition of endogenous protein domains that act as sorting signals. Experiments were performed to investigate whether insertions in secretory molecules could be identified by the subcellular localisation of βgal activity. Surprisingly, βgal activity is not detected if it is fused to a signal sequence Figure 1.5A; (Skarnes *et al.*, 1995). Immunofluorescence using βgal antibodies showed that such fusion proteins are translocated to the lumen of the ER

Figure 1.5 Insertions into genes encoding an N-terminal signal sequence are detected by the secretory trap vector

(A) Localisation of enzyme activity and βgeo fusion protein in ES cells transiently transfected with CD4-βgeo expression constructs. **I** The human β-actin (act) promoter was used to drive βgeo alone or **II**, fused in frame with the SS of CD4 , or **III**, fused to the SS and TM of CD4. Enzyme activity was assayed using X-gal as a substrate and photographed under brightfield illumination. The subcellular localisation of βgeo protein was determined by immunoflorescence and photographed using a fluorescein epifluorescence filter set. **(B)** Model to explain the selective activation of βgal fusion proteins using the secretory trap vector. **I** Insertion of pGT1.8TM in genes that contain a signal sequence place βgeo in a type I orientation in the ER membrane. The TM of the vector retains βgeo in the cytosol where it remains active. **II** Insertion of the vector into genes which do not contain a signal sequence result in a fusion proteins with an internal TM domain. These proteins are inserted in a type II orientation in the ER membrane resulting in the βgeo being sequestered in the lumen of the ER where βgal activity is lost.
Figures adapted from Skarnes *et al*, 1995.


βgal activity
subcellular localisation
I. pActβgeo
βgeo
act
pA
II. pActSSβgeo
βgeo
act
ss
pA
III. pActSSTMβgeo
βgeo
act
ss
TM
pA
A.


I. Insertion into secretory molecule
SS
TM
pA
ER
TM
cytosol
Type I orientation
βgal in cytosol
ACTIVE
II. Insertion into non-secretory molecule
TM
pA
ER
TM
cytosol
Type II orientation
βgal in ER lumen
INACTIVE
B

where they are inactive. However, addition of a transmembrane (TM) domain downstream of the signal sequence restores βgal activity presumably by placing βgal in the cytosol.

These results suggested that conventional gene trap vector insertions which occur downstream of a signal sequence but upstream of a TM domain would not be detected as βgal would be inactive. To try to recover such insertions in secretory molecules a TM domain was added to the gene trap vector to make the secretory trap vector pGT1.8TM. Cloning data from the first six lines generated using pGT1.8TM supports the prediction that βgal activity is dependent on acquiring a signal sequence from the endogenous gene carrying the insertion. A model to describe how the selection is thought to occur is presented in Figure 1.5B. In essence, the addition of a transmembrane domain upstream of βgeo in the vector confers a type II membrane orientation on βgal fusion proteins which do not acquire signal sequences from the endogenous gene into which it has inserted. The type II orientation is thought to lead to the retention of the βgal reporter in the lumen of the endoplasmic reticulum (ER) where the βgal is inactive. In contrast signal sequences acquired by insertions in secreted and type I membrane proteins place the fusion protein in a type I orientation keeping the βgal outside the ER lumen where it is active.

1.4 Experimental approach

The aim of this project was to use the secretory trap technology to identify and mutate novel secretory molecules involved in mouse development. Chapter 2 describes the characterisation of lines generated from an electroporation of the secretory trap vector pGT1.8TM. To overcome the inherent bias of the pGT1.8TM gene trap design to insert into genes with large intronic targets an exon trap version of the secretory trap was designed. Chapter 3 describes the characterisation of lines generated by the new vector design. In Chapter 4 a detailed analysis of a secretory

trap insertion into a novel multiple membrane spanning gene required for lung development is presented.

Chapter 2

Secretory Trap Screen

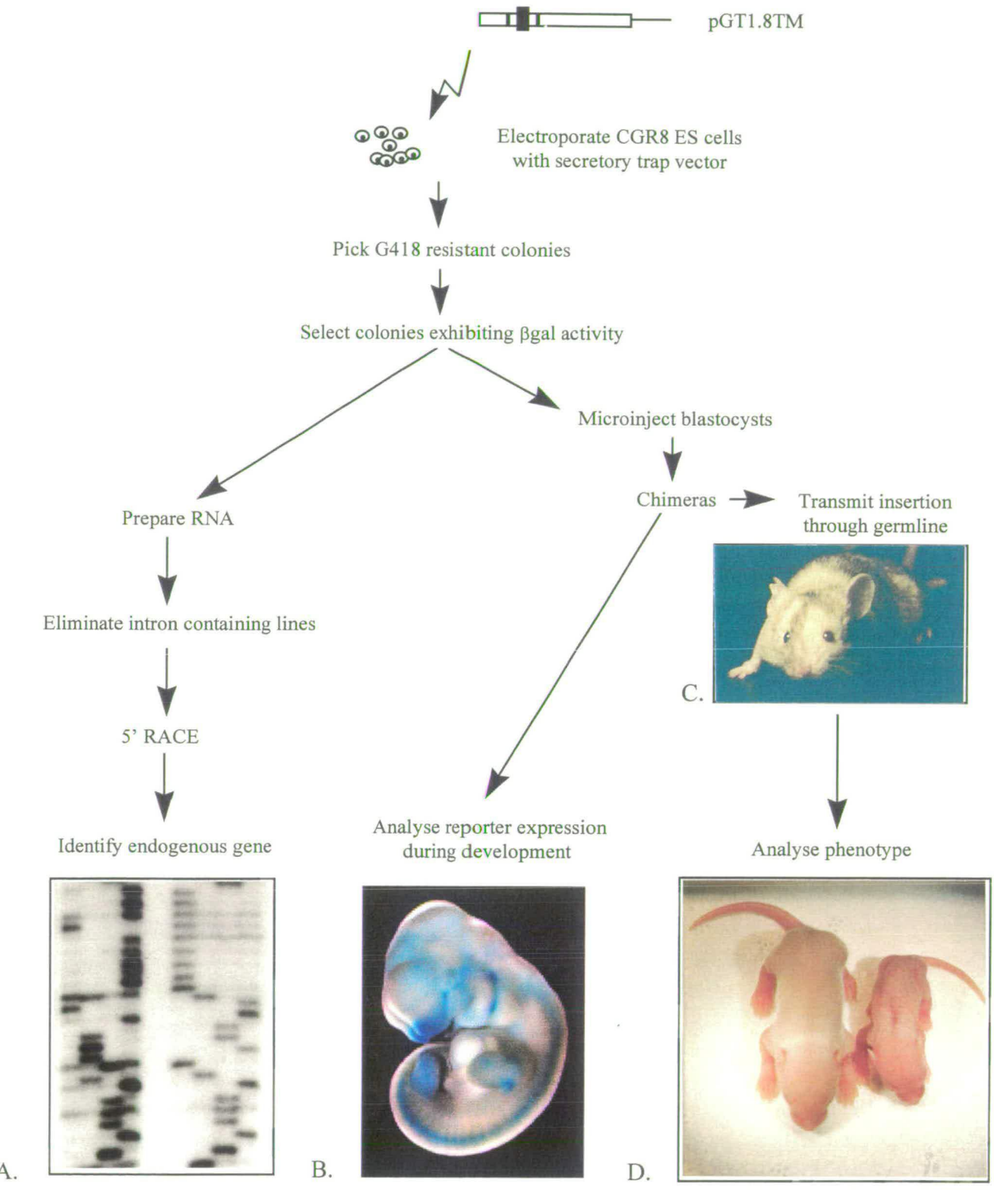
2.1 Introduction

The secretory trap vector (pGT1.8TM) has been shown to detect insertions in genes encoding secreted and membrane-spanning molecules (Skarnes *et al.*, 1995); see also Section 1.3.5. This pilot study identified insertions in the extracellular domain of a novel cadherin-related gene, the secreted molecule *netrin-1*, the *Sek* receptor tyrosine kinase and two protein tyrosine phosphatases, *LAR* and *PTPκ*. These initial results support the prediction that βgal activity is dependent on acquiring a signal sequence from the endogenous gene at the site of insertion.

This chapter describes the second more comprehensive screen using pGT1.8TM. Figure 2.1 outlines the strategy employed to identify genes important for embryonic development. X-gal positive clones were subjected to two pre-screens before transmitting the insertion through the germline. First, the insertion site was

Figure 2.1 Selection criteria for JST lines

(**A**) Direct sequencing of 5' RACE products. (**B**) βgal reporter expression in 10.5dpc JST191 chimera. (**C**) Strong coat-colour chimera shown to transmit a secretory trap insertion through its germ line. (**D**) Growth retarded JST185 homozygote alongside normal littermate.


pGT1.8TM
Electroporate CGR8 ES cells
with secretory trap vector
Pick G418 resistant colonies
Select colonies exhibiting βgal activity
Microinject blastocysts
Chimeras
Transmit insertion
through germline
Prepare RNA
Eliminate intron containing lines
5' RACE
Identify endogenous gene
Analyse reporter expression
during development
Analyse phenotype
A.
B.
C.
D.

characterised, and second, the expression pattern during early (8.5dpc) and late (10.5dpc) organogenesis was analysed in chimeras (Figure 2.2 and reviewed in Hogan *et al.*, 1994). Only bone fide gene trap insertions which were expressed in chimeras were passed through the germline and intercrossed to look for recessive embryonic phenotypes.

2.2 Isolation of JST lines and expression in ES cells

The secretory trap vector pGT1.8TM was introduced into CGR-8 ES cells by electroporation. After 10 days in G418 selection 205 resistant colonies were picked into 24 well plates. Confluent wells were passaged into duplicate 24 well plates. One of these was maintained as a master whilst the other was stained for βgal activity. βgal activity was assayed using the chromogenic substrate X-gal and 57 (28%) lines showed varying degrees of expression. Most of the X-gal positive clones (77%) exhibited patterns similar to the characteristic 'secretory pattern' of βgal staining in the peri-nuclear compartment of ES cells and multiple cytoplasmic dots described in the original report (Figure 2.3, A-D). This pattern is thought to reflect the targeting of fusion proteins to the ER/golgi apparatus. The remaining colonies showed either faint or strong staining in indistinct subcellular locations (Figure 2.3, E and F). Twenty-four of the βgal staining lines were expanded for further analysis and are referred to as JST lines.

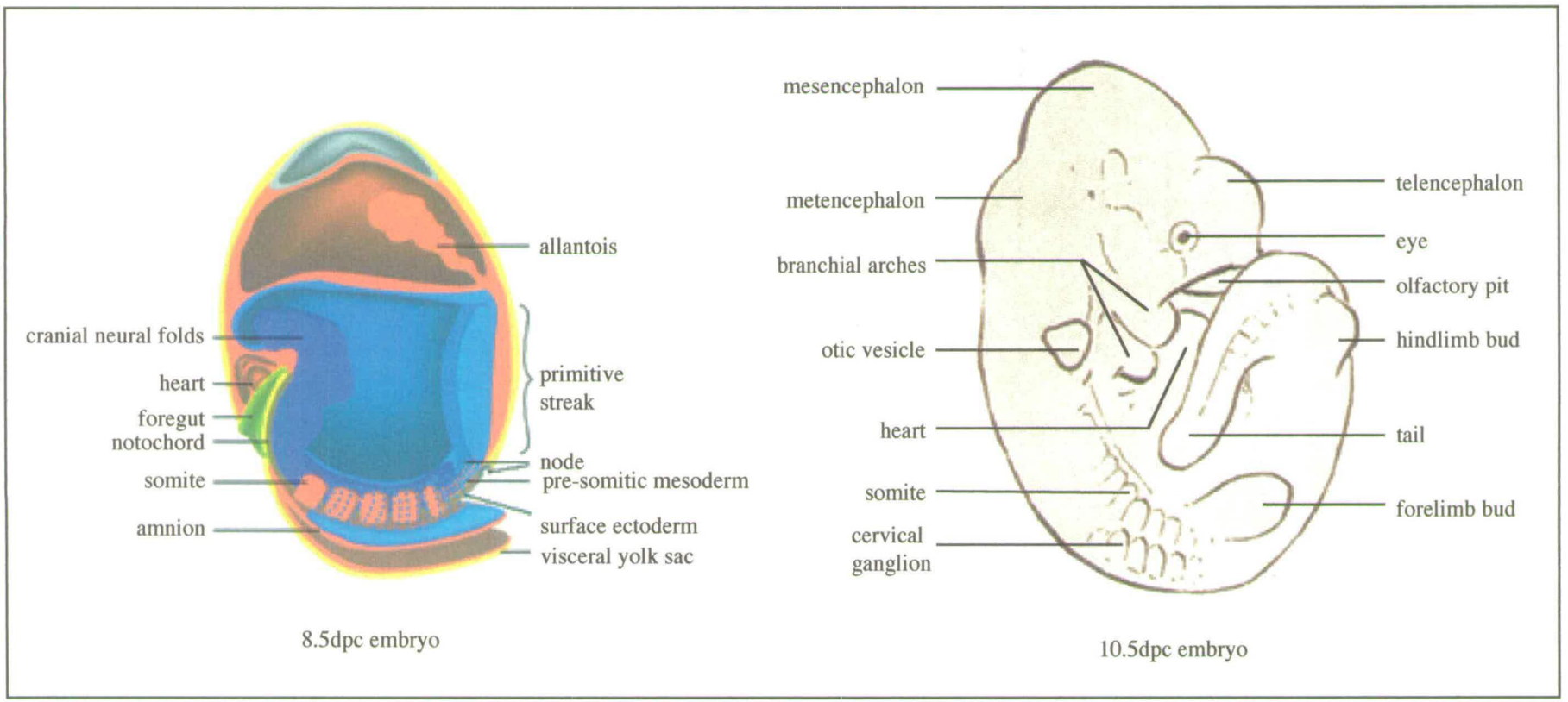
2.3 LacZ fusion transcript analysis

Previous results from our lab have shown that a significant proportion of gene trap insertions fail to utilise the vector's splice acceptor properly and so produce fusion transcripts containing vector-specific intron sequences. Inefficient splicing of the reporter gene is predicted to occur if the vector inserts in polymerase I or III transcription units or into exons of polymerase II transcription units (Figure 2.4). This prediction was borne out when two insertions producing intron-containing fusion

Figure 2.2 Summary of mouse development

(A) Reporter expression during the JST screen was analysed at 8.5 dpc and 10.5 dpc. The different lineages at 8.5d are coloured as follows, trophectoderm (grey); primitive endoderm (beige); epiblast (blue); mesoderm (red); notochor (brown); definitive gut endoderm (yellow) and neurectoderm (purple). Anterior is to the left and posterior to the right. 8.5dpc embryo figure taken from Hogan*et al*, 1994 and 10.5dpc embryo figure taken fromTheiler, 1972. **(B)** Time course of development. (0-5 days) Cleavage and blastulation; (5-10 days) implantation, gastrulation and early organogenesis; (10-14 days) organogenesis; (14-19 days) fetal growth and development. Taken from Hogan*et al*, 1994.

A.



B.

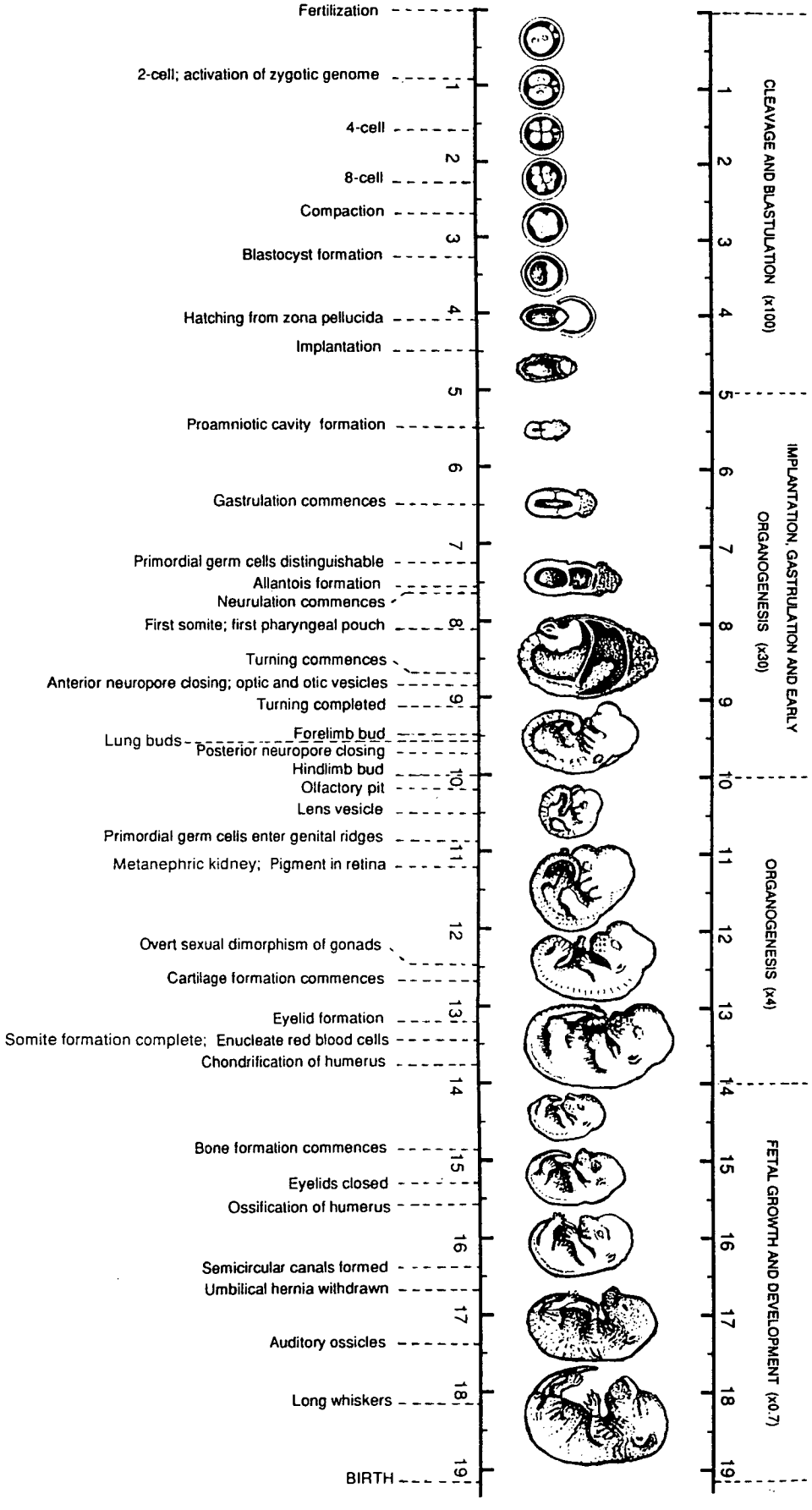
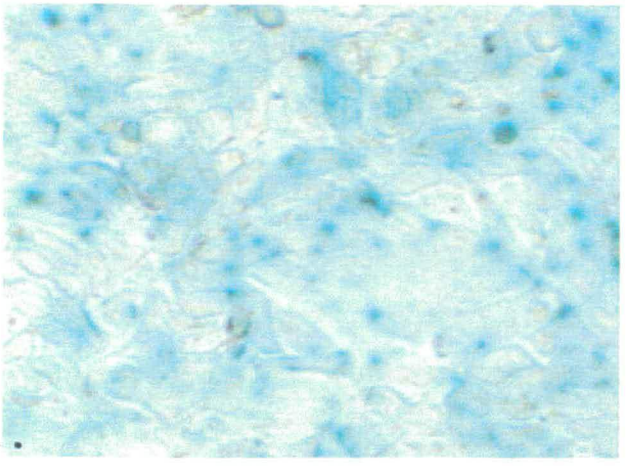
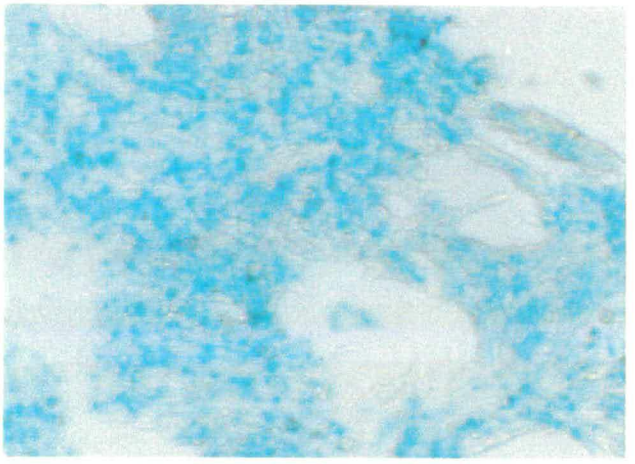

Fertilization
2-cell; activation of zygotic genome
4-cell
8-cell
Compaction
Blastocyst formation
Hatching from zona pellucida
Implantation
Proamniotic cavity formation
Gastrulation commences
Primordial germ cells distinguishable
Allantois formation
Neurulation commences
First somite; first pharyngeal pouch
Turning commences
Anterior neuropore closing; optic and otic vesicles
Turning completed
Forelimb bud
Lung buds
Posterior neuropore closing
Hindlimb bud
Olfactory pit
Lens vesicle
Primordial germ cells enter genital ridges
Metanephric kidney; Pigment in retina
Overt sexual dimorphism of gonads
Cartilage formation commences
Eyelid formation
Somite formation complete; Enucleate red blood cells
Chondrification of humerus
Bone formation commences
Eyelids closed
Ossification of humerus
Semicircular canals formed
Umbilical hernia withdrawn
Auditory ossicles
Long whiskers
BIRTH
CLEAVAGE AND BLASTULATION (x100)
IMPLANTATION, GASTRULATION AND EARLY ORGANOGENESIS (x30)
ORGANOGENESIS (x4)
FETAL GROWTH AND DEVELOPMENT (x0.7)

Figure 2.3 βgal activity in JST ES cell lines

ES cell lines containing pGT1.8TM stained for βgal activity. (**A**) JST51 and (**B**) JST219 show intense dots of cytoplasmic expression amidst a background of lower level expression in the cytoplasm. The pattern is similar to the typical 'secretory pattern' of expression reported in Skarnes *et al*, 1995. In (**C**) JST85 and (**D**) JST136 patchy low level expression is detected showing a similar pattern to A and B. (**E**) JST71 shows very faint βgal expression and (**F**) JST100 shows high level expression in nuclei and cytoplasm.

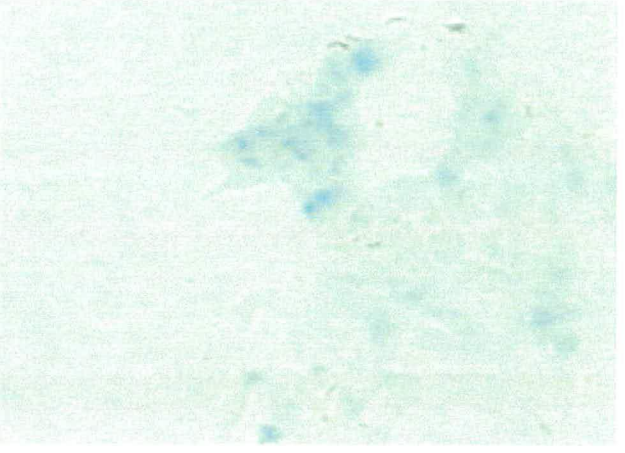


A. JST51(LDL receptor -secretory molecule)

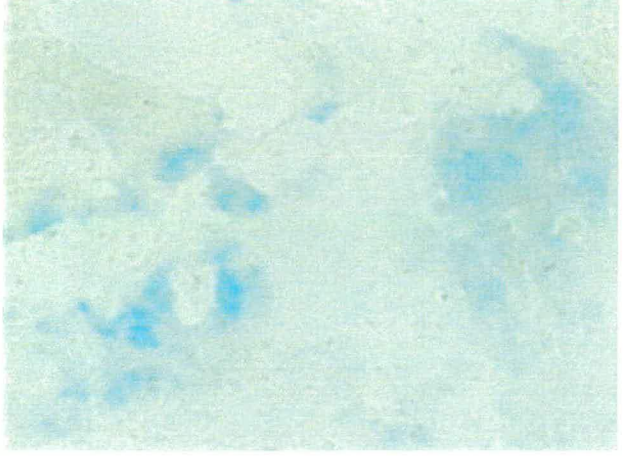


B. JST219 (cadherin - secretory molecule)

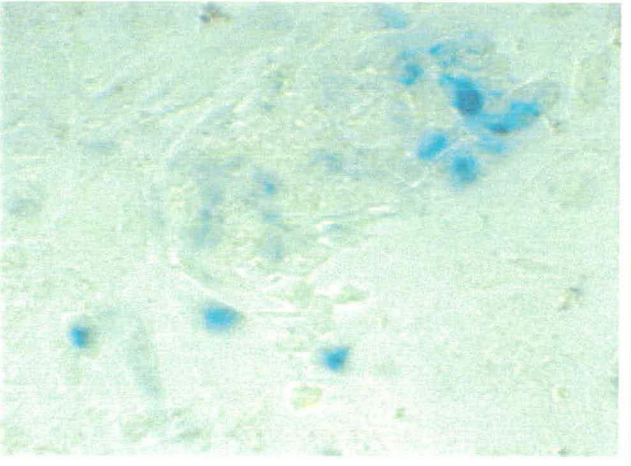
C. JST85 (5'UTR insertion)



D. JST136 (5'UTR insertion)



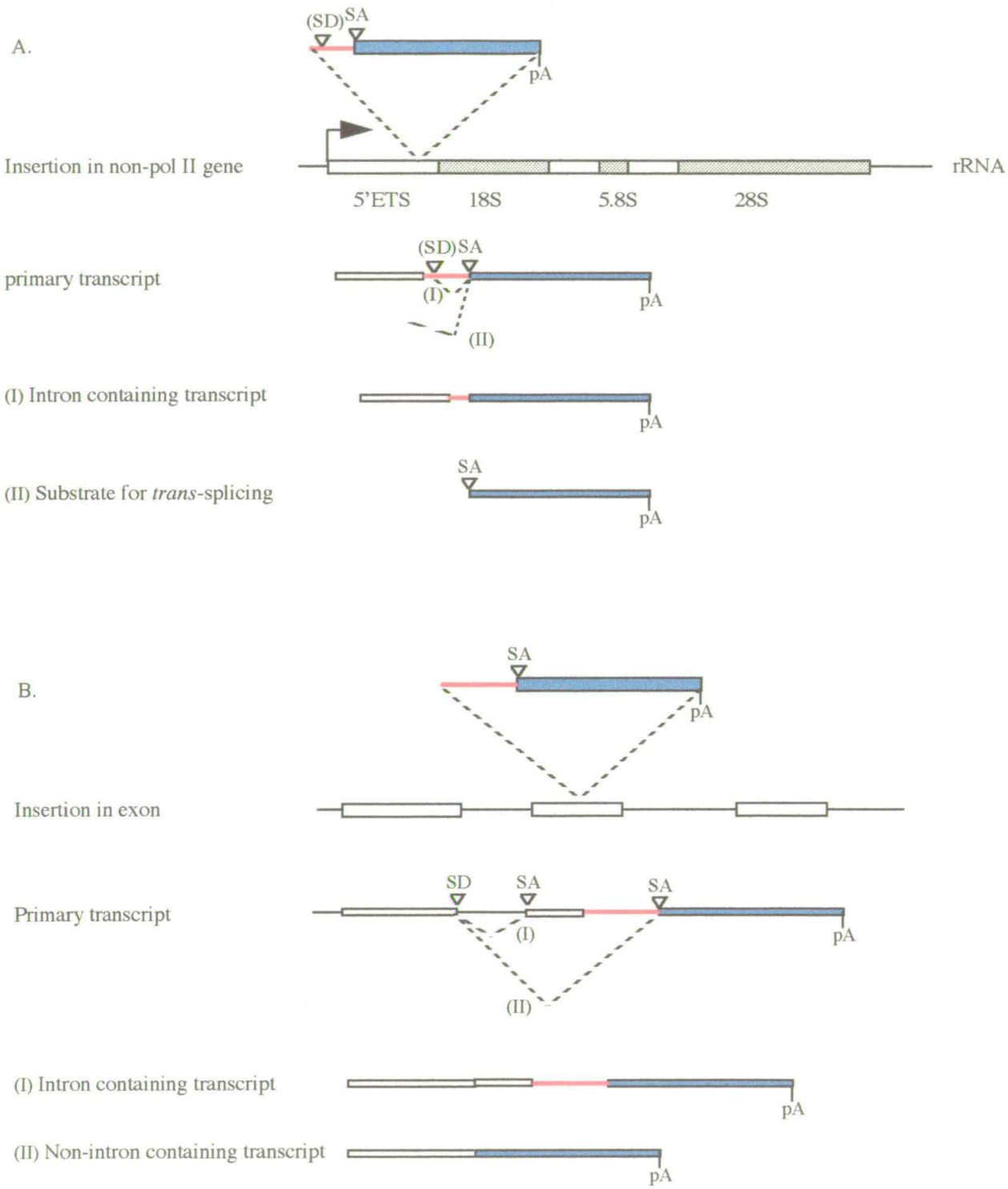
E. JST71 (non-secretory molecule)



F. JST100 (rRNA insertion)

Figure 2.4 Predicted integrations producing intron-containing transcripts

The 30% of secretory trap lines which produce intron-containing transcripts can arise in two ways shown in **(A)** and **(B)**. An uncapped pol I primary fusion transcript is produced. **(I)** *Cis*-splicing to a cryptic splice donor site within the *En2* intron produces a transcript containing the 5' region of the intron. Such transcripts are predicted not to be translated as they are uncapped and the *En2* intron contains stop codons in all three frames. **(II)** *Trans*-splicing of the *En2* splice acceptor to the splice donor site produces a capped mRNA that can be translated to give a βgeo fusion protein. **(B)** Insertion of the secretory trap vector into an exon of a pol II would produce a primary transcript in which the introduced splice acceptor would compete for the endogenous splice donor with the endogenous splice acceptor. **(I)** Splicing of the endogenous splice acceptor to the splice donor would produce an intron containing transcript. A functional fusion protein would not be produced from this transcript due to the stop codons in the *En2* intron. **(II)** Splicing of the vector splice acceptor to the splice donor would result in the removal of the partial endogenous exon from the transcript together with the *En2* intron sequences. The transcript is capable of being translated into a functional βgeo fusion protein.


A.
(SD)
SA
pA
Insertion in non-pol II gene
rRNA
5'ETS
18S
5.8S
28S
primary transcript
(SD)
SA
(I)
pA
(II)
(I) Intron containing transcript
pA
(II) Substrate for *trans*-splicing
SA
pA
B.
SA
pA
Insertion in exon
Primary transcript
SD
SA
SA
(I)
pA
(II)
(I) Intron containing transcript
pA
(II) Non-intron containing transcript
pA

transcripts were characterised. Both clones contained vector insertions in rRNA genes which are transcribed by RNA polymerase I (J.E. Sleeman & W.C. Skarnes, unpublished results). Insertions into rRNA genes appear to participate in *trans*-splicing as 5'RACE products contain not one, but several different endogenous gene sequences joined to the splice acceptor of the vector. It is thought that the uncapped RNA polymerase I transcripts are very poor substrates for *cis*-splicing, however the splicing machinery assembled at the splice acceptor of the vector is free to interact with other splice donors *in trans* (W.C. Skarnes, pers communication). Upon insertion into an exon of a polymerase II gene, the splice acceptor of the vector is expected to compete with the splice acceptor of the disrupted exon. Although this second class of event is likely to be mutagenic, it is likely that wild type transcript would be present and thus not represent a complete loss-of-function allele.

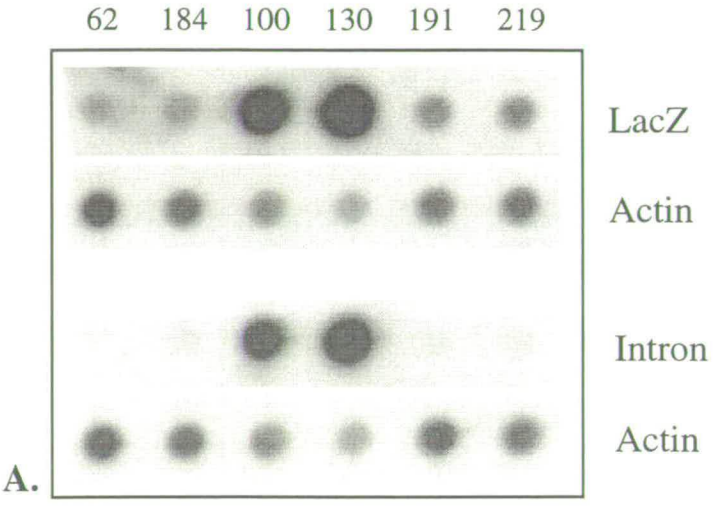
Lines producing fusion transcripts containing vector intron sequences were eliminated from the screen using RNA dot blot analysis. Twenty-four JST lines were analysed and 7 (29 %) were shown to produce intron containing transcripts. To assess whether a unique fusion transcript was produced in the efficiently spliced lines, 7 were analysed by Northern hybridisation using a Lac Z probe (Figure 2.5B and data not shown). A single major fusion transcript hybridised to LacZ in all lines analysed although the signal intensity varied between lines. Transcripts ranged between 4 kb and over 9 kb in size. Since the vector sequences compose 4kb of the fusion transcript, fusion transcripts included up to 5kb of the endogenous gene.

2.4 Identification of endogenous genes carrying insertions

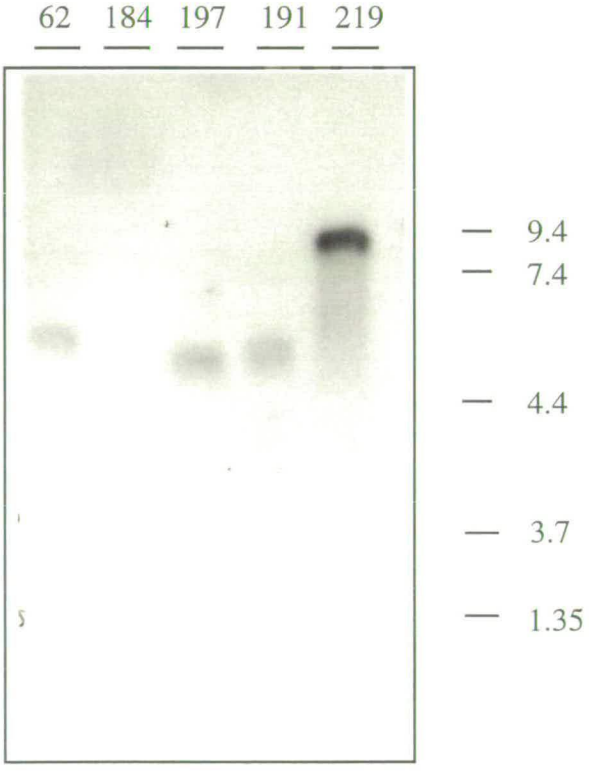
5'RACE was used to clone cDNA sequences spanning the vector splice junction in 13 intron negative lines (outlined in Figure 2.6). To monitor the efficiency of 5'RACE an aliquot of the second round products were dot blotted and hybridised with a vector specific exon probe which spans the splice acceptor site of the vector and

Figure 2.5 Analysis of transcripts in JST lines

(A) RNA dot blot analysis was used to eliminate lines producing intron containing transcripts. RNA dot blotted in duplicate and transferred to filters and hybridised to a LacZ probe or *En2* intron probe (See map on Figure 2.7). All six lines hybridised with varying intensities to LacZ but only two, JST 100 and JST130, hybridised to *En2* intron. To control for differences in amounts of sample loaded, both filters were stripped and rehybridised to a β-actin probe. **(B)** The size of individual fusion transcripts, analysed by Northern hybridisation using a LacZ probe. Single fusion transcripts were detected in each line which varied in size from 4kb to over 9kb. Expression levels are very different between lines (compare JST184 to JST219). The contribution from the vector to the fusion transcript is estimated to be 4 kb (0.18 kb of En2, 0.68 kb of CD4 fused to 3.03 kb of βgeo and the 0.1-0.2 kb poly A tail). **(C)** Ethidium bromide staining of RNA samples showing 28S rRNA which acts as a loading control.

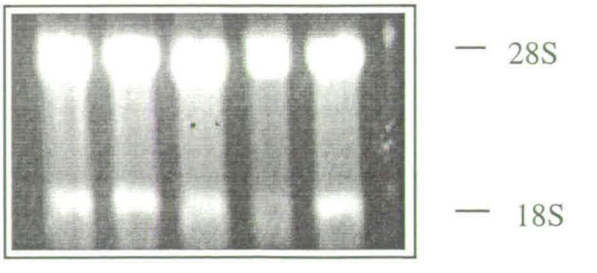


A.



B.

LacZ

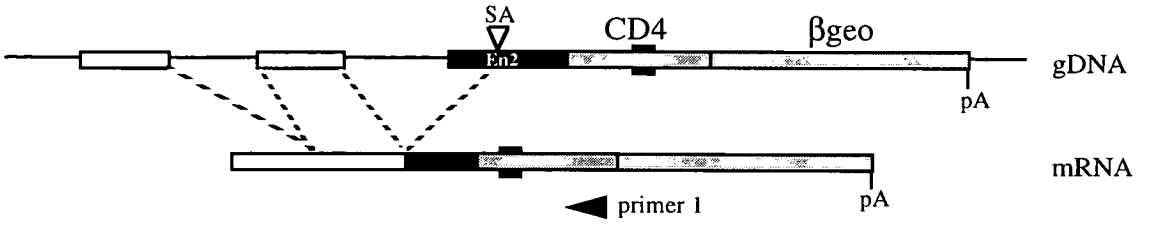


C.

EtBr

Figure 2.6 5'RACE used to identify endogenous genes

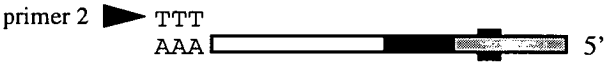
Two routes to identify endogenous genes using a 5'RACE strategy.



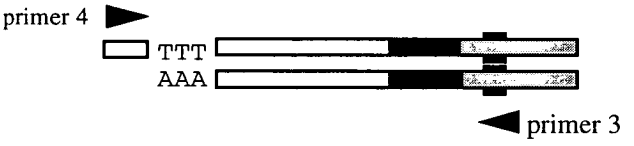
1st strand cDNA synthesis with primer 1

Tailing of 1st strand cDNA with dATP

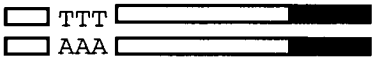
Anneal primer 2 containing oligo dT stretch



Second strand synthesis



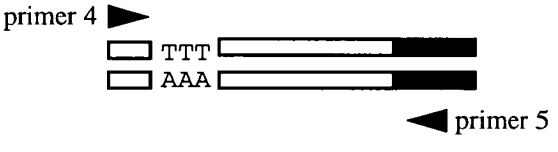
1st round PCR with primers 3 and 4



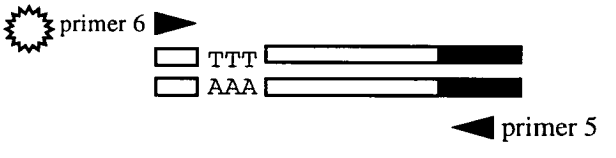
Cloning route

Direct sequencing route

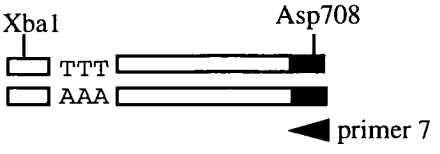
Second round PCR with primer 4 and 5



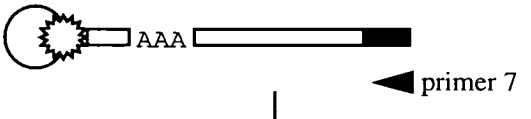
Second round PCR with primer 5 and biotinylated primer 6



Digest second round PCR products



Capture biotinylated products on streptavidin coated beads and denature



Clone products into XbaI/Asp708 cut pBluescript II KS+

Cycle sequence using ^{32}P end labelled primer 7

Sequence clones with primer 7 by dideoxy-termination method

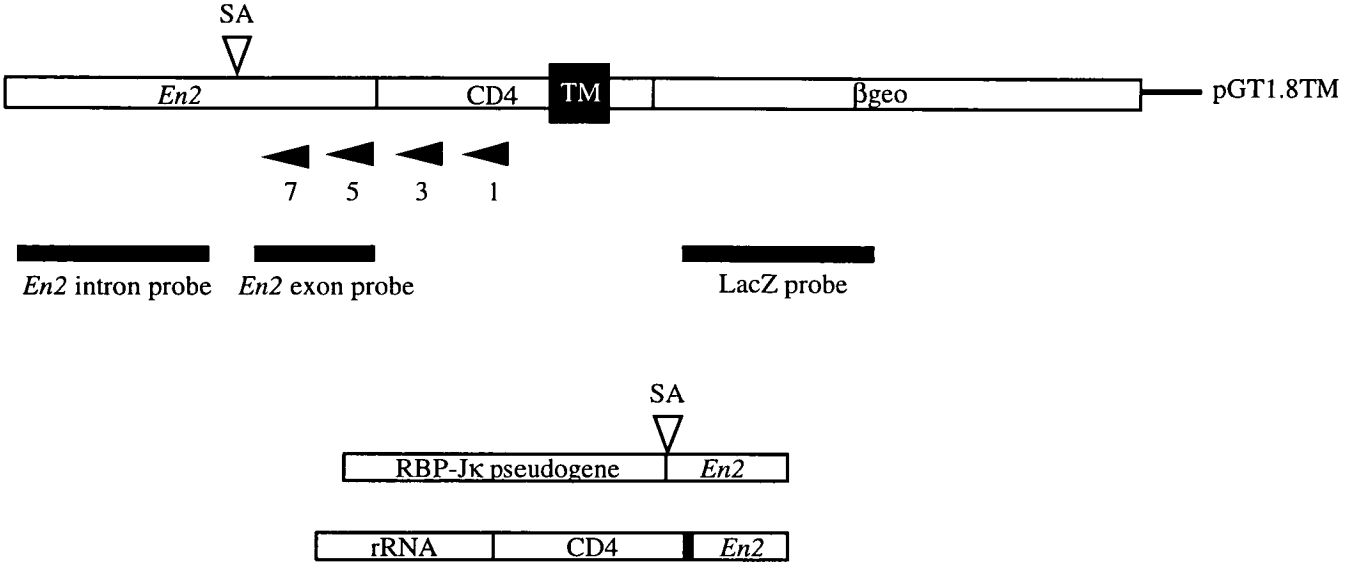
is expected to be present in fusion transcripts (Figure 2.7I). A strong exon-specific signal was present in all but one line (JST184; data not shown). Further analysis of JST184 is reported in section 2.6.

Initially five lines were cloned into pBluescript IIKS+. The resulting clones gave between 73 and 515 bp of endogenous sequence (Table 2.1). At least two independent RACE clones from each line were sequenced. Subsequently the remaining 8 lines were sequenced directly from RACE PCR products using a protocol developed in the laboratory (Townley *et al.*, 1997). Second round products for these 8 JST lines were biotinylated and immobilised on streptavidin beads. Following denaturation, single stranded products were cycle sequenced using a ^{32}P end-labelled *En-2* primer.

A single unique sequence ladder was obtained for each line except JST232. In this line there were clearly two sequences superimposed on each other. Careful scrutiny of the double sequence in line JST232 revealed that one sequence corresponded to a properly spliced fusion transcript in which sequences from the RBP-Jκ pseudogene were joined to the splice acceptor (Figure 2.7B). The second superimposed sequence was identified as vector CD4 sequence which was also joined to *En-2* sequence but on this occasion the genuine splice acceptor had not been utilised. Separating the CD4 and *En-2* sequence were 9bp of unknown sequence, upstream of CD4 were rRNA sequences (Figure 2.7B). It is unclear how this product arose but the most likely explanation is the integration of, and splicing within, tandem copies of the vector. It remains possible that JST232 carries multiple insertion sites or represents a mixed colony.

2.5 Sequence analysis of endogenous genes

In 8 JST lines, the βgal ORF continued upstream into novel endogenous sequence; however 4 lines demonstrated no upstream ORF continuity (Figure 2.7C,



A. En2 splice acceptor site

```
                                                    ↓
acctctgcccttttctcctccatgacaaccagGTCCCAGGTCCCGAAAACCAAAGAAGAAG
                                      P  R  S  R  K  P  K  K  K
```

B. JST203
In frame fusion

```
                                                    ↓
CCAGCTTGCCCTGAGGAGTCCCCGCTGCTCGGTCCCAGGTCCCGAAAACCAAAGAAGAAG
P  A  C  P  E  E  S  P  L  L  G  P  R  S  R  K  P  K  K  K
```

C. JST85
No open reading frame

```
                                                    ↓
CCCAACCACAAGACGACTCAATAGGAGGAAGGTCCCAGGTCCCGAAAACCAAAGAAGAAG
                        *  E  E  G  P  R  S  R  K  P  K  K  K
   A  P  Q  D  D  S  I  G  G  R  S  Q  V  P  K  T  K  E  E
```

Figure 2.7 Endogenous genes utilise the vector splice acceptor site

(I) Schematic of pGT1.8TM showing sites of RACE primers and probes. The primer numbers are based on Figure 2.6. **(II)** Structures of the two products directly sequenced from JST232. The top product is correctly spliced, the lower may be a result of splicing within vector sequences or may be a consequence of vector rearrangment. **(III)** Nucleotide and amino acid sequence of the secretory trap splice acceptor and the endogenous genes in two JST lines. (A) Sequence of *En2* splice acceptor in pGT1.8TM (intron sequences shown in lowercase and exon sequences in uppercase). The consensus splice acceptor nucleotides including the pyrimidine stretch are underlined. The splice site is marked with a vertical arrow. Nine amino acids of the *En2* reading frame which are fused in-frame with the βgeo reporter are shown. (B) Eight out of twelve JST lines produce a fusion transcript with endogenous sequences encoding a protein in the same frame as the vector reporter gene. For example the endogenous cDNA cloned from JST203 is shown to encode a protein in the same frame as βgeo. (C) The lack of an open reading frame in the same frame as βgeo in three of the JST lines is due to the insertions having occurred in 5'UTRs. In the remaining line, JST85, the endogenous protein reading frame (in italics) is in a different frame to βgeo. The reporter activity seen in this line is thought to be a product of internal initiation of translation from one of the methionines within the vector.

Table 2.1 5' RACE sequence information for JST lines

Line	5'RACE sequence (amount in bp)	ORF	Database accession number	BLAST score*	Insertion site (nucleotide)	Gene
Identified by cloning						
JST184	57	5'UTR	no match	n/a	n/a	novel
JST185	448	in frame	U40060	4.3×10^{-18}*	1595**	novel
JST191	73	in frame	AA118295 (EST)	2.5×10^{-12}	n/a	novel
JST197	51	5'UTR	U06939	3.6×10^{-08}	162	non-muscle myosin heavy chain
JST219	515	in frame	X87241	3.6×10^{-135}	3763	*fat*-like cadherin
Identified by direct sequencing						
JST51	85	in frame	Z19521	1.5×10^{-25}	433	LDL receptor
JST62	100	in frame	D31887	1.6×10^{-22}	575	myeloblast cDNA
JST71	120	in frame	J04633	9.4×10^{-35}	1749	heat shock protein 86
JST85	145	out of frame	D28476	1.7×10^{-52}	208	myeloid cDNA
JST136	46	5'UTR	D21729***	6.0×10^{-06}	70	p23
JST203	80	in frame	M27917	2.0×10^{-23}	1023	β1,4-galactosyltransferase
JST213	71	in frame	U95727	8.0×10^{-17}	192	DNA-J2 homologue
JST232	50	5'UTR	M81871	5.5×10^{-10}	236	RBP-Jκ/Su(H) pseudogene

* All are BLASTN scores except JST185 which refers to BLASTX score
** Insertion site within JST185 cDNA sequence (See Chapter 4)
*** Murine est which is 97 % identical to human p23
n/a not applicable

Table 2.1 & Appendix II). The endogenous sequences were compared to the non-redundant GenBank databases using the BLASTN and BLASTX algorithms (Altschul *et al.*, 1990). The BLASTN program searches for stretches of nearly identical sequence, and matches are scored according to the probability of their occurrence by chance alone. BLASTX compares a nucleotide query sequence translated in all reading frames against a protein sequence database. Ten sequences were perfect or close matches to database entries (Table 2.1). JST191 did not show any homology to any cDNA entries but closely matched a human and mouse expressed sequence tag (EST) in the DBEST GenBank database. JST185 did not show homology to any entries in either the non-redundant or DBEST databases but showed structural similarity to two hypothetical *C.elegans* proteins using the BLASTX algorithm. A full description of the JST185 sequence is presented in Chapter 4.

Three of the 12 insertions occurred in known secretory molecules; the LDL receptor (JST51), the murine homologue of *Drosophila fat* (JST219) and β1,4-galactosyltransferase (JST203). Using a program available on the world wide web, TMpred, (http://ulrec3.unil.ch/software/TMPRED_form.html) the predicted open reading frames of the novel set of JST sequences were analysed for hydrophobic sequences that might act as signal sequences or transmembrane domains. This program identifies putative membrane-spanning regions in the primary amino acid sequence and also predicts the possible orientation of these segments in the membrane. TMpred uses a combination of multiple weight-matrices that have been extracted from a statistical analysis of the annotated transmembrane proteins present in SwissProt. Another three sequences (JST62, 185 and 213) showed homology to cDNA sequences of unknown function that contain hallmarks of secreted gene products.

JST62 shows 85% identity at the protein level to a human myeloblast cDNA and therefore may represent either the murine homologue of this gene or a closely related family member. The myeloblast cDNA is predicted to have 9 transmembrane

domains but there is no obvious signal sequence. The lack of the signal sequence may be genuine or may reflect the fact that the database entry for the myeloblast cDNA does not encode an initiating methionine, suggesting that the open reading frame is incomplete. Comparison between JST62 and the myeloblast protein predicts that the secretory trap insertion occurred after the second transmembrane domain. Similarly, the full-length JST185 coding sequence is predicted to encode 9 membrane-spanning domains upstream of the site of insertion (see Chapter 4).

The gene interrupted in JST213 shows 95% identity over 23 amino acids to both a rat sequence, DnaJ homologue 2 (U95727), and a human sequence, Dnj3/Cpr3 (AF011793). Over their entire lengths, the human and the rat proteins are 57% identical to each other. Such similarity suggests that these two genes and JST213 may be homologues. The human and rat sequences show 73% identity over a 73 amino acid motif (the J domain) that is conserved in proteins of the DnaJ family of chaperonins. The TMpred program predicts a strong transmembrane domain in the both the human and rat sequences upstream of the region of JST213 homology.

The RACE sequence for JST191 does not show homology to any cDNAs in the database but shows 80 % identity to a murine EST isolated from a 13.5dpc embryo (AA118295). The predicted open reading frame for the EST contains multiple proline residues. The βgal staining pattern in JST191 ES cells suggests that it carries an insertion in a secretory molecule, however, the available sequence is too short to be informative.

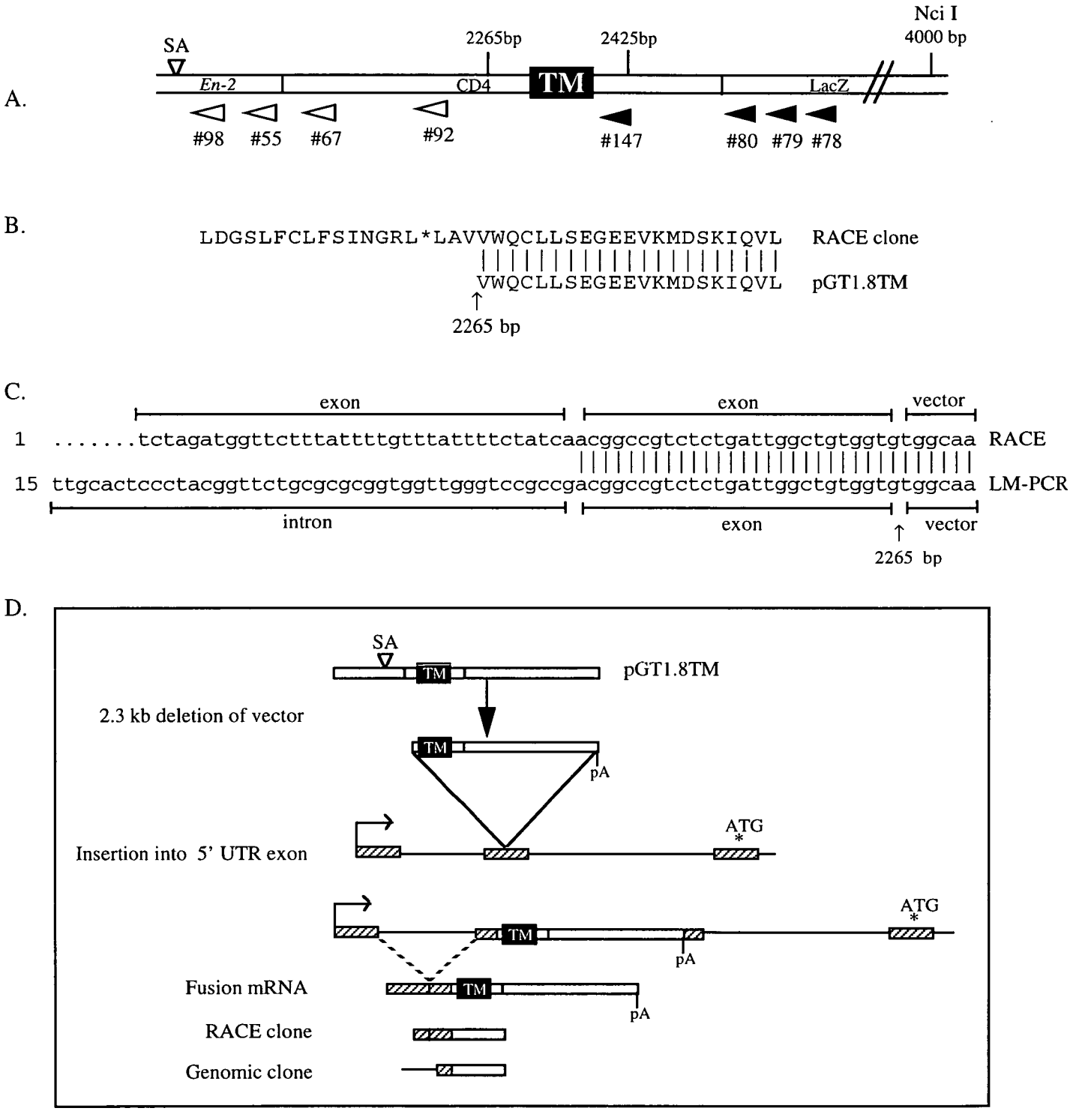
Unexpectedly, five of twelve insertions occurred in known or predicted non-secretory molecules; heat shock protein 86 (JST71), cytoplasmic p23 (JST136), non-muscle myosin heavy chain (JST197) and a pseudogene of a murine homologue of the *Drosophila* transcription factor *Suppressor of Hairless*, (JST232). The latter three insertions occurred in 5'UTRs. JST85 carries an insertion in a novel cDNA from

myeloid cells which has no hallmarks of a secretory molecule. However the myeloid protein coding sequence is fused out of frame with βgeo. In four out of five of these non-secretory insertions there is no endogenous open reading frame in the same frame as βgal. These results indicate that βgal activity is detected in these lines via internal initiation of translation from one of the vector's methionines

2.6 Vector deletion prior to insertion

Dot blot analysis of second round amplified products revealed that JST184 RNA could not be amplified with the primers routinely used for RACE. Northern blot analysis confirms the presence of a fusion transcript in the RNA prepared from JST184 ES cells (Figure 2.5B). This suggests that the lack of amplification may be due to deletion of one or more of the primer sites in the vector prior to its insertion in the genome. The observation that the JST184 fusion transcript is only slightly larger than the 4 kb expected to be contributed by the vector is consistent with partial deletion of vector sequences.

A second set of primers were designed which are located further 3' in vector sequences than the primers routinely used for 5' RACE of secretory trap lines (Figure 2.8A). Several RACE clones of different sizes were obtained and sequencing revealed that each clone contained 160 bp of CD4 sequence and up to 60 bp of novel sequence. The novel sequence did not encode an open reading frame with βgeo. This suggests that either part of the vector has been deleted, or an alternative splice acceptor site in the CD4 sequence is being used. To investigate these alternatives the sequences flanking the insertion site were cloned from genomic DNA using ligation-mediated (LM) PCR (outlined in Figure 3.4). This is a single-sided PCR method that requires specification of only one primer hybridisation site, the second is defined by the ligation-based addition of a unique DNA linker. This linker, together with the vector-specific primer, allows amplification of DNA fragments flanking insertion sites.



igure 2.8 Vector deletion occurs prior to insertion in JST184

A) The primers routinely used to amplify cDNAs from secretory trap lines are indicated on a map of GT1.8TM as open arrow-heads, (1st round synthesis, #92; 1st round PCR, #67; 2nd round PCR, #55; equencing, #98). These primers failed to amplify a product from JST184 RNA. Primers were designed rther 3' in the vector sequence to examine whether the failure to amplify was due to deletion of the primer tes. The new primers are indicated as filled arrowheads (1st round synthesis, #78; 1st round PCR, #79; 2nd ound PCR, #80; direct sequencing, #147). **(B)** RACE clones contained only 160 bp of CD4 sequence dicating that 2265 bp of vector sequence had been deleted. **(C)** A junction fragment was amplified using M-PCR (see Chapter 6 for details). The fragment contained the same amount of CD4 sequence as the RACE one confirming that 2265 bp of vector had been deleted prior to insertion. Immediately upstream of the CD4 as 23 bp of sequence which was identical to that found in the RACE clone. The next 55 bp of sequence was fferent. Presumably the genomic clone extends into intron sequence and the RACE clone extends into further xon sequence. **(D)** Representation of proposed insertion in JST184.

Genomic DNA prepared from JST184 ES cells was digested with a frequent cutter enzyme and denatured. Nci I was used for the digestion as it does not cut at the 5' end of the pGT1.8TM vector and because its target site is made up of C and G nucleotides. Northern blot analysis suggested that the vector insertion had occurred towards the 5' end of JST184 possibly in the promoter region which may be expected to be rich in C and G nucleotides. DNA was synthesized using a primer directed against LacZ sequences and the adapter ligated. PCR was performed using primers directed against vector CD4 sequences and an adapter primer. The products were digested with XbaI which cuts within the linker and cloned into pBluescript IIKS+ digested with Xba and EcoRV. The sequence revealed the same amount of vector was present as observed with the RACE clone (Figure 2.8C). This suggests that 2.3 kb of sequence, including the *En-2* splice acceptor, had been deleted from the vector prior to insertion in the genome.

In the genomic clone the first 22 bp of sequence fused to the vector CD4 was identical to that found in the RACE clone. This suggested that the vector insertion had occurred in an exon. A further 66 bp of sequence differing from that found in the RACE clone was fused to the exon in the genomic clone. This may correspond to intron sequence although there is no consensus splice site between the two sequences. In the RACE clone 35 bp of novel sequence is fused to the 22bp of exon sequence which may correspond to an upstream exon. The lack of an open reading frame in the RACE clone suggests the JST 184 insertion occurred in an untranslated exon. The small number of base pairs which contribute to the fusion transcript suggests that the insertion occurred in the 5'UTR (Figure 2.8 D). The JST184 sequence did not match any entries in GenBank.

2.7 Expression patterns associated with JST insertions

To identify developmentally regulated genes the expression pattern of the reporter was monitored in chimeric embryos. Expression was analysed in 11 cloned lines between 8.5 and 10.5 dpc. To investigate the potential for active βgal in inefficiently spliced lines, 3 additional intron-containing lines were analysed. Embryos were stained whole with X-gal to visualise lacZ expression. Different chimeric embryos produced from any particular cell line showed identical staining patterns.

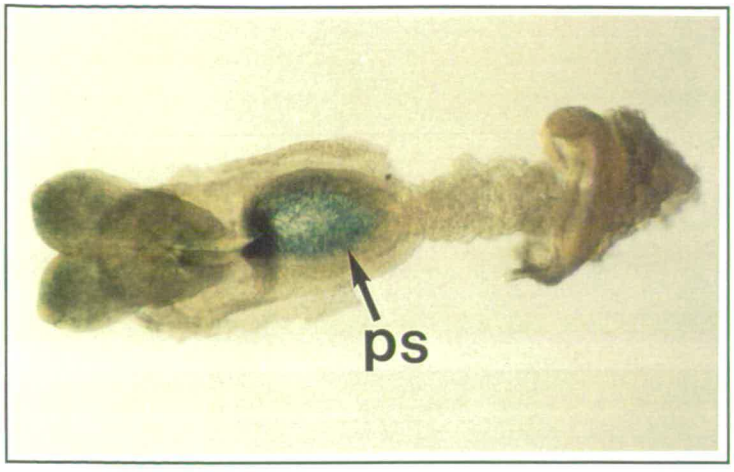
A variety of X-gal staining patterns were observed in chimeras generated from the different cell lines (Figure 2.9 and Table 2.2). Four of the lines, including the three lines which produce intron containing transcripts, did not express lacZ at the developmental stages analysed. In three lines, lacZ expression was detected in all tissues. Seven lines exhibited developmentally regulated patterns of expression.

2.8 Germline transmission of JST insertions

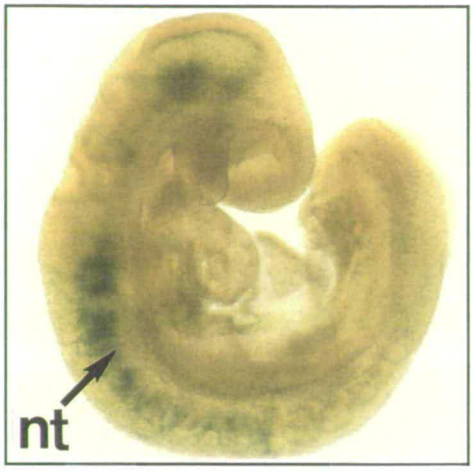
C57Bl/6 blastocysts injected with JST ES cells were allowed to develop to term within the uteri of pseudopregnant mothers. The ES cell contribution to offspring was estimated on the basis of coat colour. Male coat-colour chimeras were test bred to C57Bl/6 females and generation of agouti offspring indicated successful germline transmission. Offspring were genotyped by dot blotting tail biopsy DNA and hybridising with a *neo* probe. Offspring heterozygous for secretory trap insertions were back crossed to C57Bl/6 to dilute the CGR8 genetic background, which may have accumulated mutations during culture. The resultant F2 offspring were then intercrossed and offspring were genotyped by dot blotting. Clear differences in signal intensity were seen between heterozygotes and homozygotes as long as the protocol for tail biopsy size was strictly adhered to (data not shown).

Figure 2.9 Reporter expression in JST ES cell line derived chimeras

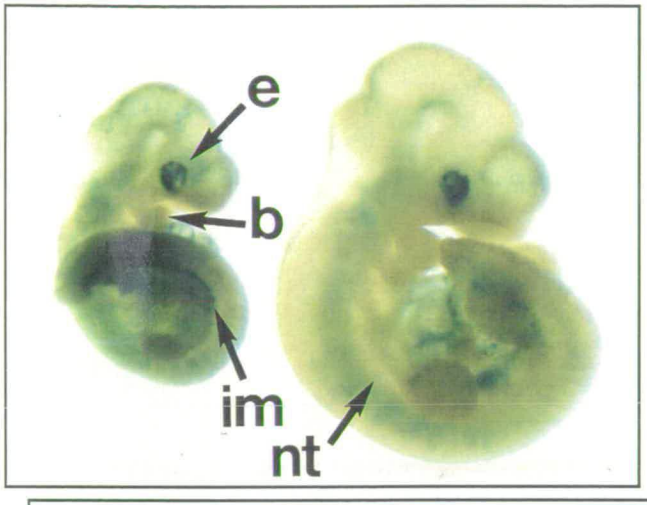
Reporter expression was analysed between 8.5 - 10.5dpc chimeras (A-E,G-I) or heterozygotes (F). **(A)** JST62. **I** Dorsal view of 8.5dpc embryo reveals LacZ expressed at high levels in the primitive streak (ps). **II** Lateral view of a late 9dpc embryo. Expression in ventral half of the neural tube (nt). **(B)** JST184. **I** Lateral views of 9dpc and 10.5dpc embryos. High levels of expression in the developing eye (e) and in the intermediate mesoderm (im). Expression is absent from the branchial arches (b). A day later expression is seen in the ventral neural tube (nt). **(C)** JST85. Lateral and dorsal views of 8.5dpc embryos. **(D)** JST197. Semi-lateral view of 10dpc thoracic regions show expression restricted to the developing heart (h). **(E)** JST203. **I** Ventral views of 8.5dpc embryos, expression restricted to the border of the closing yolk sac (ys). **II** Lateral views at 10.5dpc show expression at high levels in the developing heart. Clusters of staining cells around somitic vessels in most posterior regions (arrows).



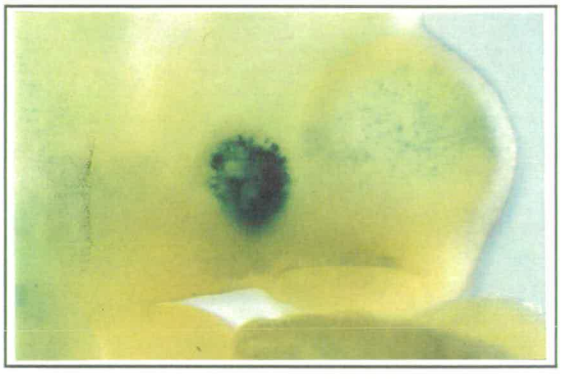
I



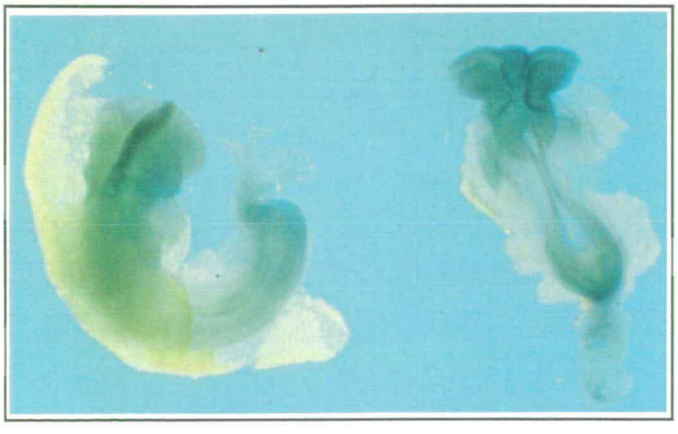
AII



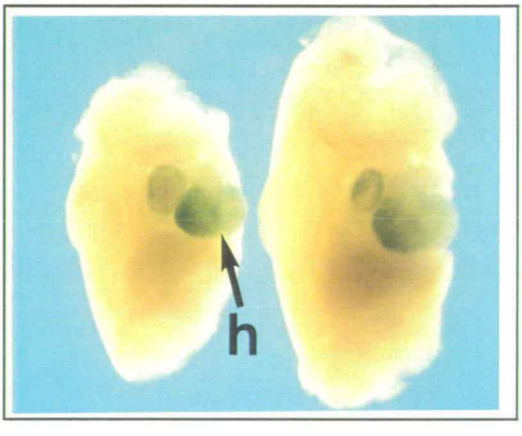
I



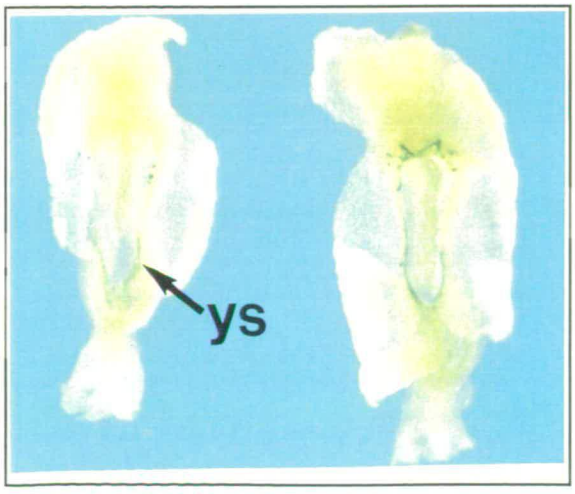
BII

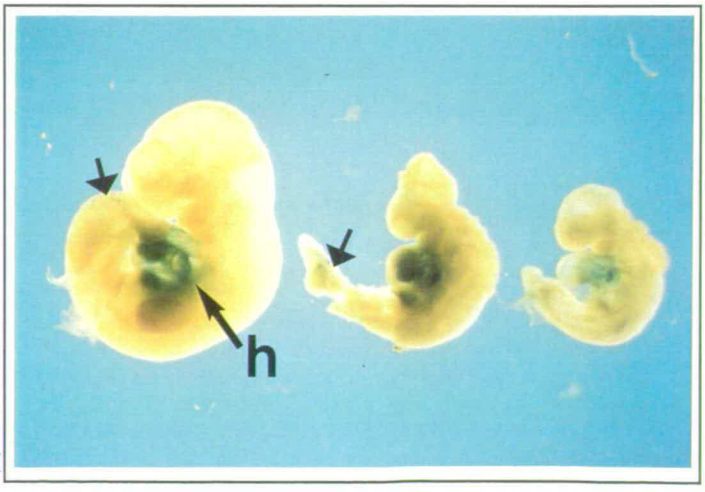


C



D

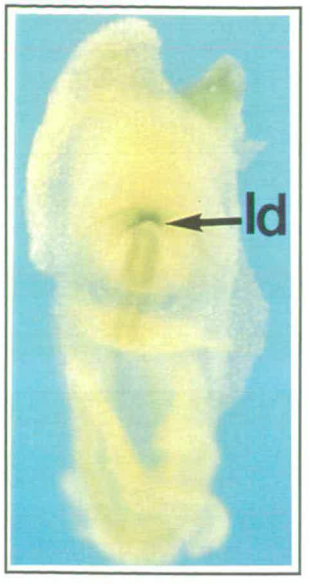




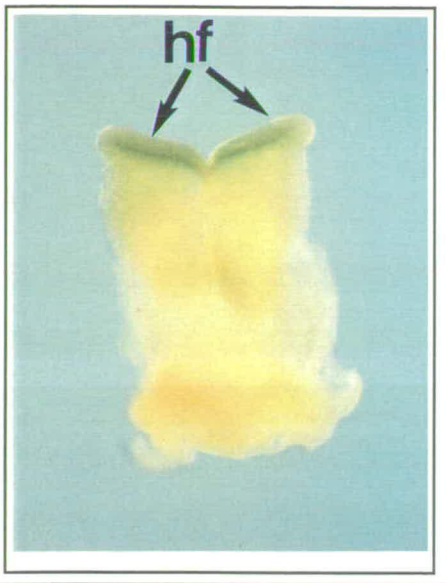
E II

Figure 2.9 cont

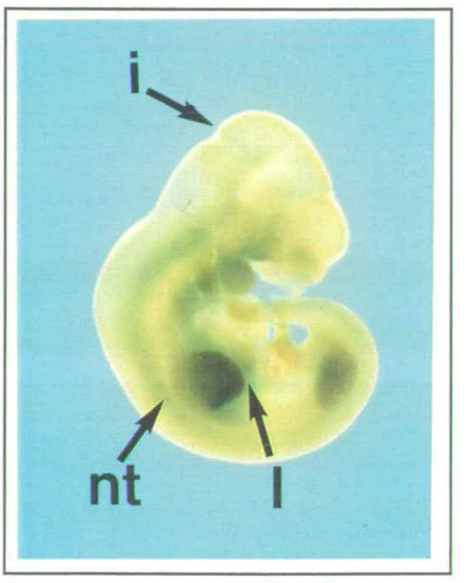
(F) JST213. **I** Ventral view of 8.5dpc embryo shows high level of expression in the region of the foregut which corresponds to the liver diverticulum (ld). **II** A dorsal view of the same embryo shows high levels of expression in the closing head folds (hf). **III** Lateral view of 10.5dpc embryo reveals expression throughout the neurectoderm with higher levels in the ventral neural tube (nt) and at the border of the hindbrain and midbrain, called the isthmus (i). The highest level of expression is seen in the liver (l). **(G)** JST191. **I** Lateral view of 11dpc embryo. High levels of LacZ expression are seen in the nasal placode (np), the limb buds (lb) and between the first and second branchial arches. Low level staining in seen in the neural tube. Sections reveal high levels of expression in the spinal, trigeminal and facial ganglia (sg, tg & fg). **II** Dorsal view of forelimb bud with anterior at the top of the figure. Two regions of high expression are seen in the proximal region of the limb bud in both a posterior (P) and anterior (A) region. In more distal regions expression is absent from the apical ectodermal ridge but is seen in the mesenchyme (mes) surrounding forming digits. **III** Transverse section through the proximal limb bud shows the anterior and posterior regions of expression (dorsal, d and ventral, v). **(H)** JST51. Lateral view of 8.5dpc embryo showing widespread expression. **(I)** JST219. Dorsal and lateral views of 8.5dpc embryos showing widespread expression.



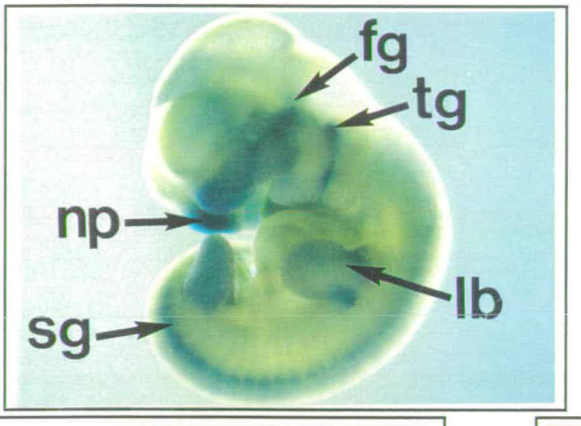
F I



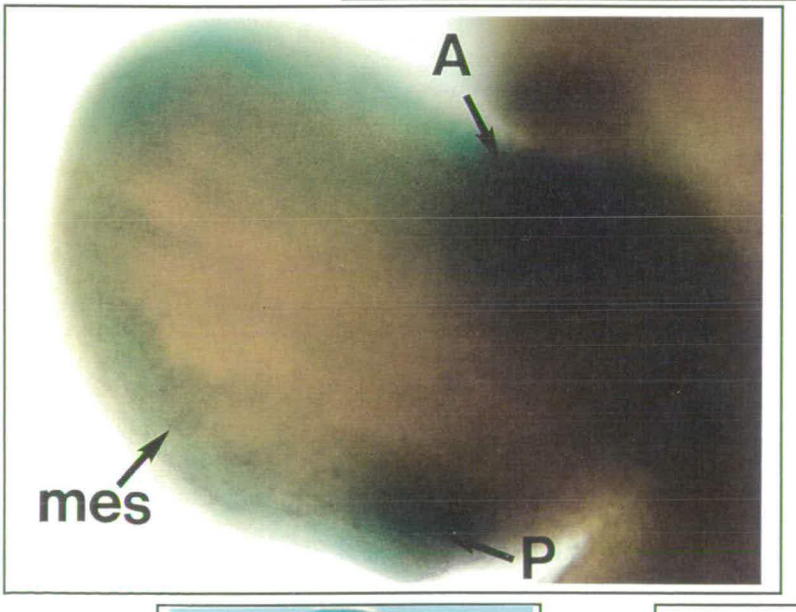
F II



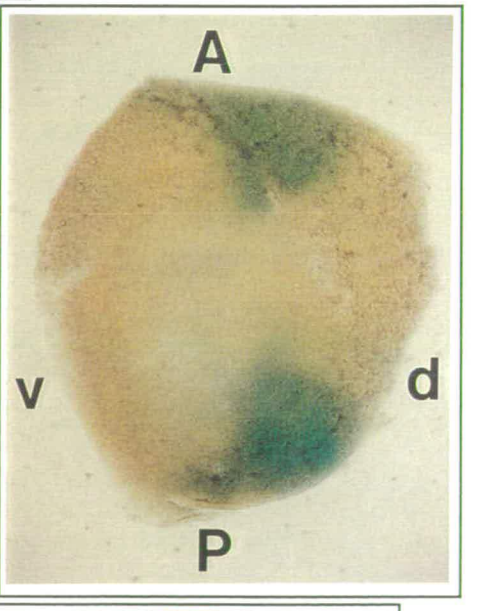
F III



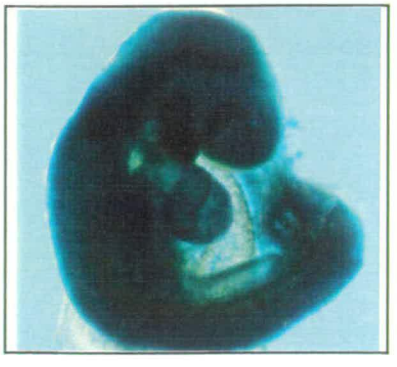
G I



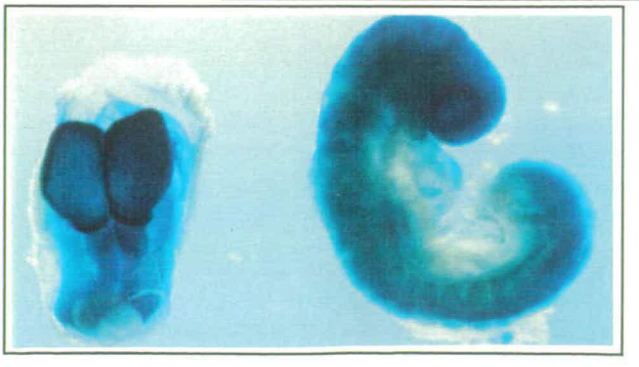
G II



G III



H



I

Table 2.2 Summary of JST expression patterns

Line	Gene	βgal expression pattern
JST51	LDL receptor	Widespread at 9dpc
JST62	novel	At 8.5dpc high levels of expression are seen in the primitive streak ectoderm, lower levels are present in anterior neurectoderm. Expression is absent from extra embryonic tissues. At 9.5dpc expression is restricted to the ventral half of the neural tube.
JST85	novel (myeloid cDNA)	LacZ is expressed throughout embryonic but not extraembryoinc regions. In most chimeras the anterior neural folds failed to close. The severity of the head phenotype is proportional to the level of LacZ expression and therefore proportional to the contribution of ES cells to the chimera
JST184	novel	At 9pc high levels of expression are seen in the developing eye and in the intermediate mesoderm. Lower levels of expression are seen elsewhere in the neurectoderm and brain but expression is absent from the branchial arches. A day later expression continues at high levels in the eye, the developing urogenital system and the ventral neural tube.
JST185	novel	At 8.5dpc expression is restricted to lateral mesoderm which provides mesenchyme for the developing viscera. At 9.5 and 10.5dpc expression is seen in the mesenchyme surrounding endodermal epithelia. Lower levels of expression are seen in the ventral midline of the neural plate. (Further details in Chapter 4).
JST191	novel	AT 11dpc high levels of LacZ expression are seen in the nasal placode, the limb buds and between the first and second branchial arches. Low level staining is seen in the neural tube. Sections reveal high levels of expression in the spinal ganglia, the trigeminal ganglia and in specific cranial ganglia. Two regions of high expression are seen in the proximal region of the limb bud in both a posterior and anterior domain. In more distal regions expression is absent from the apical ectodermal ridge but is seen in the mesenchyme surrounding forming digits.
JST197	non-muscle myosin	At 10.5dpc expression is restricted to the heart.
JST203	β1,4 galactosyltransferase	At 8.5dpc expression is restricted to the border of the closing yolk sac. By 10.5dpc expression is restricted to the developing heart. Sections show lower levels expressed in the liver. Higher magnification shows expression in isolated round cells of unknown origin throughout the embryo, they are often found clustered around blood vessels suggesting they may be blood or capillary cells. Most of the chimeras are severely abnormal.
JST213	novel (DNA-J like)	At 8.5dpc high levels of expression are found in the region of the foregut which corresponds to the liver diverticulum. Expression is also seen at the edges of the closing neural folds. By 10.5dpc expression becomes more widespread and is seen throughout the neurectoderm with higher levels in the ventral neural tube and at the border of the rhombencephalon and midbrain. The highest level of expression is seen in the liver.
JST219	novel *fat*-like cadherin	Widespread expression at 8.5 and 9.5 dpc. Some chimeric embryos are abnormal.

Eight JST lines were injected into blastocysts and tested for germline transmission of which only three lines were successfully transmitted; JST51 (LDL receptor), JST185 (novel) and JST213 (DNA-J like) (Table 2.3). This rate of germline transmission compares unfavourably with a previous experiment using the same CGR8 ES cell line. 1 germline male was produced per 20 blastocysts injected and only 40% of lines contributed to the germline, whereas previously 1 germline male was produced per 10 blastocysts injected and 80% of lines contributed to the germline (W.C. Skarnes, unpublished results). This drop in germline efficiency is thought to result from the later passage ES cells used in this experiment than previously. Culture may bias towards the growth of a population of germline incompetent ES cells. This hypothesis is borne out by the fact that a clonal subline of earlier passage CGR8 cells gives a much higher efficiency; 1 germline male is produced per 5 blastocysts injected and 90% of gene trap lines contribute to the germline (W.C. Skarnes, unpublished results).

The JST51 insertion was bred to homozygosity and, in agreement with published data for a null mutation in the LDL receptor, homozygous animals were viable and fertile (data not shown). JST213 homozygous animals were also found to be viable and fertile. The JST185 insertion results in neonatal lethality and is presented in Chapter 4.

2.9 Discussion

Insertions detected in type I, type II and multiple membrane-spanning proteins

In the original report describing the secretory trap vector, 6 out of 6 insertions occurred in extracellular domains of secreted and type I membrane-spanning proteins showing that the vector selects for insertions in genes encoding N-terminal signal sequences (Skarnes *et al.*, 1995). The results presented in this chapter now extend

Table 2.3 Summary of information on JST lines

Line	Insertion site[f]	Gene	Gene function	Expression pattern in chimeras	Germline transmission	Phenotype
Secretory proteins						
JST51	coding (intron)	LDL receptor	uptake of lipoproteins into cell	widespread	yes	viable
JST62	coding (intron)	myeloblast cDNA	unknown (multiple TM domains)	restricted	no	nd
JST185	coding (intron)	novel	unknown (multiple TM domains)	restricted	yes	perinatal lethal
JST191[a]	coding (intron)	novel	unknown (proline rich protein)	restricted	no	nd
JST203	coding (intron)	β1,4-galactosyltransferase	multiple (inc. cell adhesion)	restricted	not tried	neonatal lethal[d]
JST213	coding (intron)	DNA-J2 homologue[b]	molecular chaperone	restricted	yes	viable
JST219	coding (intron)	*fat*-like cadherin	cell adhesion molecule	widespread	no	nd
Non-secretory proteins						
JST71	coding (intron)	heat shock protein 86	unknown	nd	not tried	nd
JST85	coding (intron)[c]	myeloid cDNA	unknown (contains ubiquitin ligase domain)	widespread	no	nd
JST136	5'UTR (intron)	p23	complexes with progesterone receptor	nd	not tried	nd
JST197	5'UTR (intron)	non-muscle myosin heavy chain	mediates actin-based motility	restricted	no	nd
JST232	5'UTR (intron)	RBP-Jκ/Su(H) pseudogene	transcription factor pseudogene	no stain	not tried	nd
Unknown protein						
JST184	5'UTR (exon)[e]	novel	unknown	restricted	not tried	nd

a βgal staining in ES cells exhibits a typical secretory pattern
b predicted to have a transmembrane domain
c endogenous gene in a different frame to reporter
d reported in Lu et al., 1997
e 2.3 kb of vector deleted prior to insertion
f coding intron signifies an insertion into an intron within the open reading frame
nd not determined
TM transmembrane

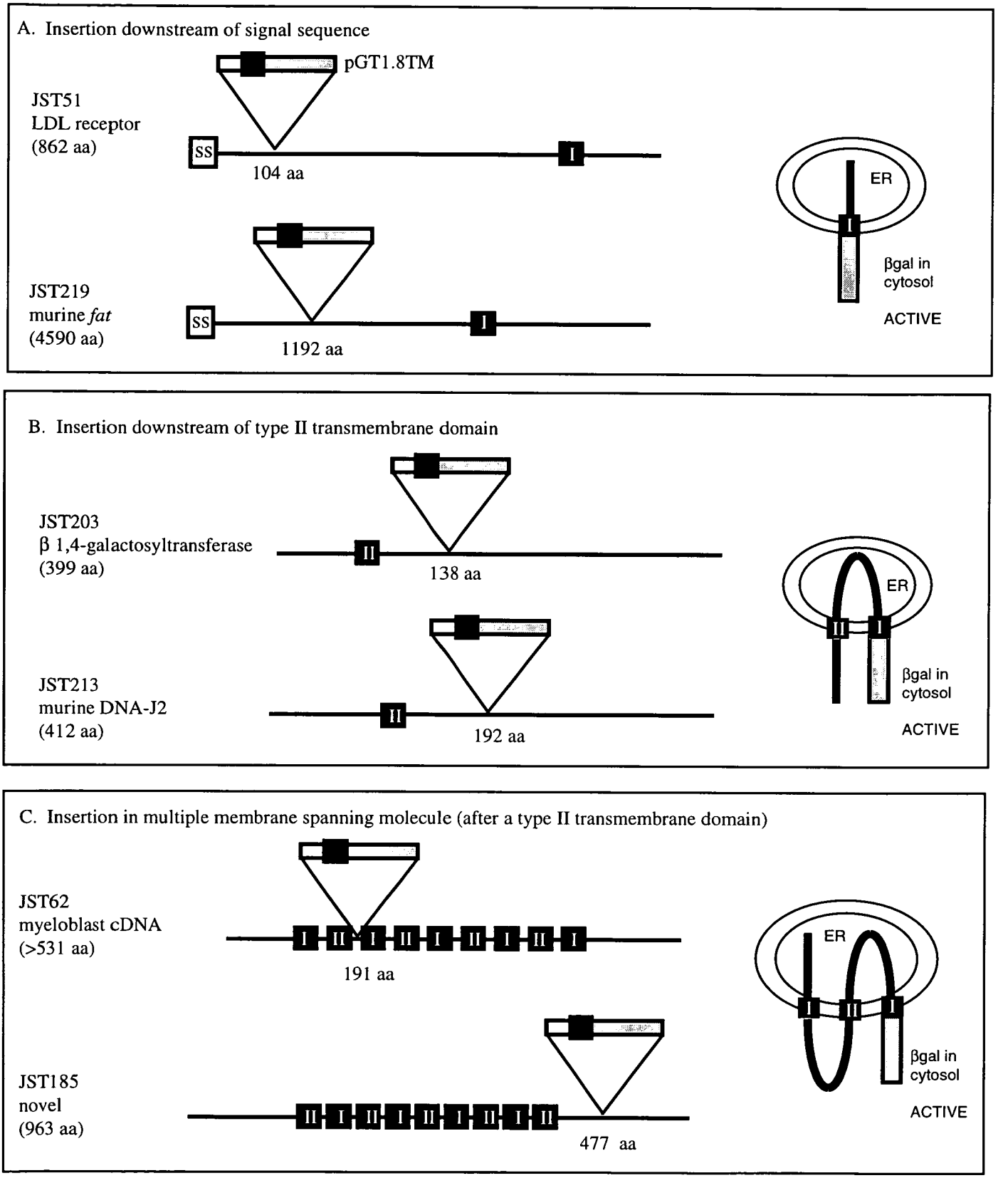


Figure 2.10 Membrane topology of JST fusion proteins
Fusion proteins which place the vector CD4 TM domain in a type I orientation in the ER membrane are thought to expose βgal to the cytosol where it is active. In single membrane spanning proteins insertions are detected if they occur downstream of a signal sequence (**A**) or downstream of a type II transmembrane domain (**B**). If the insertion occurs in a multimembrane spanning molecule it must occur downstream of a type II domain to allow βgal to be detected (**C**).

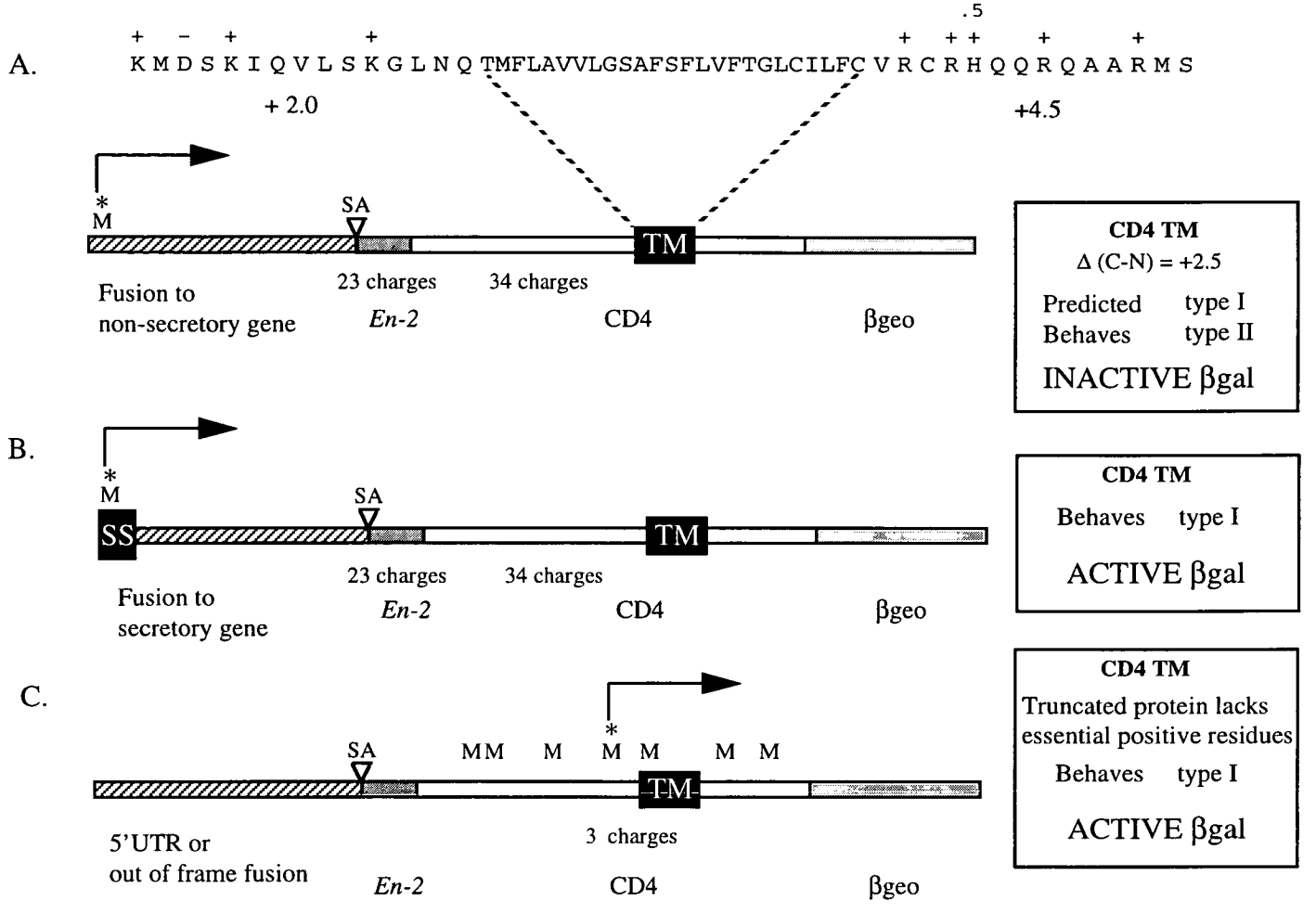


Figure 2.11. Membrane topology of CD4 fusion proteins

(**A**) The difference in the charge between the first 15 residues flanking a transmembrane domain can be used to predict its orientation, with the more positive portion facing the cytosol (Hartmann *et al*, 1989). Arginine (R) and lysine (K) are given a value of +1, histidine +0.5, and aspartatic acid (D) and glutamic acid (E) -1. The charge differential, Δ(C-N), for CD4 is +2.5. This positive value predicts that CD4 is a type I transmembrane domain. However, the highly charged N-terminal portion of the vector, which includes 23 charges from *En-2* and 34 charges from CD4, makes the transmembrane domain behave in a type II manner. Inframe fusions to non-secretory genes will result in the βgal reporter downstream of the TM being placed in the lumen of the ER where it is inactive. (**B**) Fusions to secretory molecules containing a signal sequence are always placed in a type I orientation which places βgal in the cytosol where it is active. (**C**) Fusions to 5'UTRs or those which are out of frame with the reporter can only be detected if translation is initiated within the vector. A Kozak consensus methionine upstream of the transmembrane domain produces a truncated fusion protein which only contains 3 charges in its N-terminus. The paucity of charges allows the N-terminus of the fusion protein to be translocated across the ER membrane leaving βgal in the cytosol where it is active. In this manner insertions into non-secretory molecules can be detected.

the range of secretory molecules to include type II membrane receptors as well as multiple membrane-spanning proteins (Table 2.3). As was predicted for N-terminal signal sequences, insertions downstream of a type II membrane-spanning region are also expected to place ßgeo in the cytosol where it is active (Figure 2.10). The same holds true for multiple membrane spanning proteins if the vector inserts downstream of a type II membrane-spanning region.

Insertions detected in non-secretory proteins

My results also indicate that the secretory trap vector enriches but does not absolutely select for insertions in secretory molecules. Five out of twelve JST insertions were associated with non-secretory molecules. It is noteworthy that three insertions had occurred in 5' untranslated region of the target gene and a forth was out of frame with the target gene. This suggests that ßgal protein is produced via internal initiation of translation from one of the 7 in frame methionines within the CD4 sequences of the vector. One of the methionines located upstream of the CD4 transmembrane domain has an excellent Kozak consensus (Kozak, 1987). Although internal initiation of translation at this site offers an explanation for how the reporter might be made such a product should be inactive.

To investigate this apparent anomaly the CD4 transmembrane domain of the vector was re-examined. The Hartmann charge-difference rule predicts that CD4 should in fact adopt a type I orientation in the membrane (Figure 2.11 and Hartmann, 1989). This contradicts the entire basis of the selective property of the vector as the CD4 transmembrane domain is thought to place fusions that do not acquire a signal sequence in a type II orientation by default. However, not all transmembrane domains fit the charge-difference rule: analysis of 126 proteins revealed 7 proteins that do not fit the rule (Hartmann, 1989). All 7 proteins were predicted to be type I but in reality behaved as though they were type II. Most of the exceptions contained a total of 17 or

more charged residues in their N-termini. These highly charged segments are thought to prevent the translocation of the amino terminus across the membrane thus placing the protein in a type II orientation. Indeed the addition of positive charged residues to the type I membrane protein cytochrome P-450 has been shown to convert it to a type II protein (Szczesna-Skorupa *et al.*, 1995). In the secretory trap vector both CD4 and *En-2* contribute highly positively charged segments to the N-terminus of the fusion protein (Figure 2.11A). The presence of this charge may account for the fact that the CD4 transmembrane domain behaves as a type II domain within the context of the vector.

Internal initiation of translation would lead to the production of a truncated fusion protein which lacks positive charges encoded by the N-terminal part of the vector (Figure 2.11). Removal of these positive charges may cause the CD4 transmembrane domain to revert back to type I behaviour. Similar types of reversion have been documented but they were only partial. For example inversion of the charge difference between the segments flanking the membrane spanning domain of the asialoglycoprotein (ASGP) from -3 to +5 yielded mixtures of type II and type I topologies (Beltzer *et al.*, 1991). Thus internal initiation of translation allows insertions into non-secretory molecules to be detected as βgal activity no longer depends on the acquisition of a signal sequence. The βgal activity in such lines tends to be expressed at very low levels and may reflect a mixture of active and inactive enzyme topologies (Figure 2.3). Only those lines showing a clear "secretory" pattern were selected in the preliminary study, explaining why no insertions were detected in non-secretory molecules (W.C. Skarnes, pers. comm).

It is unclear why the in-frame insertion in the heat shock protein (hsp) 86 in line JST71 produces detectable βgal activity. However, TMpred predicts a type II membrane-spanning domain in hsp86 upstream of the insertion site which would place the fusion protein in an active type I orientation. Alternatively there is evidence to

indicate that the orientation of the transmembrane domain can be affected by the folding state of the N-terminal domain (Denzer *et al.*, 1995). Tightly folded N-terminal domains hinder and can prevent translocation whereas destabilisation or unfolding facilitates translocation. It is possible that the fusion to βgeo may somehow destabilise the N-terminus of hsp 86 allowing it to be translocated.

Developmental regulation of reporter activity in JST lines

All the JST lines which show detectable levels of expression in ES cells were expressed in either a widespread or restricted pattern in embryos during early gestation. The four JST lines which did not express in the embryo include the three intron containing lines and JST232. The latter carries an insertion in a pseudogene of RBP-Jκ/Su(H). Although the endogenous RBP-Jκ/Su(H) is expressed during development the pseudogene is expected not to be expressed as promoter elements are usually missing from pseudogenes. It is not clear why the three lines which produce intron containing transcripts express βgal activity in ES cells, yet do not express in the early embryo. Either the level of expression is higher in ES cells than embryos or the staining method is less sensitive in embryos. However it is clear that these events always represent very low levels of βgal activity.

A number of novel genes associated with interesting developmentally regulated patterns of reporter gene expression were identified in the JST screen. For example, JST62 carries an insertion into a novel multiple membrane spanning molecule which is expressed at high levels in the developing nervous system. JST191 carries an insertion in a novel proline rich protein which is expressed at high levels in the limbs, various ganglia and the nasal placode. JST184 is expressed almost exclusively in the developing eye.

JST213 carries an insertion in a membrane spanning DnaJ-like protein. The prototypical bacterial DnaJ protein is a molecular chaperone which is involved in

protein folding and translocation across biological membranes (Ziegelhoffer *et al.*, 1996). The gene interrupted in JST213 shows 95% identity to a rat and human sequence. Unlike bacterial DnaJ both these sequences are predicted to have transmembrane domains. JST213 also shows 61% identity to a murine EST, Nedd-7 (accession D10917), which was isolated on the basis of its high expression in the mouse neuronal precursor cell (NPC) and its down-regulation in adult brain (Kumar *et al*, 1992). Two human homologues of Dna-J have also been isolated which show restricted expression patterns in the brain, both are preferentially expressed in neurons (Cheetham *et al.*, 1992). Interestingly, the expression of the βgal reporter in 8.5dpc JST213 embryos is detected in the anterior closing neural folds. The JST213 insertion was transmitted through the germ-line and homozygous animals are viable and fertile.

JST203 carries an insertion in β1,4-galactosyltransferase (GalTase). Several studies have suggested that GalTase acts as a surface receptor for extracellular ligands during a variety of cellular interactions (Hathaway & Shur, 1996 and references within). To date, the βgal reporter in JST203-derived chimaeric embryos is expressed solely in the developing heart. Studies have shown that GalTase is expressed in individual cells of the morula before compaction and later in migrating neural crest cells (Bayner *et al.*, 1988; Hathaway and Shur, 1992). Intriguingly, mice carrying a targeted mutation in GalTase die a few weeks after birth due to a pituitary insufficiency (Lu *et al.*, 1997). Unfortunately the JST203 insertion was not transmitted through the germline. Further studies should address whether βgal is also detectable in the pituitary.

JST219 carries an insertion in the murine homologue of the *Drosophila* tumour suppressor *fat* which is essential for controlling cell proliferation during *Drosophila* development (Mahoney *et al.*, 1991). The gene encodes a large transmembrane protein with 34 tandem cadherin repeats, five EGF-like repeats, and a laminin A-G domain. JST219 is 88% identical to the human homologue of *FAT* which is expressed

in epithelia and in some mesenchymal compartments (Dunne *et al.*, 1995). Higher levels of human *FAT* expression were observed in fetal, as opposed to adult tissue, suggesting that its expression may be developmentally regulated in these tissues. Surprisingly, the reporter in JST219 was expressed ubiquitously and at very high levels in embryos analysed between 8.5 and 10 dpc. Line ST519, which was identified in the original secretory trap screen, carries an insertion in another *Fat*-like cadherin (Skarnes *et al.*, 1995). ST519 is 81% identical to *FAT*. Unlike JST219, the reporter in ST519 is not expressed in ES cells, suggesting that these lines carry insertions in different but related genes. Several attempts have been made to transmit both these insertions through the germline but none have been successful. This failure may reflect the compromised germline transmission rate of the particular ES cells used in the electroporation, or may reflect a role for these *Fat*-like molecules in early development which is sensitive to a gene dosage effect. Alternatively the mutant *Fat*-like molecules carrying the secretory trap insertions may be exerting a dominant negative effect on the endogenous molecules.

In summary, the results presented in this chapter validate the secretory trap approach as a means to identify novel secretory molecules involved in mouse development. In future screens, secretory trap lines which do not show a clear secretory pattern of βgal activity should not be analysed, as these often represent insertions in non-secretory genes. An alternative approach to eliminate non-secretory insertions would be to remove internal methionines from within vector sequences. This may prevent the detection of insertions into 5'UTRs which make up the majority of non-secretory insertions.

Chapter 3

EXON TRAP SCREEN

3.1 Introduction

The requirement for pGT1.8TM to insert into introns is likely to favour insertions into genes composed of large intronic regions. Consequently, the vector will be biased against detecting insertions in genes containing few or no introns. There are several reports in the literature of intronless genes which play roles in cell signalling. Many of the genes encoding G protein-coupled receptors including members of the somatostatin receptor family (Patel, 1995) and serotonergic receptor family (Li *et al.*, 1995) are intronless. Coding sequences of the recently discovered olfactory multigene family of serpentine receptors lack introns although members do contain introns in untranslated regions (Drutel *et al*, 1995). A recent report suggests that imprinted genes, including the secreted growth factor IGF2, contain fewer and smaller introns than non-imprinted genes (Hurst *et al.*, 1996). Since genes composed of few or no introns represent a class of molecule that cannot be accessed with the

intron targeting vector, a new vector was designed to access insertions in these genes. This chapter describes the construction and preliminary experiments with an exon trapping version of the secretory trap vector, pET1.8TM, which relies on insertion into the exons of secreted genes.

3.2 Construction of exon trapping vector, pET1.8TM

The pGT1.8TM vector contains the *En-2* splice acceptor site upstream of the CD4 transmembrane domain and the βgeo reporter. To construct the exon trap version of pGT1.8TM the En-2 sequences were deleted by digestion with Hind III and Bgl II. The vector backbone was gel purified and circularised by ligating adapters which re-introduce a unique Hind III site. The new vector, pET1.8TM, now relies on inserting into the exons of genes in the correct orientation and reading frame to generate a fusion transcript (Figure 3.1).

3.3 Comparison between pGT1.8TM and pET1.8TM electroporations

To compare the efficiency of the original and derivative vectors, pET1.8TM and pGT1.8TM, parallel electroporations were carried out under identical conditions. Vectors were electroporated into 5 x 10^7 CGR-8 ES cells. As expected, the pET1.8TM electroporation resulted in a reduction in G418 resistant colonies compared to pGT1.8TM (123 versus 205). Similar to previous observations 28% of pGT1.8TM colonies stained for βgal activity whereas 47% of pET1.8TM colonies stained (Table 3.1). Most of the βgal positive pGT1.8TM lines showed the characteristic secretory pattern of βgal expression, in multiple cytoplasmic inclusions and in the peri-nuclear compartment of the cell. Surprisingly, only 4 pET1.8TM colonies showed this pattern, the remaining 54 displayed a variety of subcellular localisations (Figure 3.1). Roughly equal numbers of colonies showed nuclear, cytoplasmic, single dots and dots with cytoplasmic staining patterns (Figure 3.2). The different patterns of βgal activity are thought to reflect the differential sorting properties of the endogenous genes

Figure 3.1 Comparison of the structures of gene trap and exon trap vectors. The exon trap vector, pET1.8TM, is derived from pGT1.8TM by removal of the *En2* splice acceptor sequences.

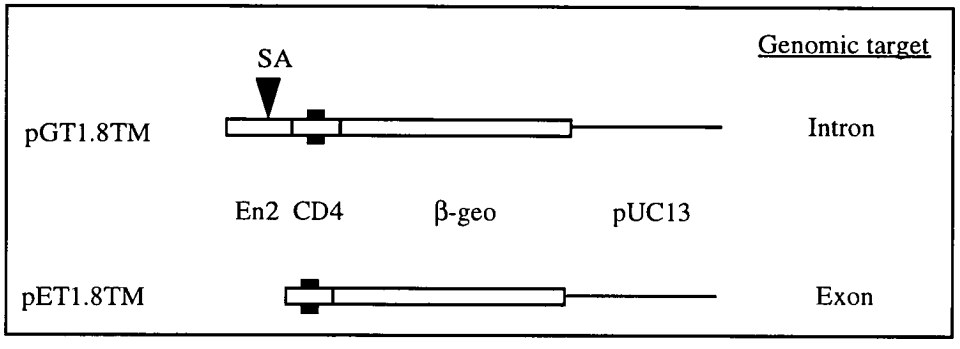
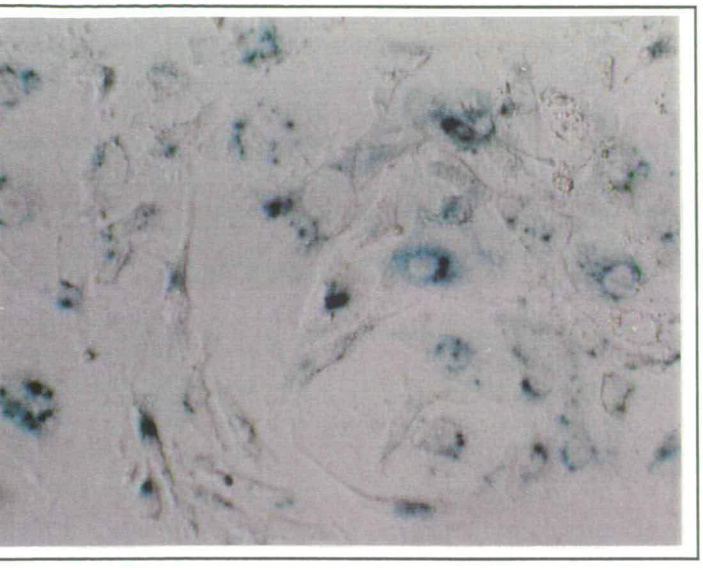


Table 3.1 Colony numbers generated with pGT1.8TM and pET1.8TM

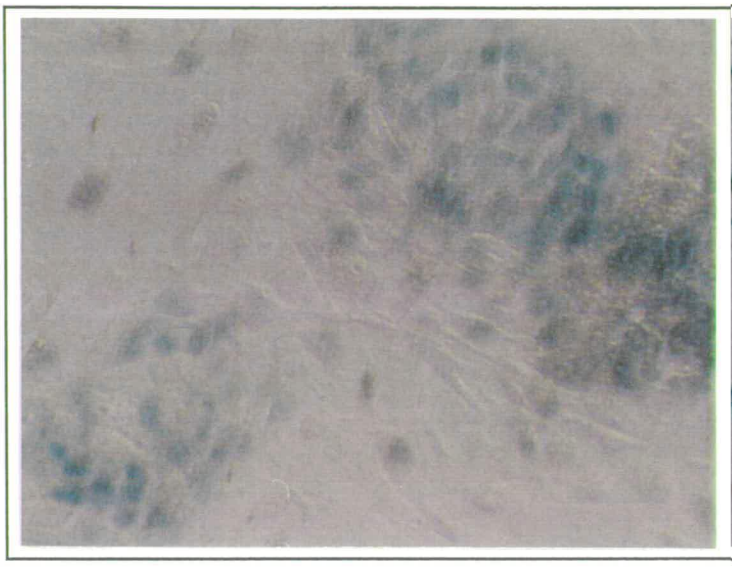
Number of colonies	pGT1.8TM	pET1.8TM
neo^R	205	124
β-gal staining	57 (27%)	58 (47%)
secretory pattern	44 (21%)	4 (3%)
other patterns	13 (6%)	54 (44%)

Figure 3.2 βgal activity ES cell lines carrying pET1.8TM insertions

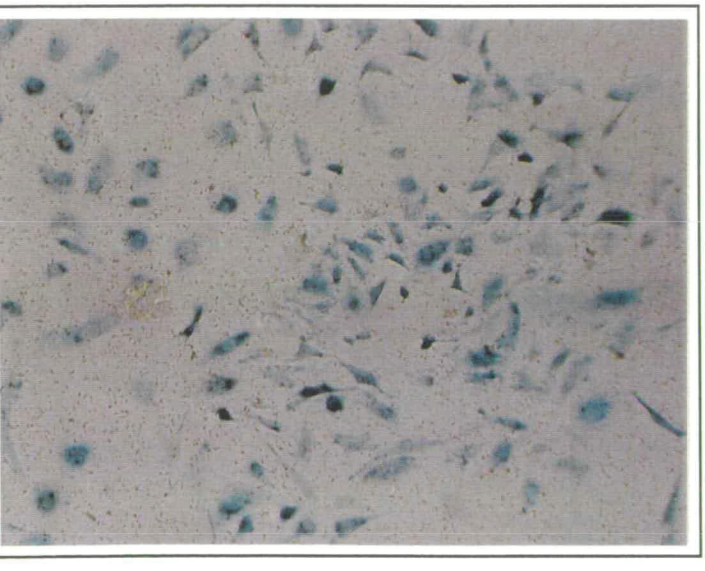
ES cell colonies carrying insertions of pET1.8TM show βgal activity in a variety of subcellular localisations. **(A)** 'Secretory' pattern (ET74) **(B)** Nuclear (ET71) **(C)** Cytoplasmic (ET95) **(D)** Single dots per cell (ET50). **(E)** Shows small patches of expression (ET113) and **(F)** shows larges patches of stain with a tight border seperating the expressing and non-expressing cells (ET98).



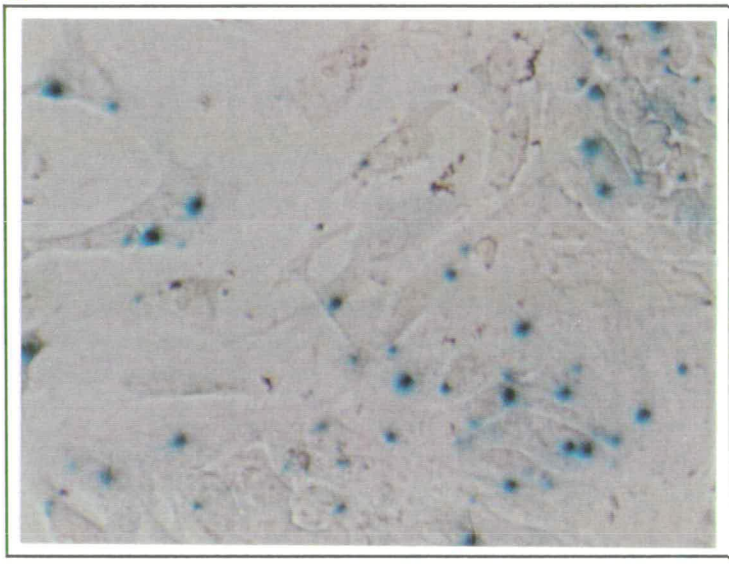
A



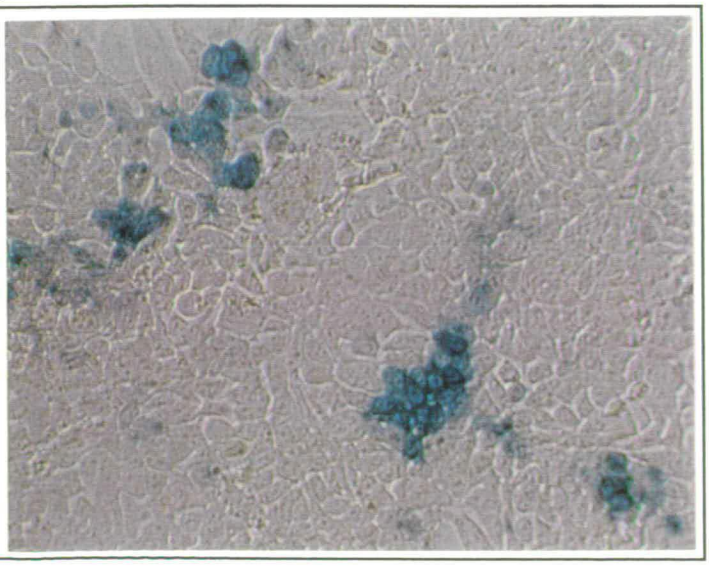
B



C



D



E

F

carrying insertions. Such a variety of patterns was unexpected and suggested that the selective property of the transmembrane domain in pET1.8TM has been lost, so that insertions were possibly being detected in non-secretory molecules. Unfortunately, the 4 colonies showing the secretory pattern were lost to contamination and could not be studied further.

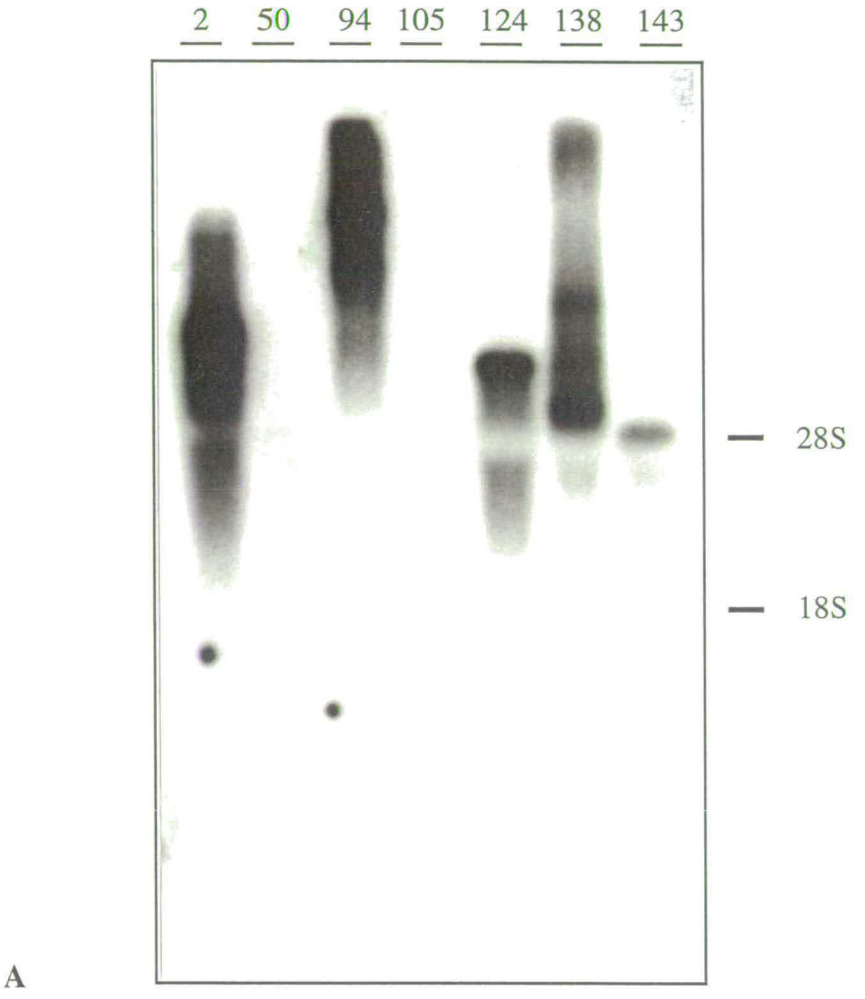
3.4 ET lines produce fusion transcripts to endogenous genes

Total RNA was prepared from 16 ET lines expressing detectable βgal and analysed by Northern blot hybridisation with a Lac Z probe. A single major fusion transcript of varying size hybridises to lac Z in 11 out of the 16 lines analysed. Short auto radiographic exposures of Northern blots reveal that in the remaining 5 lines a complex array of minor transcripts hybridise as well as the major transcript (Figure 3.3 and data not shown). Such a pattern of transcripts is often seen when gene trap insertions occur in rRNA genes and undergo *trans*-splicing (Figure 2.4). The complex array of transcripts may reflect a variety of *trans*-spliced fusion transcripts (J.E. Sleeman & W.C. Skarnes, unpublished results).

Fusion transcripts varied in size from 3.5 kb to > 12 kb, although most (9 out of 16) were around 5 kb. The contribution from the vector to the fusion transcript is estimated to be 3.8 kb (0.68 kb of CD4 fused to 3.03 kb of βgeo and the 0.1-0.2 kb poly A tail). In ET93 over 7 kb of endogenous sequence contributed to the fusion transcript confirming that βgal activity can be detected in the presence of large N-terminal fusions. In most cases, less than 700 bp of endogenous sequence contributed, suggesting that insertions tended to occur towards the 5' ends of genes (Table 3.2). In ET105, the transcript was estimated to be 3.5kb. As this is 300 bp less than the minimum contribution expected from the vector, this suggests that sequences were deleted from the vector prior to insertion in the genome.

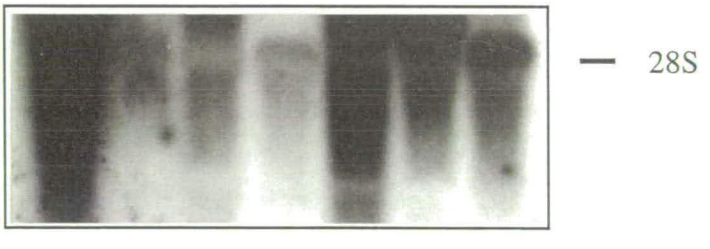
Figure 3.3 Northern blot analysis of ET lines

20μg of total RNA was separated on a denaturing agarose gel and analysed by Northern blot hybridisation to a LacZ probe. (**A**) 24 hour exposure to autoradiographic film. Three out of the seven lines contain a complex array of fusion transcripts (**B**) Seven day exposure. The remaining four lines contain a single fusion transcript between 4.5 and 6kb in size. (**C**) Loading control. Ethidium bromide stained gel.



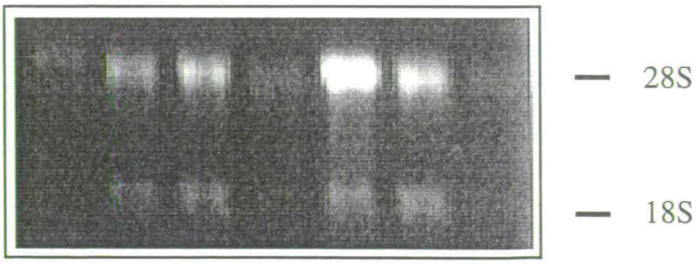
A

LacZ



B

LacZ



C

EtBr

Table 3.2. Summary of information on pET1.8TM lines

Cell line	βgal in ES cells	βgal in 9.5dpc chimeras	Vector deleted (kb)[a]	Lac Z fusion transcript (kb)[b]	Endogenous gene contribution to fusion transcript (kb)[c]	Method of identification[d]	Insertion site
ET 2	dots & cytoplasm	restricted	0.528	8	3.70	L	5'ETS
ET 8	cytoplasm	widespread	nd	nd	nd	nd	nd
ET 50	dots	nd	0.521	5	0.7	L	novel
ET 80	nuclear	nd	0.300	5	0.5	R	novel (ORF)
ET 82	nuclear	widespread	nd	nd	nd	nd	nd
ET 94	dots & cytoplasm	nd	0.496	>12	>7.7	L	28S rRNA
ET 105	dots	restricted	0.365	4.5	0.1	L + R	novel (5'UTR)
ET 111	dots & cytoplasm	no stain	nd	nd	nd	nd	nd
ET 119	nuclear	widespread	nd	nd	nd	nd	nd
ET 124	dots & cytoplasm	nd	0.559	6	1.8	R	28S rRNA
ET 138	cytoplasm	nd	0.449	5	0.6	L + R	novel (5'UTR)
ET 143	dots	restricted	0.565	4.8	0.6	R	novel (5'UTR)

a Determined by sequencing LM-PCR products except for ET 80, 124 and 143 where amount deleted is estimated from RACE sequence

b Estimated from Northern blot by plotting the migration of the rRNAs against molecular weight and extrapolating

c Calculated by subtracting the amount of vector deletion (a) from the 4.8 kb expected to be contributed by the exon trap vector and then subtracting this value from the size of the fusion transcript (b)

d Insertion sites identified by LM-PCR (L) or 5'RACE (R)

nd not determined

3.5 Identification of insertion sites reveal vector deletions

Initially 5'RACE was used to identify the endogenous genes in 4 lines using a lacZ first strand primer. However, none gave rise to second round PCR products. Indeed, only one, ET80, gave rise to first round products. Use of a more 3' primer gave rise to 2nd round RACE products for ET80 which were cloned. Sequencing confirmed that vector sequences were missing from the amplified cDNA. The absence of 5' vector sequences may be due to the use of a cryptic splice acceptor within the vector in ET80, or may reflect a deletion from the vector before it inserted into the genome. To investigate these possibilities insertion sites in exon trap lines were identified from genomic DNA.

A ligation mediated PCR approach (LM-PCR, Garrity & Wold, 1992) was adapted to allow direct sequencing of junction fragments which span the insertion site (Figure 3.4). Sequence information should reveal whether or not deletions occurred from the vector before insertion in the genome. In addition, owing to the requirement of pET1.8TM to insert into exons, it should be possible to identify coding sequence of interrupted genes directly from junction fragments.

Before attempting LM-PCR, ET lines were confirmed to carry single vector insertion sites. Genomic DNA was digested with Hind III, which does not cut within the vector, and analysed by Southern hybridisation with a neo vector probe. A single band whose size varied between lines hybridised to the probe suggesting that each line carried a single insertion site (Figure 3.5A). A frequent cutting restriction enzyme, NlaIII, was then used to generate junction fragments for amplification. NlaIII was chosen as it cuts within the 5' region of lacZ but does not cut in CD4 sequences. Digested genomic DNA was analysed by Southern blot hybridisation using a lacZ probe located 5' to the NlaIII site. Junction fragments between 0.5 and 3 kb, which

Figure 3.4 Ligation mediated PCR.

Genomic DNA is digested with a frequent cutter enzyme to produce a junction fragment that spans the secretory trap insertion site. Digested DNA is denatured and blunt ended molecules are synthesised. A unidirectional adapter is ligated and the DNA denatured. As the adapter is non-phosphorylated only the top adapter strand (5'-3') is bound to the DNA. A nested vector primer is used to synthesize DNA which forms the template for a first round of PCR amplification. The second round PCR uses a biotinylated vector primer so that amplified products can be bound to magnetic streptavidin coated beads, denatured and directly sequenced without the need for cloning.

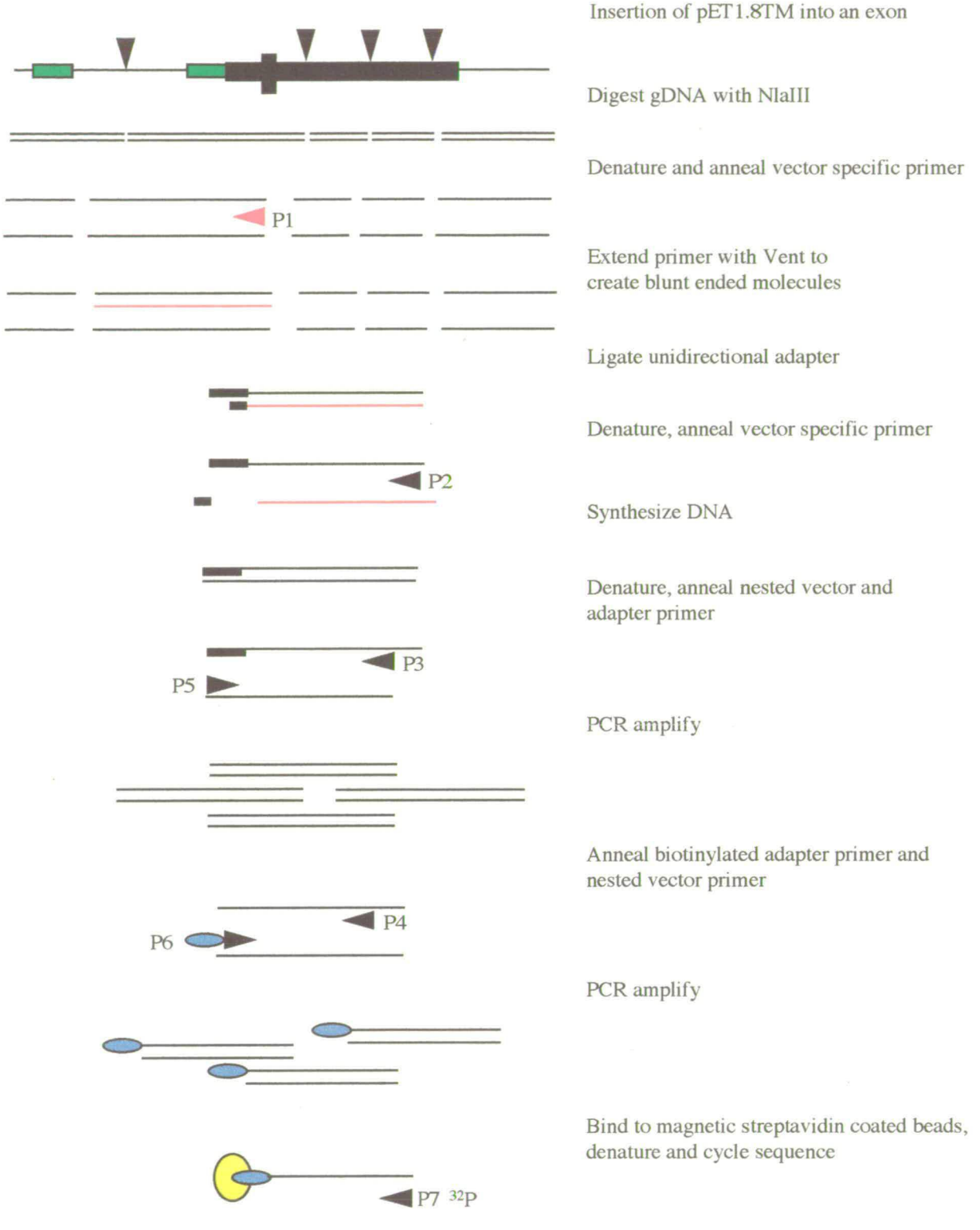
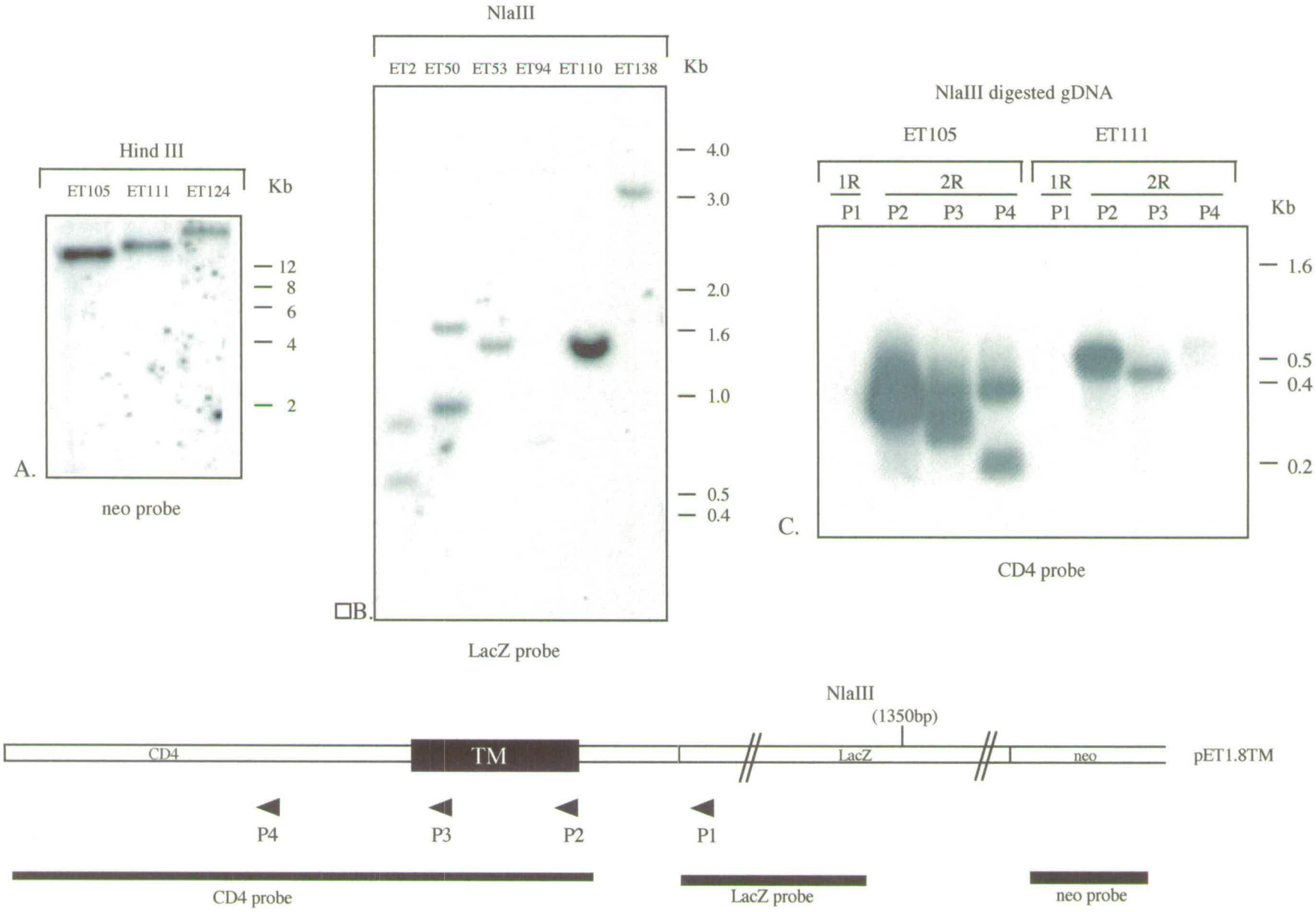

Insertion of pET1.8TM into an exon
Digest gDNA with NlaIII
Denature and anneal vector specific primer
P1
Extend primer with Vent to
create blunt ended molecules
Ligate unidirectional adapter
Denature, anneal vector specific primer
P2
Synthesize DNA
Denature, anneal nested vector and
adapter primer
P3
P5
PCR amplify
Anneal biotinylated adapter primer and
nested vector primer
P4
P6
PCR amplify
Bind to magnetic streptavidin coated beads,
denature and cycle sequence
P7 32P
Key
pET1.8TM
exon/intron structure
NlaIII site
5'-3' primer
non-phosphorylated
adapter oligo
5' biotin
streptavidin

Figure 3.5 Ligation mediated-PCR on ET lines

(A) Schematic representation of primers and probes used for LM-PCR, numbering of primers is based on Figure 3.4. **(B)** Genomic DNA prepared from ET ES cells digested with Hind III, an enzyme which does not cut within the vector. Southern hybridisation with a neo probe reveals a unique fragment containing the vector in each line. **(C)** Genomic DNA digested with NlaIII and analysed by Southern hybridisation using a LacZ probe. Hybridising bands indicate the size of NlaIII junction fragments which span the insertion site. **(D)** First round LM-PCR products were re-amplified using nested primers (P2-P4 shown on the diagram). Products were analysed by Southern hybridisation with a CD4 probe. In ET105 all three primer sites are present in the vector whereas in ET111 the p4 primer site has been deleted from the vector. The larger second band seen in ET105 2nd round PCR products is due to re-amplification of the junction fragment with the first round primer pair which are present in the reaction.


Hind III
ET105 ET111 ET124
Kb
12
8
6
4
2
A.
neo probe
NlaIII
ET2 ET50 ET53 ET94 ET110 ET138
Kb
4.0
3.0
2.0
1.6
1.0
0.5
0.4
B.
LacZ probe
NlaIII digested gDNA
ET105
ET111
1R
2R
P1 P2 P3 P4
Kb
1.6
0.5
0.4
0.2
C.
CD4 probe
NlaIII
(1350bp)
CD4
TM
LacZ
neo
pET1.8TM
P4
P3
P2
P1
CD4 probe
LacZ probe
neo probe

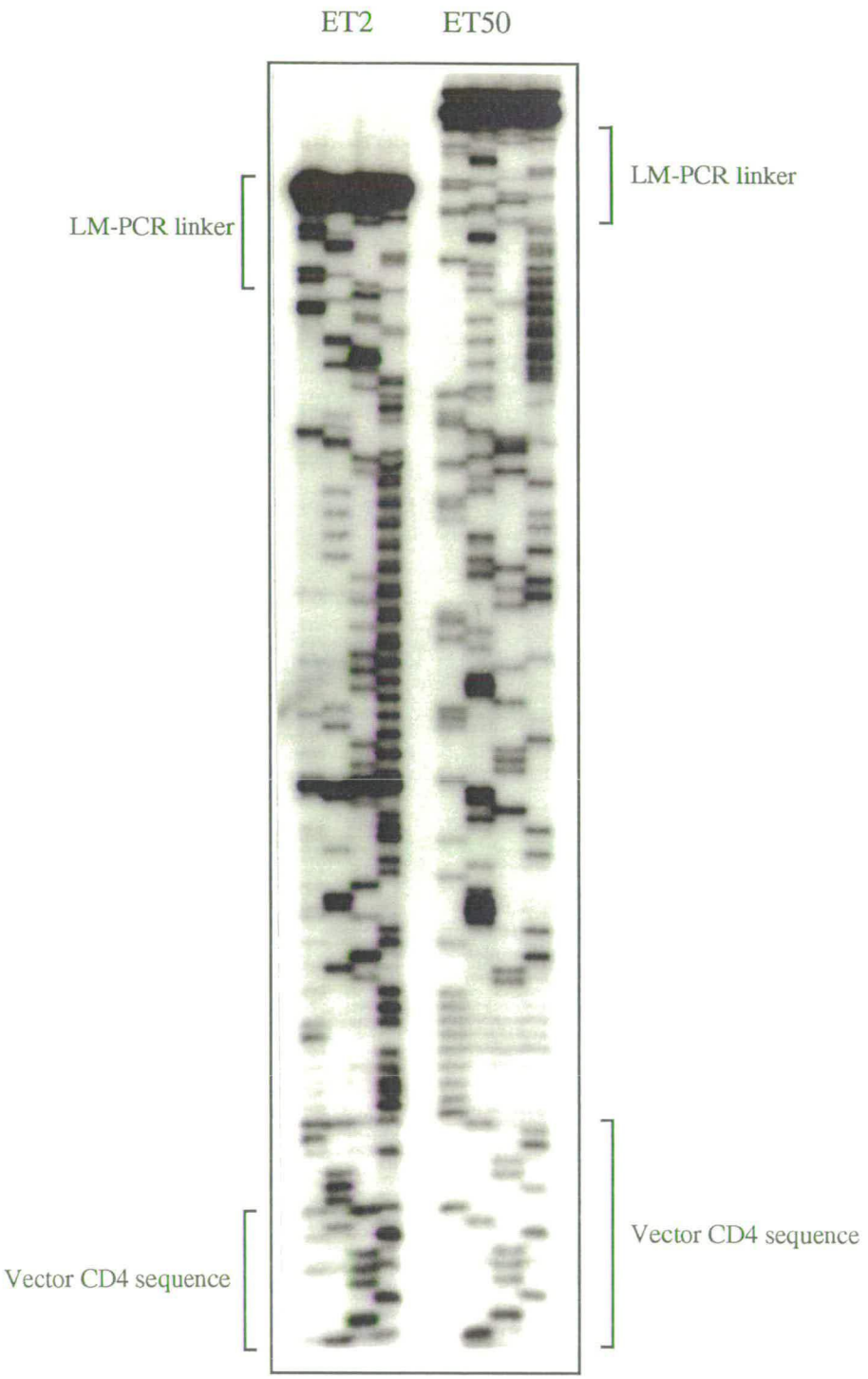
are a suitable size for amplification, were produced (Figure 3.5C). Two fragments hybridised in some lines suggesting that multiple copies of the vector had inserted.

LM-PCR was performed on NlaIII digested genomic DNA using the primers indicated in Figure 3.5. A primer against lacZ sequence was used to generate the first round products. These were further amplified using one of a set of three diagnostic primers. The presence or absence of products hybridising with CD4 on a Southern blot was used to estimate the amount of CD4 sequence that had been deleted from the vector prior to insertion (Figure 3.5C). This enabled a suitable primer to be selected for direct sequencing from second round products (Figure 3.6). In two lines carrying multiple copies of the vector, ET2 and ET50, single endogenous sequences were identified from the junction fragment, suggesting that one or more of the LM-PCR primer sites were absent from subsequent copies of inserted vector. However, multiple copies in ET111 precluded the identification of endogenous sequences, only sequence corresponding to the 3' end of the exon trap vector could be read upstream of CD4. This indicated that at least two copies of the vector had inserted in a head to tail arrangement.

Sequence information from 5 ET lines revealed that between 365 and 528 bp had been deleted from the 5' end of the vector prior to its insertion in the genome (Table 3.2). Endogenous sequences are recorded in Appendix III. They were compared to the non-redundant GenBank database using the BLASTN program. Two of the insertions occurred in rRNA genes. Sequence from the remaining three lines failed to generate a significant match with any database entry. The lack of homology may be due to the vector having inserted towards the 5' end of an exon. In this situation the sequence flanking the insertion site would mostly be composed of intron, this would result in insufficient coding sequence to generate a match to any database entries. The observation that the endogenous gene sequences from 3 lines lack open

Figure 3.6 Direct sequencing of LM-PCR products

Direct sequencing of ET2 and ET50 LM-PCR products. Length depends on the position of the NlaIII site within genomic sequences. Various amounts of vector CD4 sequence are deleted before insertion. In ET2 528bp has been deleted whereas in ET50 521bp was deleted. The novel sequence upstream of vector CD4 corresponds to genomic sequence flanking the insertion.


ET2
ET50
LM-PCR linker
LM-PCR linker
Vector CD4 sequence
Vector CD4 sequence

reading frames in the same frame as βgal may likewise be explained by insertion at 5'ends of exons.

3.6 5'RACE confirms pET1.8TM inserts into exons

Coding sequence from endogenous genes carrying vector insertions were identified by cloning cDNAs generated from fusion transcripts using 5' RACE. Comparison between the sequences generated by 5' RACE to those generated by LM-PCR from genomic DNA should indicate whether insertions occurred in introns or exons. Identical sequence would suggest an insertion into an exon.

Total RNA was prepared from five ET lines and first strand cDNA was synthesized using a primer complementary to lacZ. In the lines where the extent of vector deletion had not been determined by LM-PCR, a series of PCR amplifications were performed on cDNAs using the same approach as that for LM-PCR. Direct sequencing of second round RACE products revealed that in the two lines where genomic sequences flanking the insertion site were novel, the RACE products gave identical sequence. This suggested that the insertions had occurred in exons. In addition, the lack of an open reading frame suggests that the insertions occurred in UTRs. The 5'RACE sequences are recorded in Appendix 3 and summarised in Table 3.2.

3.7 Expression in ES cell derived chimeras

To identify developmentally regulated genes, the expression patterns of trapped genes were determined by monitoring reporter activity in chimeric embryos. Seven ET lines were selected for analysis on the basis of high levels of βgal expression in ES cells. Chimeric embryos were dissected between 8.5 and 9.5d of development and reporter expression was monitored by staining for βgal activity. Of the 7 lines

analysed, 3 showed restricted expression patterns, 3 were widespread and 1 failed to express (Figure 3.7).

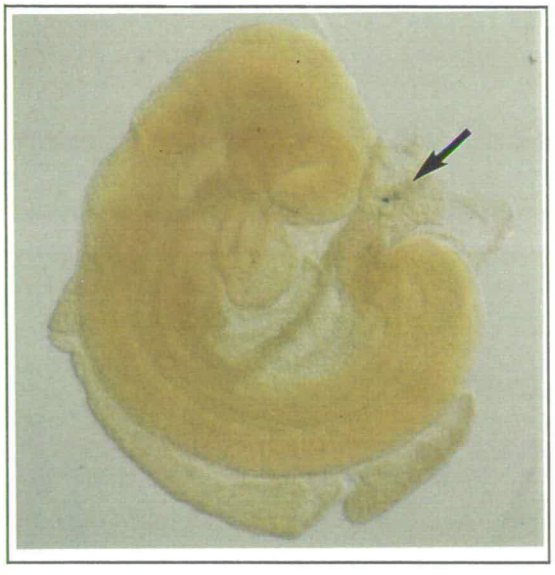
Out of three lines where chimeras were generated, ET105, was the only one successfully transmitted through the germ line. Heterozygous intercrosses were set up after one backcross to C57Bl/6 males. Offspring were genotyped by dot blotting tail DNA and hybridising to a neo probe (data not shown). Five suspected homozygotes were identified whose DNA hybridised to neo with twice the intensity of the suspected heterozygotes. To confirm the genotype of the homozygous animals a restriction fragment length polymorphism (RFLP) was identified to distinguish between the wild type and the disrupted ET105 alleles (Figure 3.8). The probe used to identify the RFLP was cloned from LM-PCR products which flank the genomic insertion site. The Southern blot genotypes were identical to the dot blot results confirming that homozygous animals are viable (Figure 3.8C).

3.8 Discussion

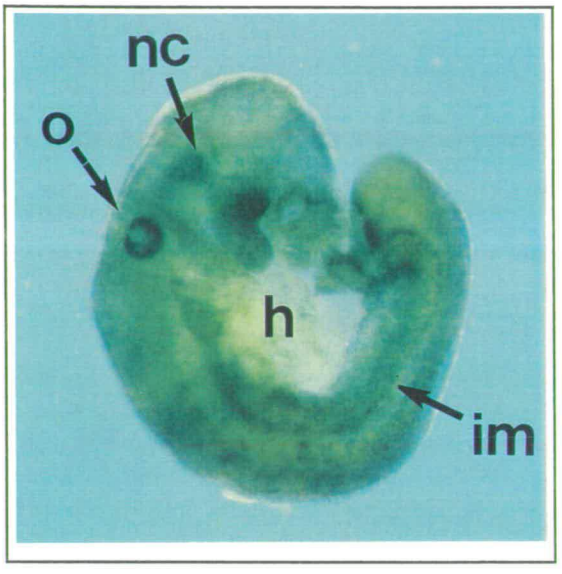
The results in this chapter demonstrate the feasibility of using an exon trap approach to screen for developmentally regulated genes. In two published reports the efficiency of detecting insertions into transcription units was 12-50 fold lower using an exon trap vector compared to a vector designed to insert into introns (Gossler *et al.*, 1989; Friedrich and Soriano, 1991). The decrease in efficiency is thought to reflect the smaller exon target size in the genome compared to the intron target. The recovery of similar numbers of colonies for both types of vector in our screen was therefore unexpected. Molecular characterisation of 8 pET1.8TM insertions revealed that between 300 - 600 bp of sequence had been deleted from the 5' end of the vector prior to its insertion into the genome. Deletions from the ends of electroporated DNA that destroy reporter activity may explain the poor frequency of gene trap events in

Figure 3.7 Reporter expression of ET lines during development

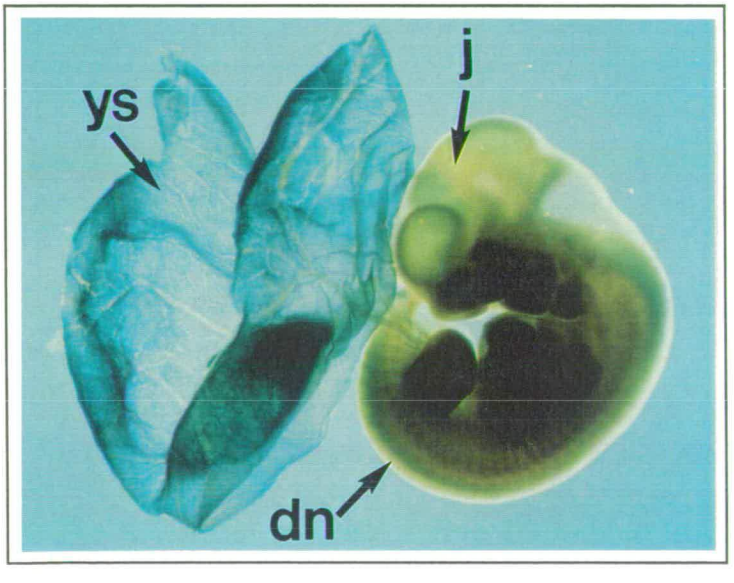
LacZ expression was monitored by staining for βgal activity in whole mount embryos. Lateral views of 8.5dpc (A and B), 9.5dpc (D) and 10.5dpc (C) embryos. (**A**) ET2 chimera. LacZ is expressed in the allantois (arrow). (**B**) ET82 chimera. Widespread expression is observed with greatly reduced levels in the heart (h). Higher levels are seen in the otic vesicle (o), intermediate mesoderm (im) and the neural crest (nc). (**C**) ET105 (heterozygote). Fairly widespread expression although absent from the dorsal neural tube (dn) and the anterior/midbrain junction (j). Highest levels of expression are seen in the yolk sac (ys). (**D**) ET143. LacZ expression restricted to the vascular system (v), otic vesicle (o), limb buds (lb) and branchial arches (b).



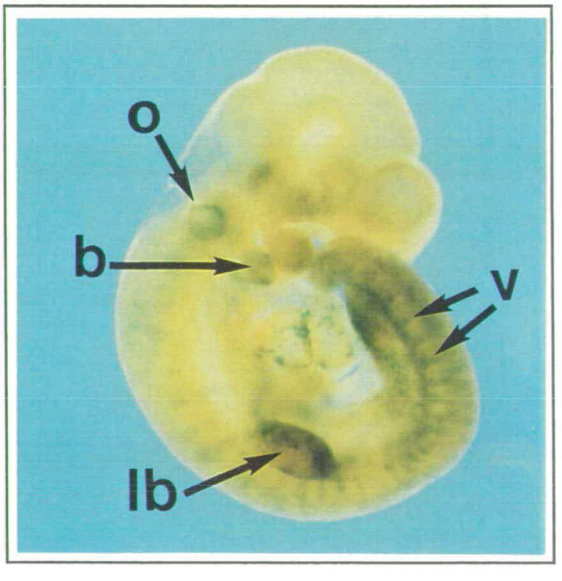
A



B



C

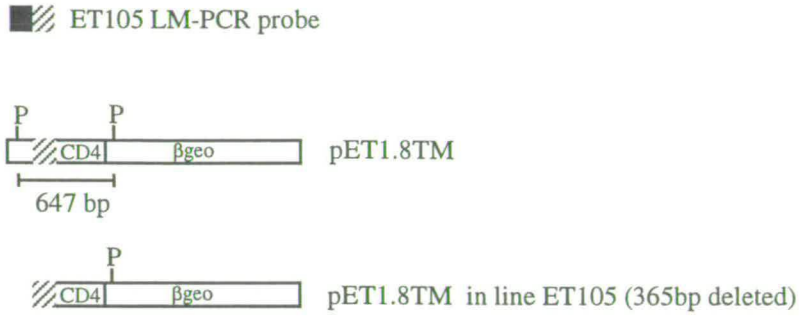


D

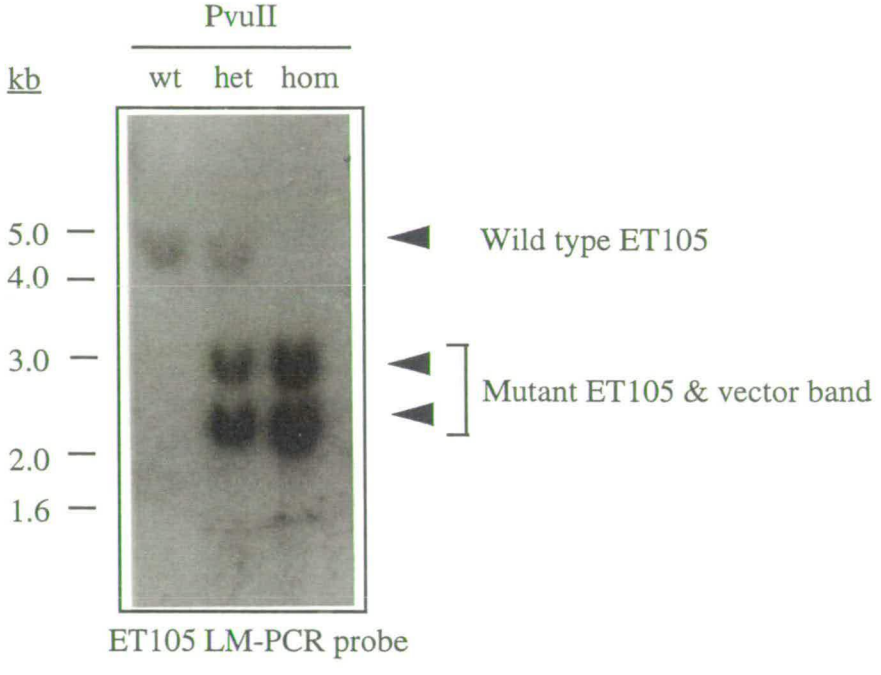
Figure 3.8 Characterisation of ET105 insertion

(A) The probe used to identify the RFLP was cloned from the LM-PCR products and contains 130bp of ET105 genomic sequences (black shading) flanking the secretory trap insertion site and 155bp of vector CD4 (hatched shading). In ET105, 365bp of vector has been deleted, including one of the PvuII sites. This makes PvuII a suitable enzyme to digest genomic DNA with to look for restriction fragment length polymorphisms (RFLPs). **(B)** A single fragment of 5kb hybridises in the wild type DNA which corresponds to the ET105 locus. An endogenous CD4 fragment does not hybridise because the vector CD4 sequence originates from rat DNA which does not cross-hybridise to mouse DNA. The wildtype ET105 band is eliminated due to the vector insertion and two bands of approximately 2.5 and 3.0kb arise. One of these bands corresponds to the PvuII fragment flanking the insertion site and the other corresponds to an internal PvuII vector fragment. The internal fragment is generated as a consequence of multiple copies of the vector having inserted at the ET105 locus. **(C)** Genotype data of ET105 weaning age offspring. (minus (-) signifies that no homozygotes are expected from that particular cross).

A.



B.



C.

	Genotype		
Cross	WT	HET	HOM
Test	6	6	-
Back	9	9	-
Inter	8	5	5

previous attempts at exon trapping. The improved efficiency of pET1.8TM is likely due to the buffering property of the CD4 TM domain upstream of the reporter.

The pET1.8TM exontrap vector was designed to detect insertions in secretory molecules yet it is clear that deletion of CD4 sequences abolishes its selective property. This was first suspected when the subcellular localisation of βgal activity was analysed in ES cells. Rather than colonies showing the secretory pattern of staining which is indicative of fusion proteins having been targeted to the secretory pathway, exontrap colonies exhibited a variety of localisations. Similar to observations with the pGT1.8βgeo gene trap vector, this variety probably reflects differential sorting of various classes of endogenous genes fused to βgeo. Sequence information from exontrap lines confirmed the deletion of transmembrane sequences from the vector accounting for non-secretory patterns.

These results highlight the importance of protecting the 5' end of a vector from deletion if exon trapping is to be successful. The ends of electroporated DNA are thought to be deleted by the action of endogenous exonucleases (Kumar & Simons, 1993). This exonuclease activity causes problems with positive/negative gene targeting constructs where the TK negative selection cassette is often damaged prior to or upon integration, leading to a background of non targeted events (Mansour, Thomas & Capecchi, 1988). A recent report shows how the addition of a nucleic acid hairpin structure to the end of a targeting vector increases the enrichment for targeted events, presumably by protecting the TK gene from exonuclease activity (Nehls *et al.*, 1996). Alternatively, random integration of these protected vectors may have been prevented by the absence of a free end. We are exploring the possibility of ligating a similar hairpin structure to the end of pET1.8TM to protect the CD4 sequences from deletion and thus preserve the selective property of the vector. An alternative strategy to protect the vector's selection is to include buffer sequences upstream of CD4. These sequences would need to contain an open reading which does not interfere with

the type II property of the transmembrane domain. It is difficult to know how much buffer to include as deletions of up to 2.3kb have been observed with gene trap vectors (Chapter 2). A drawback to this approach is that deletions will still occur within the buffer, and to some degree within the CD4, so that the extent of deletion will need to be mapped by RT-PCR before endogenous genes can be identified by 5' RACE.

Another way to overcome the problems associated with deletions is to use a retroviral based exon trap vector. However, retroviral promoter traps have been shown to insert mostly in the 5'UTRs of genes (Friedrich & Soriano, 1991; von Melchner *et al*, 1992). For the purpose of secretory trapping, it is essential that there be no bias towards inserting in the 5'UTRs as the selective procedure is dependent on insertions occurring downstream of a signal sequence.

Three out of the five insertions identified from fusion transcripts by RACE occurred in 5'UTRs (Table 3.2). The evidence for such insertions comes from the observation that fusion transcripts do not encode open reading frames in the same frame as βgal, suggesting insertions in untranslated regions. Northern data indicates that the contribution of endogenous genes to fusion transcripts is on average less than 500bp suggesting that the insertions occurred in 5'UTRs rather than 3'UTRs. Conceivably the RACE products may represent insertions in non-translated RNAs but it is unlikely this is the case for all the insertions. As the endogenous gene does not provide a translation start site when insertions occur in 5'UTRs, it is likely that reporter activity is detected as a result of internal initiation of translation from within vector sequences as discussed in Chapter 2.

It is not known why pET1.8TM insertions seem to preferentially occur in 5' UTRs particularly as 5' exons are often the smallest exons within transcription units (Hawkins, 1988). The same unknown mechanisms which cause retroviral gene trap insertions to occur in 5' UTRs (Friedrich & Soriano, 1991; Jonsson *et al.*, 1996), may

be operating with plasmid based vectors. Evidence for preferred retroviral integration into 5'UTRs comes from data showing that the number of colonies generated with a promoter trap construct containing an ATG translation start codon in frame with a lacZ gene was 5-10 times greater than the same construct lacking an ATG (Brenner *et al*, 1989). Both constructs had a splice acceptor site and they only differed in the ability of the construct containing the ATG to detect insertions into introns downstream of untranslated exons. Previous analysis of insertion sites generated by electroporation of an exontrap construct showed a similar preference for 5'UTRs in that 6 out of 7 integrations occurred within, or adjacent to, CpG islands (Macleod *et al.*, 1991). The authors proposed that DNA of CpG islands has a more open chromatin structure which allows a high level of recombination with exogenous DNA. However, the construct used in this experiment may be intrinsically biased towards detecting insertions which generate short fusion transcripts. The results may not reflect a bias towards insertions in 5' regions but rather the fact that the neo selection gene in the construct cannot tolerate large 5' fusions. A more convincing argument that electroporated DNA may preferentially integrate into 5' UTRs comes from a comparison between two gene trap constructs in which an initiating ATG in lacZ is either present or absent (Gossler, 1993). As with the retroviral experiment, addition of the ATG allows insertions into 5' UTRs to be detected. In both constructs the neo selection gene is driven by its own promoter. Four times as many blue colonies were recovered with the construct which included the ATG indicating a preference for inserting into 5'UTRs.

The 3 lines carrying insertions in rRNA genes are predicted to produce chimaeric fusion transcripts, comprised of the rRNA polymerase I transcript and the vector polymerase II transcript. The production of functional βgal activity from such non-polymerase II transcripts could be due to promiscuous activity of RNA polymerase II which produces a capped transcript capable of being translated (Palmer

et al., 1993). Alternatively there are precedents for translation from uncapped transcripts (Gunnery and Matthews, 1995).

There are a number of potential advantages of exon trapping over the more widely used gene trap approach. Firstly, insertions that occur in exons are unlikely to be spliced around, such that the insertion will completely abolish the production of endogenous transcripts. In contrast, splicing around intronic gene trap insertions which gives rise to wild type transcripts, has been observed in a number of homozygous lines (Chapter 1). Secondly, insertions into exons allow sequences flanking insertion sites to be directly cloned from genomic DNA by PCR. These junction fragments allow the simple identification of RFLPs associated with vector integrations. RFLPs can be used to unequivocally genotype offspring and, more importantly, can prove that the cloned sequence represents the gene carrying the insertion. In contrast, identifying RFLPs in gene trap lines can be a lengthy process. This is because gene trap insertions often occur in large introns, therefore cloned 5'RACE products are unlikely to contain DNA which recognises a polymorphism due to the insertion.

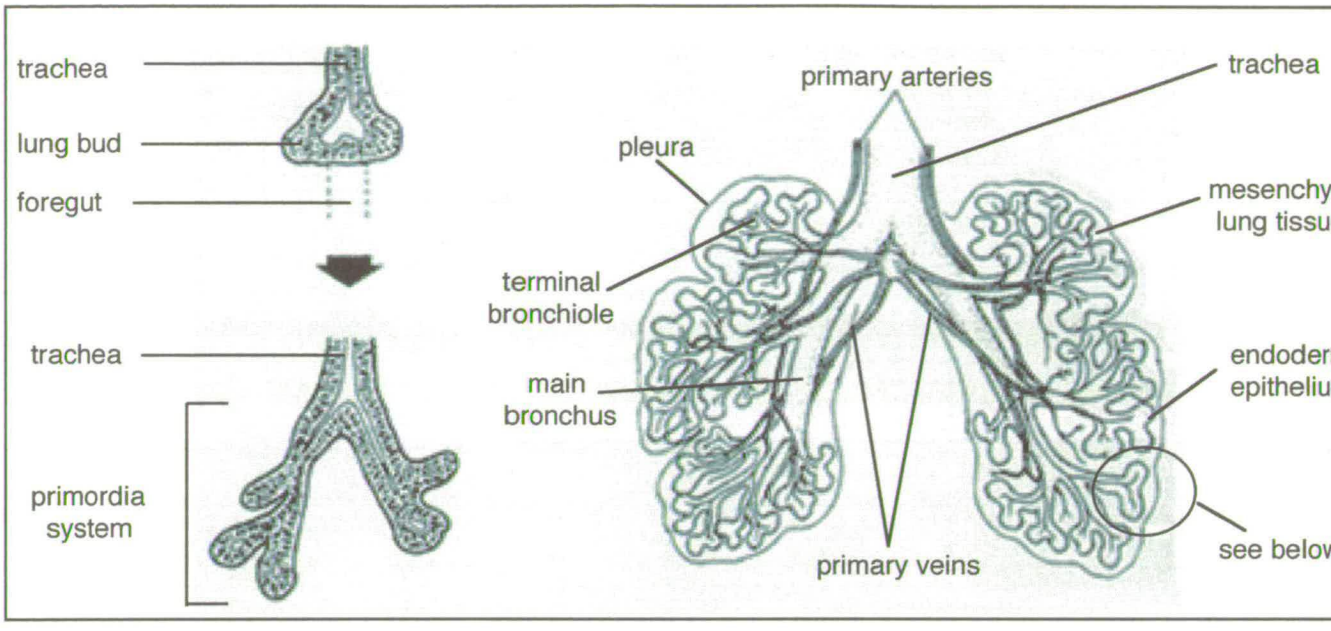
Chapter 4

DETAILED ANALYSIS OF LINE JST185

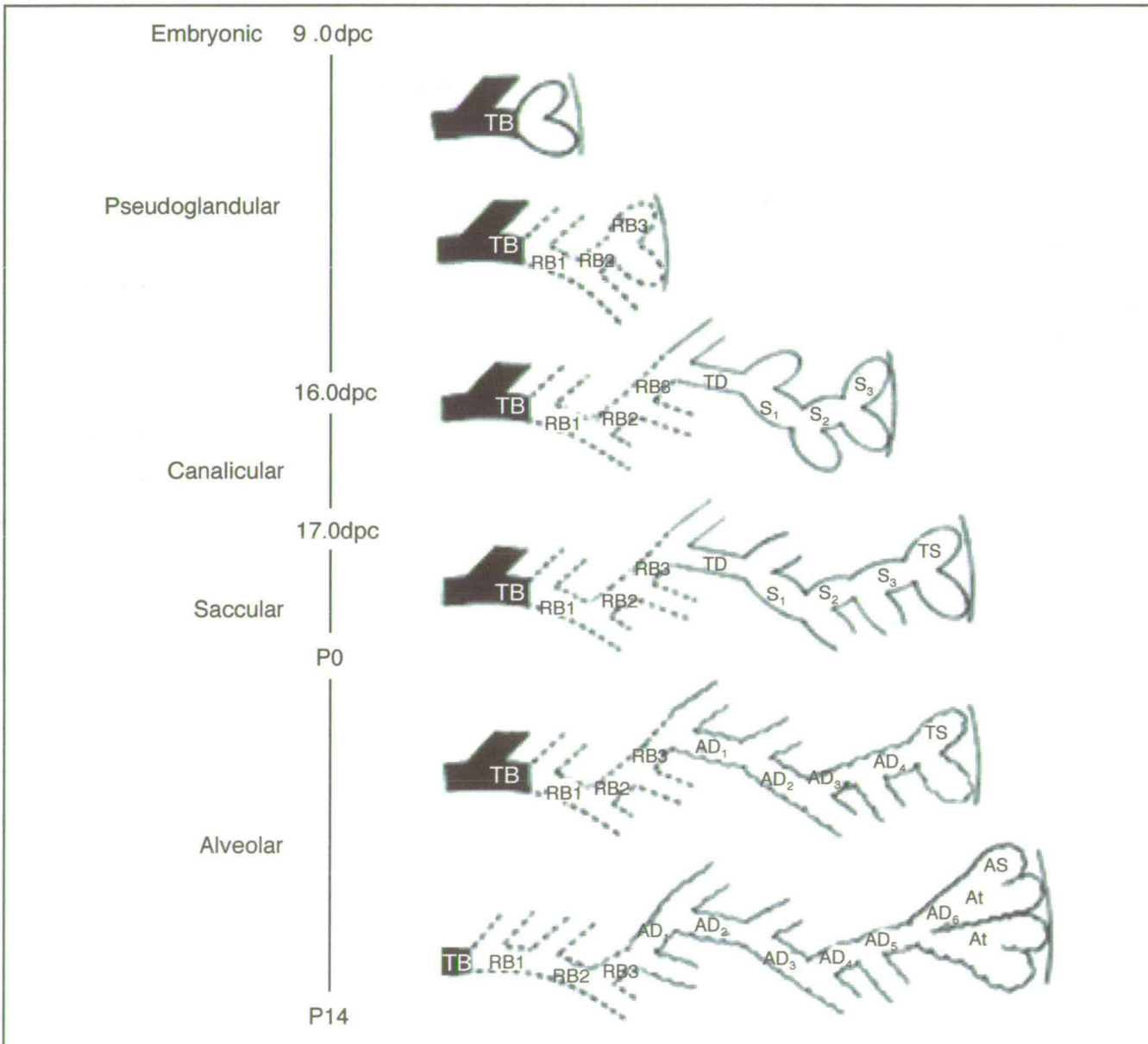
JST185 is a novel multiple membrane spanning molecule which was identified in the secretory trap screen presented in Chapter 2. In this chapter the cDNA cloning and structure of the disrupted gene is described as well as its expression pattern during embryogenesis. Animals homozygous for the insertion die shortly after birth due to respiratory failure. However, a few homozygous animals survive beyond weaning and display a striking lung pathology. The following introduction provides a brief background to lung development.

4.1 Introduction to lung development

Lung development starts around 9.5d with an outgrowth of epithelial cells from the foregut endoderm into mesenchyme, the splanchnic mesoderm, generating the lung primordium. Shortly after its appearance, the primordia develop into the prospective trachea and two lung buds (Figure 4.1A). Each bud gives rise to a


trachea
lung bud
foregut
trachea
primordia
system
primary arteries
trachea
pleura
mesenchy
lung tissu
terminal
bronchiole
main
bronchus
endoder
epitheliu
see belo
primary veins

A


Embryonic
9 .0dpc
Pseudoglandular
16.0dpc
Canalicular
17.0dpc
Saccular
P0
Alveolar
P14
TB
RB1
RB2
RB3
TD
TS
AS
At

B

C

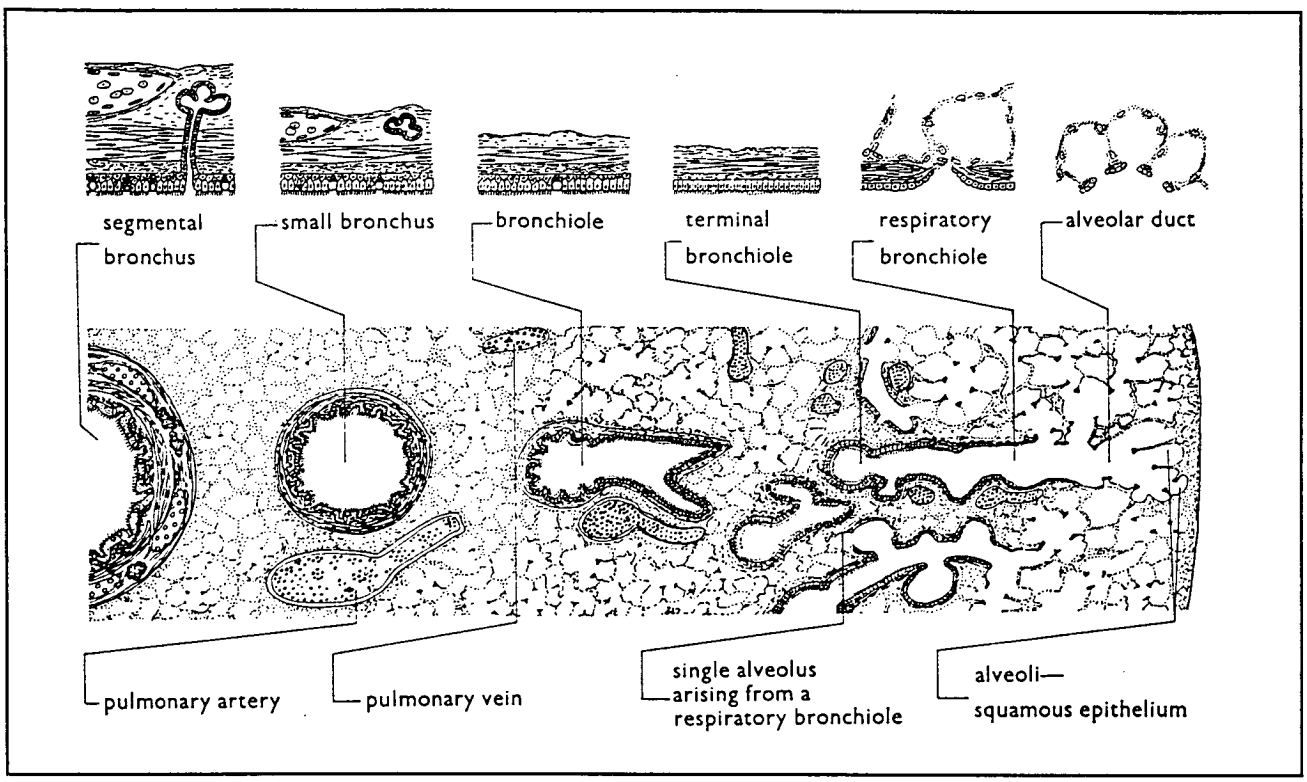


Figure 4.1 Lung development

(A) The two lung buds derive from the primitive foregut at around 9dpc. During the pseudoglandular stage of development the primordial tubes undergo a defined pattern of branching morphogenesis to form the bronchial tree which consists of a proximal conducting system and a distal respiratory system. The circled structure shows a terminal bronchiole which has terminated close to the pleura and goes on to form the pulmonary acinus (respiratory portion of lung). **(B)** Shown is a schematic representation of the growth of the pulmonary acinus. The last generation of airway at the pseudoglandular stage shows a thinned epithelium and forms the first respiratory bronchiolus (RB). These respiratory bronchioles branch further and eventually form the transitional duct (TD) which in turn gives rise to two primitive saccules (S). By birth the terminal sacs are established. No true alveoli are present although indentations representing future alveoli are found. During post-natal development the alveoli develop in the walls of the respiratory bronchioli, transitional ducts and saccules, which change these last two into alveolar ducts (AD). By about 2 weeks of development the architecture of the adult lung is acheived in which the terminal sac has formed the adult atrium (At) which has given rise to a number of alveolar sacs (AS). (Adapted from Hislop and Reid, 1974). **(C)** Schematic of adult lung histology. (Taken from Freeman and Bracegirdle, 1966)

branching tubular system which eventually forms the proximal conducting system and the distal respiratory system (reviewed in Ten Have-Opbroek, 1991; Hilfer, 1996). Lung development is described in four stages based on characteristic morphological changes: the pseudoglandular, canalicular, terminal sac and alveolar stages (Table 4.1 and Figure 4.1). During the pseudoglandular stage, the epithelial tissue proliferates and branches into the surrounding mesenchyme to form the bronchial tree (Figure 4.1A). This process is known as branching morphogenesis and relies on epithelial-mesenchymal interactions. In the absence of bronchial mesenchyme the epithelium fails to branch (Rudnick, 1933). Grafting experiments have shown that branching morphogenesis and differentiation of the proximal tracheal epithelium can be induced by interaction with distal mesenchyme, while the mesenchyme around the trachea inhibits the growth and branching of the distal epithelium (Alescio & Cassini, 1962); Wessels, 1970). During this stage molecular markers specifying proximal and distal epithelium are first expressed in distinct populations of cells (Hackett & Gitlin, 1997). By the end of the pseudoglandular period the conducting airways including the trachea, bronchi, bronchioles and terminal bronchioles have been established (Figure 4.1B).

Table 4.1 Chronology of stages in murine lung development. (Based on Minoo and King, 1994; Thurlbeck, 1996))

Stage	Age*	Physiological events
Embryonic	8.5 - 10.5d	Development of trachea and primordial bronchi
Pseudoglandular	10.5 - 16.0d	Branching morphogenesis
Canalicular	16.0 - 17.5d	Development of vascular bed and framework of respiratory acinus; flattening of epithelium
Saccular	16.5 - P0	Increased complexity of saccules
Alveolar	P0 - P14	Development of alveoli

* d = day of gestation, P = days after birth

The canalicular stage is characterised by the development of the vascular system, the airways divide further, the epithelium thins, and the proportion of

mesenchyme decreases. Distal to the terminal bronchioles, respiratory bronchioles develop containing both type I and II epithelial cells. The thin flat type I cells eventually facilitate gas exchange from airspace to capillary whereas the alveolar type II cells synthesise pulmonary surfactant. The surfactant spreads over a thin layer of liquid that normally coats the wall of the alveoli. In doing so, it lowers the surface tension at the air-liquid interface and thereby reduces the tendency of the alveoli to collapse at the end of expiration.

The terminal sac stage involves a dramatic increase in gas-exchanging surface area. Further thinning of the mesenchyme brings the pulmonary capillary bed in close apposition to the distal epithelium. Transitory ducts branch from the respiratory bronchioles into several saccules which end in terminal saccules. At this stage the intersaccular septa are thick. Deposits of elastin are found in regions where future alveolar structures will appear (Noguchi *et al.*, 1989). It is also during the saccular period that increased production of surfactant components yields a potentially functional lung should birth occur prematurely.

The alveolar stage begins on post-natal day 5 with a rapid rate of increase of surface area of both the airspaces and the capillaries. Secondary septa subdivide saccules and transitory ducts into alveolar sacs and alveolar ducts, respectively. The septa thin by a reduction in the volume of interstitial tissue. After about two weeks of restructuring, the mature lung histology is achieved (Figure 4.1C).

4.2 Sequence analysis of JST185

Cloning of the JST185 fusion transcript from heterozygous embryos revealed that the secretory trap insertion occurred in a novel multiple membrane spanning protein. RNA was prepared from 8.5dpc embryos derived from crosses between a JST185 heterozygous male and wild type C57Bl/6 females. The RACE protocol was used to amplify cDNA sequences spanning the splice junction from the fusion

transcript. The sequence of the longest RACE clone (480bp) was compared to the non-redundant GenBank database using the BLASTN program and no significant homology to any entry was seen (Appendix II). The RACE cDNA was used to screen a random primed D3 ES cell cDNA library (gift from Dr Hitoshi Niwa). One positive clone containing an insert of 1.8 kb was identified. Two probes corresponding to the 5' and 3' ends of this cDNA were used to re-screen the same ES cell library and two further cDNAs were isolated. The 3 overlapping cDNAs span 3510bp of JST185 sequence. The cDNAs encode the complete coding sequence of JST185 as both 5' and 3' UTRs were identified. Two in-frame methionines are encoded within the first 7 residues. The sequence preceding the second methionine fits the Kozak consensus for vertebrate initiation perfectly suggesting it represents the initiating methionine (Kozak, 1987). There is no polyA addition signal (AATAAA) or a polyA stretch, therefore sequences are missing from the 3' end of the cDNA.

Conceptual translation revealed a single open reading frame encoding a protein of 963 amino acids (Figure 4.2). BLASTP was used to compare the JST185 protein sequence to the GenBank protein sequence database. Two genes located on chromosome III of C. elegans, called F38B6.6 and F2D1, are both 31% identical to JST185. The C. elegans genes are 30% identical to each other. No information regarding structure or function of these two genes is available. Multiple alignments reveal that all three genes shared conserved blocks of sequence identity, particularly in the amino termini (Figure 4.3). Hydropathy profiles of all three proteins also reveal clear similarities (Figure 4.4A). Multiple membrane spanning domains appear to reside within the 400 amino-terminal most amino acids of the three proteins. TMpred predicts 9 TM domains in JST185 and 11 in the two C.elegans genes. The first hydrophobic stretch is predicted to be an uncleaved signal sequence which takes a type II orientation in the membrane. As the secretory trap insertion occurred downstream

Figure 4.2 Sequence of JST185 cDNA

The JST185 nucleotide sequence is derived from three overlapping cDNA clones and is shown on the upper line. The conceptual translation of the JST185 open reading frame is shown on the lower line. Nine putative transmembrane domains are underlined and TPR domains are in red and underlined. The site of the secretory trap insertion is marked with an arrow.

```
cggccgcggatccgccggccggagtccgagcgtggcaggctatgactgtgttttctgcggagtccttctccggtccaccatcccagaagctt   90
tcctgaaccgtctcctggcggacgcctcagtttgagggtgatcgatgcttgaagggaagatggctgatattaacttcaaagaagtgact   180
                                                    M  A  D  I  N  F  K  E  V  T                10

ttaatagtcagcgtggttgccgcctgctactggaacagcctgttctgtgggtttgtttttgatgatgtttcagcaatattggataataaa   270
L  I  V  S  V  V  A  A  C  Y  W  N  S  L  F  C  G  F  V  F  D  D  V  S  A  I  L  D  N  K     40

gacctgcatccatctacacctttaaaaacttttatttcagaatgacttctggggaactcctatgtctgaggagagaagccacaagtcatac   360
D  L  H  P  S  T  P  L  K  T  L  F  Q  N  D  F  W  G  T  P  M  S  E  E  R  S  H  K  S  Y     70

aggccttaacggtgttgacatttgcttaaattatctcctcagtgagctgaaaccatgtcttaccacctcctgaacacagtgttccac   450
R  P  L  T  V  L  T  F  R  L  N  Y  L  L  S  E  L  K  P  M  S  Y  H  L  L  N  T  V  F  H    100

gctgtcgtcagcgtgatcttccttaaggtctgcagactctttctggataagaggagcagcatgattgcggctctgctcttgcagtgcac   540
A  V  V  S  V  I  F  L  K  V  C  R  L  F  L  D  K  R  S  S  M  I  A  A  L  L  F  A  V  H    130

ccaatccatacccgaggcagtcacaggtgttgtaggaagagcggaactcttatcatctgtcttttcctagctgccttctctctgtatact   630
P  I  H  T  E  A  V  T  G  V  V  G  R  A  E  L  L  S  S  V  F  F  L  A  A  F  L  S  Y  T    160

aagtcaaaaggaccagataattccatagtgtggacccccgattgtattaacccgtgttttagtggctgttgcaacattgtgtaaggaacaa   720
K  S  K  G  P  D  N  S  I  V  W  T  P  I  V  L  T  V  F  L  V  A  V  A  T  L  C  K  E  Q    190

ggaataacggttgtcggaatttgttgtgtatatgaagtatttggtggcccaagggtacaccttgccgatgttatgcacccgttgccgggca   810
G  I  T  V  V  G  I  C  C  V  Y  E  V  F  G  G  P  R  V  H  L  A  D  V  M  H  R  C  R  A    220

gtttctgcgcgggaagggcagcattccgttatctatgctgcagacactggtgaagctcattgtcctgatgctcagtaccttgctactcgtt   900
V  S  A  R  K  G  S  I  P  L  S  M  L  Q  T  L  V  K  L  I  V  L  M  L  S  T  L  L  L  V    250

gtggtcagagtccaagttattcagtcacaacttccagtgtttaccaggtttgataatccagctgctgtaagcccaaccccgacacgacag   990
V  V  R  V  Q  V  I  Q  S  Q  L  P  V  F  T  R  F  D  N  P  A  A  V  S  P  T  P  T  R  Q    280

ctgacttttaactatctccttcctgtgaatgcctggcttctgctgaatccttcggagctctgctgtgattggaccatggggacaataccc  1080
L  T  F  N  Y  L  L  P  V  N  A  W  L  L  L  N  P  S  E  L  C  C  D  W  T  M  G  T  I  P    310

ctgatagaatcatttctagatgttcgaaatctcgccactttgctttcttttgctttctgggggctttgggaatattcagtctcagatac  1170
L  I  E  S  F  L  D  V  R  N  L  A  T  F  A  F  F  C  F  L  G  A  L  G  I  F  S  L  R  Y    340

cctggtgactcctcaaagactgtcctaatggcgctttgtttaatggcgttaccatttatttccgcatcaaacctgttctttcctgttgga  1260
P  G  D  S  S  K  T  V  L  M  A  L  C  L  M  A  L  P  F  I  S  A  S  N  L  F  F  P  V  G    370

tttgtggttgctgagcgagtactatatgttcctagcatggggttctgtattttagtagccatggatggcaaaagatttcaaacaaaagt  1350
F  V  V  A  E  R  V  L  Y  V  P  S  M  G  F  C  I  L  V  A  H  G  W  Q  K  I  S  N  K  S    400

gtgctgaaaaagctctcgtgggtttgtctgtccatggtgatactaacccatgccttgaaaacacttcatagaaattgggactgggagtca  1440
V  L  K  K  L  S  W  V  C  L  S  M  V  I  L  T  H  A  L  K  T  L  H  R  N  W  D  W  E  S    430

gnatatacattgtttatgtcagcctaaaggtgaataaaaacaatgccaaattatggaataatgtgggtcatgctctggagaatgagaag  1530
?  Y  T  L  F  M  S  A  L  K  V  N  K  N  N  A  K  L  W  N  N  V  G  H  A  L  E  N  E  K    460
```

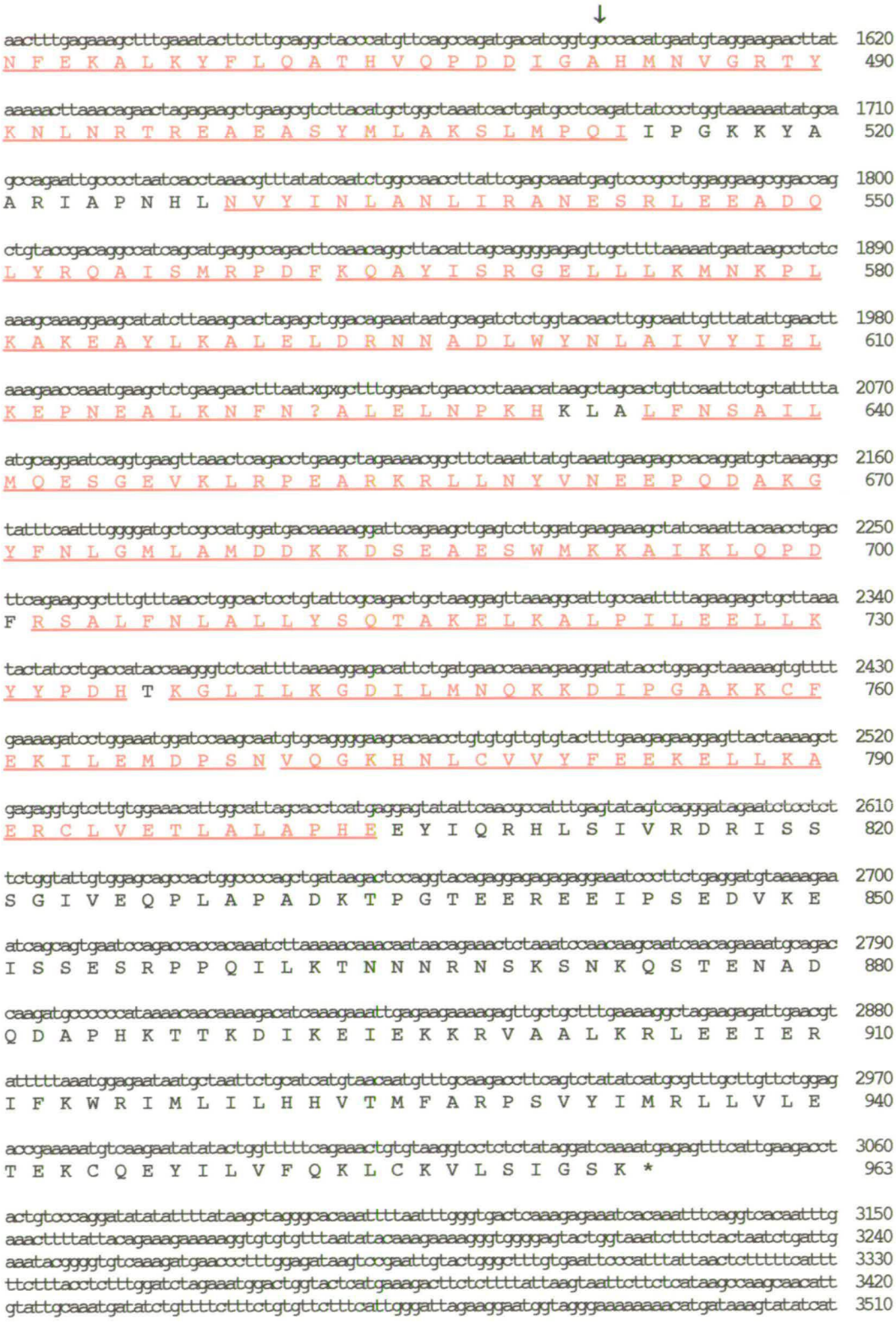
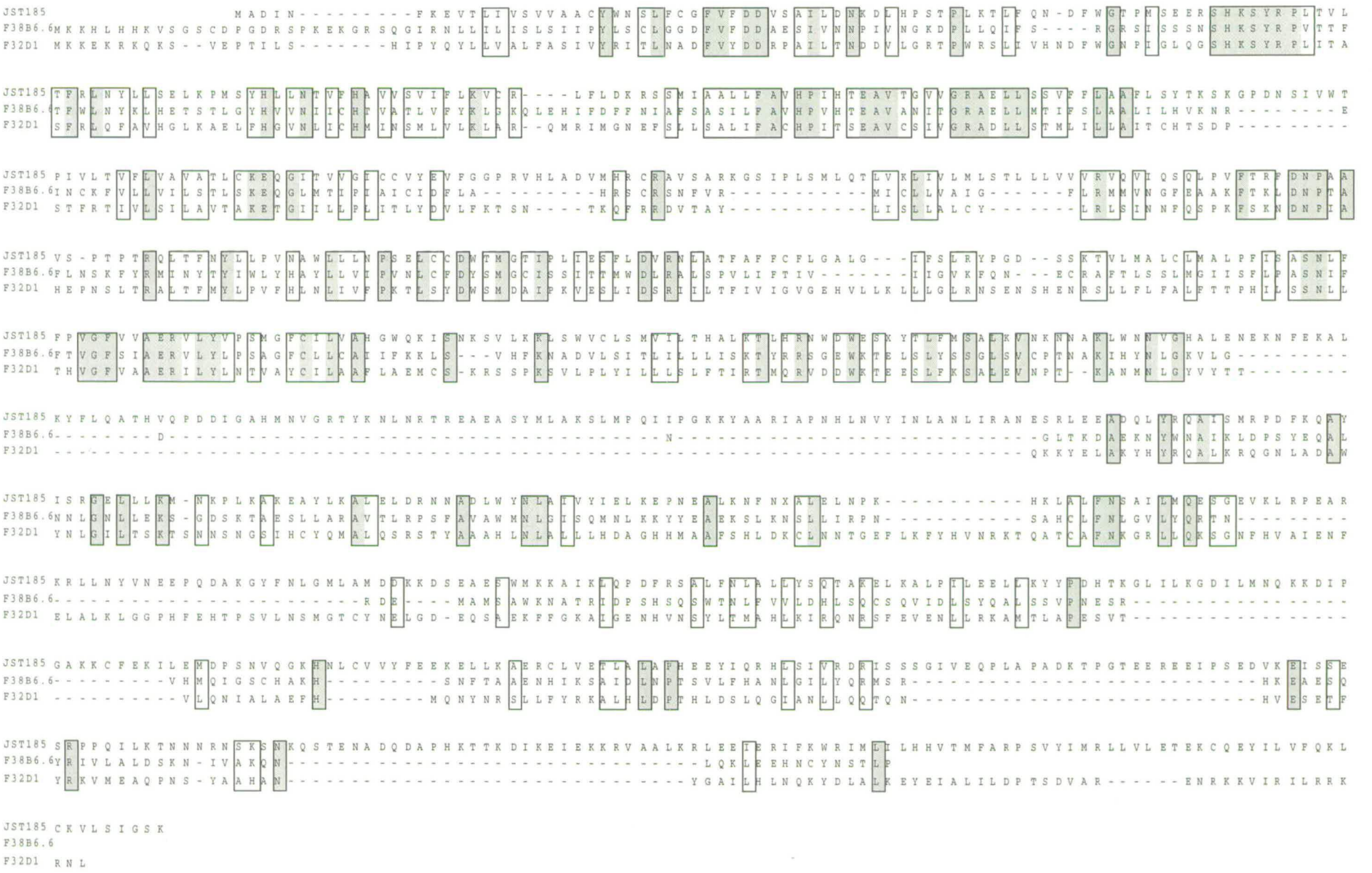


Figure 4.3 Amino acid alignment of JST185 and nematode homologues

Sequences were aligned using the ClustalW algorithm from MacVector (version 5.1 software). JST185 protein sequence aligned with two C.elegans genes, F38B6.6 (accession number U40060) and F32D1 (accession number AF016427).


JST185
F38B6.6
F32D1

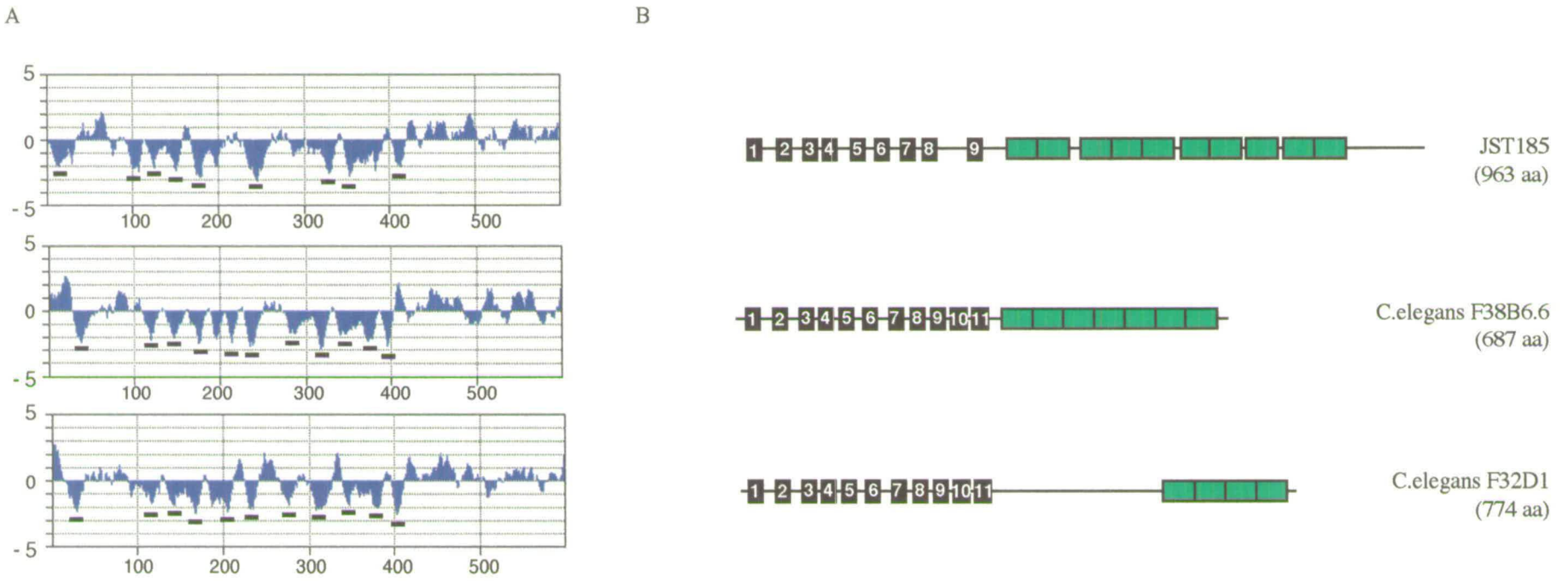


Figure 4.4 Hydropathy profiles of JST185 and nematode homologues
(A) Hydropathy values were calculated using MacVector version 5.1 software according to the Kyte-Doolittle algorithm with a window size of 12 amino acids(Kyte and Doolittle, 1982). The first solid line indicates an uncleaved signal sequence. Following lines indicate putative transmembrane domains. Only the first 600 amino acids for each protein are shown because no transmembrane domains are predicted to occur in their carboxy termini. **(B)** Cartoon structures of JST185, F38B6.6 and F32D1. Each contains between 9-11 transmembrane domains (black numbered boxes) and 4-10 TPR repeats (green boxes).

of all the TM domains, the vector CD4 TM domain will be placed in a type I orientation thus exposing βgal to the cytosol where it is active.

The carboxy-termini of the three proteins differ in length but they all show a high degree of similarity to proteins containing tetratricopeptide repeats (TPRs). These 34 amino acid motifs are found in proteins that function in diverse processes, including cell cycle control, transcriptional repression, protein transport and protein dephosphorylation (reviewed in Goebl & Yanagida, 1991). TPRs contain 8 consensus residues whose size, hydrophobicity and spacing are conserved, yet the absolute identity of each residue is flexible. Secondary structure studies predict that TPRs form a pair of amphipathic α-helical domains, punctuated by proline-induced turns, that have been proposed to mediate TPR-TPR interactions (Figure 4.5B). Domain A (W-LG-Y) is predicted to form a hydrophobic pocket into which the bulky phenylalanine (or tyrosine) side-chain of domain B (A-F-A) is predicted to fit. While there is as yet no evidence that TPR motifs interact directly with each other, they have been shown to participate in interactions between TPR-containing proteins. For example, the TPR-containing proteins CDC23 and CDC27 form part of a complex that promotes anaphase; a mutation in the TPR region of CDC27 impairs its ability to interact with CDC23 (Lamb *et al.*, 1994). There is also evidence that TPRs mediate interactions with non-TPR containing proteins: the transcriptional repression protein SSN6 (Cyc8), for example, interacts with specific DNA-binding proteins by means of its TPR region (Tzamarias & Struhl, 1995).

Careful scrutiny of the JST185, F38B6.6 and F2D1 sequences revealed that they each contain between 4 and 10 TPR domains within their C-termini (Figure 4.4B). JST185 contains 10 TPRs some of which are separated by short stretches of between 3 and 15 amino acids (Figure 4.2). F38B6.6 contains 7 tandemly arranged TPRs and F2D1 contains 4 TPR tandemly arranged domains. Comparison between the individual TPR motifs of JST185, F38B6.6 and F2D1 reveals they are more

closely related to each other than to the consensus (Figure 4.5A). The 185 consensus differs from the canonical TPR consensus sequence in that tryptophan is not well conserved at position 4, this is replaced by histidine or leucine in 10/17 cases. This tryptophan differs in a number of other TPR-containing proteins including a mammalian protein, p150, which also binds to SH2 containing proteins (Malek *et al.*, 1996) and Tg737 which is mutated in polycystic kidney disease (Moyer *et al.*, 1994).

A *Drosophila* EST (Accession nos. AA540614 & X92653) has recently been identified with 51% identity to JST185 over 420 amino acids spanning the transmembrane region. There are also two human ESTs (Accession nos. T66249 & AA540614) with 50% identity to JST185 nucleotide sequence. It is unclear how many genes these represent since they match different parts of the JST185 sequence.

4.4 Analysis of JST185 endogenous and fusion transcripts

RNase protection was used to confirm that the RACE clone, and thus the cloned cDNA, represents the interrupted gene in line JST185. Total RNA was prepared from wild type, heterozygous and homozygous 12.5dpc embryos. A 200bp fragment was released from the RACE clone and used as a template for riboprobe synthesis. The riboprobe contains the *En2* exon from the vector which acts as an internal loading control. In wild type RNA only the 120bp endogenous *En2* band was protected (Figure 4.6A). In the heterozygous and homozygous lanes endogenous *En2* was protected and also a fragment of 180bp which corresponds to the sequence from the fusion transcript which spans the splice acceptor. Unfortunately the 40bp endogenous JST185 fragment which is expected to be protected in the wild type and heterozygous RNAs, and absent from homozygous RNA, was too small to be detected.

The expression of JST185 was analysed by RNase protection during development and in adult tissues using a probe prepared from the JST185 cDNA

Figure 4.5 Alignment of JST185 tetratricopeptide repeat (TPR) domains

(A) Sequence comparison between TPR domains of JST185 and C.elegans F38B6.6 and F2D1. Consensus TPR residues are highlighted in yellow. Conserved amino acids within the repeats of the same proteins are in red. The canonical TPR consensus sequence as defined in (Lamb, 1995) is shown at the bottom. The 185 consensus represents the consensus residues between the repeats in JST185, F38B6.6 and F32D1 **(B)** Surfaces for TPR-TPR and TPR-non-TPR interactions. Domains A and B of a canonical TPR are shown as α-helices. TPR-TPR interactions are probably mediated by the hydrophobic surfaces formed by the seven consensus residues (short arrows). TPR-non-TPR interactions need not be limited to the consensus residues but could involve any face of the α-helix (long arrows).

JST185

1	A	K	L	W	N	N	V	G	H	A	L	E	N	E	K	N	F	E	K	A	L	K	Y	F	L	Q	A	T	H	V	Q	P	D	D
2	I	G	A	H	M	N	V	G	R	T	Y	K	N	L	N	R	T	R	E	A	E	A	S	Y	M	L	A	K	S	L	M	P	Q	I
3	N	V	Y	I	N	L	A	N	L	I	R	A	N	E	S	R	L	E	E	A	D	Q	L	Y	R	Q	A	I	S	M	R	P	D	F
4	K	Q	A	Y	I	S	R	G	E	L	L	L	K	M	N	K	P	L	K	A	K	E	A	Y	L	K	A	L	E	L	D	R	N	N
5	A	D	L	W	Y	N	L	A	I	V	Y	I	E	L	K	E	P	N	E	A	L	K	N	F	N	X	A	L	E	L	N	P	K	H
6	L	F	N	S	A	I	L	M	Q	E	S	G	E	V	K	L	R	P	E	A	R	K	R	L	L	N	Y	V	N	E	E	P	Q	D
7	A	K	G	Y	F	N	L	G	M	L	A	M	D	D	K	K	D	S	E	A	E	S	W	M	K	K	A	I	K	L	Q	P	D	F
8	R	S	A	L	F	N	L	A	L	L	Y	S	Q	T	A	K	E	L	K	A	L	P	I	L	E	E	L	L	K	Y	Y	P	D	H
9	K	G	L	I	L	K	G	D	I	L	M	N	Q	K	K	D	I	P	G	A	K	K	C	F	E	K	I	L	E	M	D	P	S	N
10	V	Q	G	K	H	N	L	C	V	V	Y	F	E	E	K	E	L	L	K	A	E	R	C	L	V	E	T	L	A	L	A	P	H	E

F38B6.6

1	A	K	I	H	Y	N	L	G	K	V	L	G	D	N	G	L	T	K	D	A	E	K	N	Y	W	N	A	I	K	L	D	P	S	Y
2	E	Q	A	L	N	N	L	G	N	L	L	E	K	S	G	D	S	K	T	A	E	S	L	L	A	R	A	V	T	L	R	P	S	F
3	A	V	A	W	M	N	L	G	I	S	Q	M	N	L	K	K	Y	Y	E	A	E	K	S	L	K	N	S	L	L	I	R	P	N	S
4	A	H	C	L	F	N	L	G	V	L	Y	Q	R	T	N	R	D	E	M	A	M	S	A	W	K	N	A	T	R	I	D	P	S	H
5	S	Q	S	W	T	N	L	F	V	V	L	D	H	L	S	Q	C	S	Q	V	I	D	L	S	Y	Q	A	L	S	S	V	P	N	E
6	S	R	V	H	M	Q	I	G	S	C	H	A	K	H	S	N	F	T	A	A	E	N	H	I	K	S	A	I	D	L	N	P	T	S
7	V	L	F	H	A	N	L	G	I	L	Y	Q	R	M	S	R	H	K	E	A	E	S	Q	Y	R	I	V	L	A	L	D	S	K	N

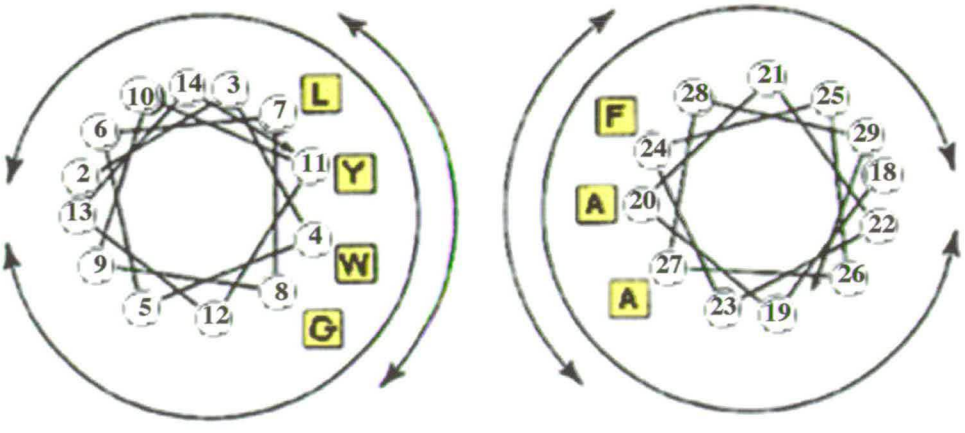
F32D1

1	V	N	S	Y	L	T	M	A	H	L	K	I	R	Q	N	R	S	F	E	V	E	N	L	L	R	K	A	M	T	L	A	P	E	S
2	V	T	V	L	Q	N	I	A	L	A	E	F	H	M	Q	N	Y	N	R	S	L	L	F	Y	R	K	A	L	H	L	D	P	T	H
3	L	D	S	L	Q	G	I	A	N	L	L	Q	Q	T	Q	N	H	V	E	S	E	T	F	Y	R	K	V	M	E	A	Q	P	N	S
4	Y	A	A	H	A	N	Y	G	A	I	L	H	L	N	Q	K	Y	D	L	A	L	K	E	Y	E	I	A	L	I	L	D	P	T	S

Consensus TPR				W			L	G			Y									A			F			A				P		
185 consensus			H/W			N	L	G			L/Y								E	A	E	K		Y/F			A	L			L	D	P

Domain A Domain B

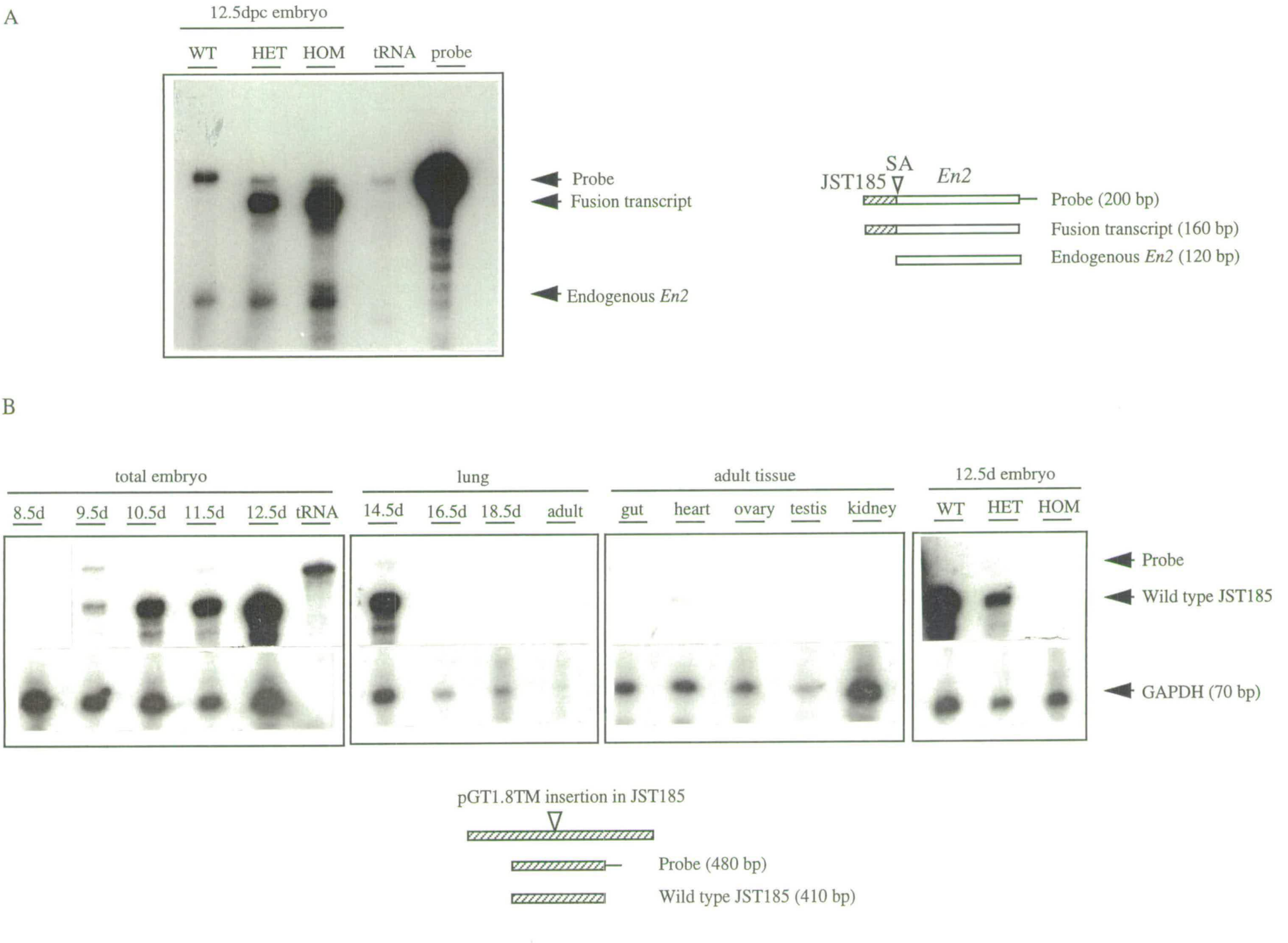
Tetratricopeptide repeat (TPR) structure



Domain A Domain B

Figure 4.6 JST185 transcript analysis by ribonuclease protection

(A) Ribonuclease protection analysis (RPA) to detect the fusion transcript in heterozygous and homozygous RNA from 8.5d embryos. The RNA probe was digested from the JST185 RACE clone. The protected RNA species corresponding to the fusion and endogenous *En2* loading control are marked. **(B)** Endogenous JST185 transcript levels were examined during development and in the adult. A fragment of the JST185 cDNA was used as a template for riboprobe synthesis. The protected RNA species is 410 bp. The last panel shows an RPA used to confirm that the synthesis of normal wild type transcripts is undetectable in homozygote RNA.


A
12.5dpc embryo
WT
HET
HOM
tRNA
probe
Probe
Fusion transcript
Endogenous En2
SA
JST185
En2
Probe (200 bp)
Fusion transcript (160 bp)
Endogenous En2 (120 bp)
B
total embryo
8.5d
9.5d
10.5d
11.5d
12.5d
tRNA
lung
14.5d
16.5d
18.5d
adult
adult tissue
gut
heart
ovary
testis
kidney
12.5d embryo
WT
HET
HOM
Probe
Wild type JST185
GAPDH (70 bp)
pGT1.8TM insertion in JST185
Probe (480 bp)
Wild type JST185 (410 bp)

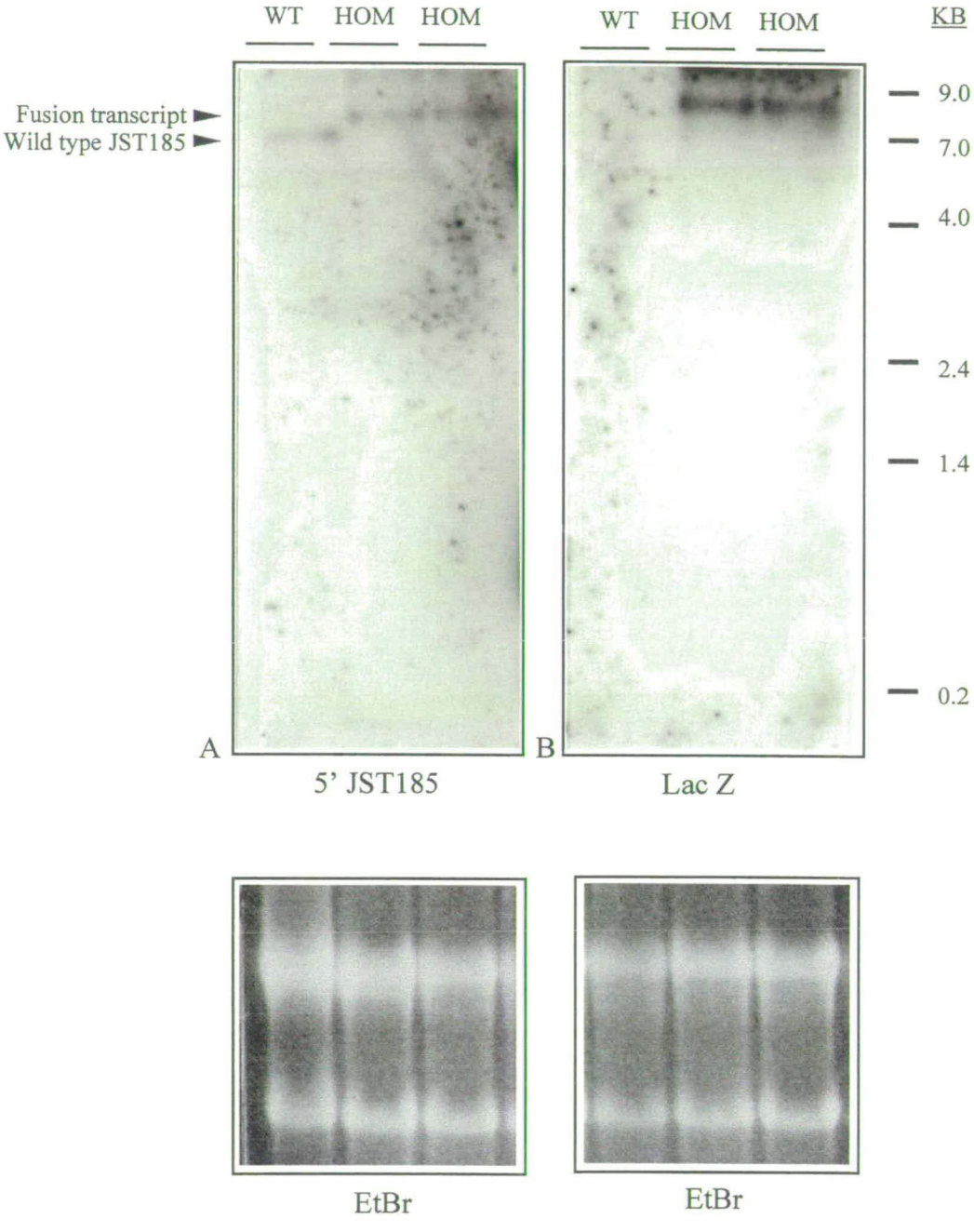
which flanks the secretory trap insertion site (Figure 4.6B). At 8.5dpc expression is undetectable by RNAse protection. A day later expression is detected and its level increases up to 12.5 dpc. Transcripts were analysed in total RNA prepared from isolated lung buds at various stages of development. JST185 is expressed at high levels during the pseudoglandular stage at 14.5dpc but by the canalicular stage at 16.5dpc expression is down-regulated. By birth and in the adult no expression is detected. Expression was not detected in any of the other adult organs tested except very low levels in the heart, gut and testis. Thus the expression pattern of the endogenous gene is consistent with that of the reporter gene.

RNase protection was used to determine whether splicing around the insertion could lead to the production of wild type transcript. Total RNA was prepared from 12.5dpc embryos, the stage at which the highest level of JST185 transcripts are expressed. This should ensure that even very low levels of wild-type transcripts will be detected. A GAPDH loading control probe was also added to the RNA samples. A single band is protected in the wild type and heterozyote (Figure 4.6B). In the homozygote no band is detected therefore no splicing around the insertion occurs to produce wild type transcript.

The sizes of the endogenous JST185 transcript and the fusion transcript were investigated by Northern blot analysis with a JST185 probe which hybridises to sequence 5' to the insertion site. Total RNA from a wild type 12.5d embryo and two homozygous embryos were analysed (Figure 4.7A). A single wild type band of approximately 7.5kb was detected. In the homozygous RNA the wild type band was absent confirming the RNase data that no splicing occurs around the insertion. A fusion transcript size of 8.0kb was detected in both homozygous samples. In a duplicate Northern blot a LacZ probe confirms the size of the fusion transcript (Figure 4.7B). The vector contribution to the fusion transcript size is predicted to be 5.3kb (Chapter 2), 1.6kb of JST185 cDNA sequence has already been cloned 5' of the

Figure 4.7 LacZ fusion and endogenous transcript in JST185

Northen blot analysis of 20μg of total RNA prepared from a wild type 12.5dpc embryo and two homozygote embryos hybridised to 5' JST185 probe (**A**) and LacZ (**B**). A fusion transcript of 8.0kb and an endogenous JST185 transcript of 7.5 kb hybridises.


WT
HOM
HOM
WT
HOM
HOM
KB
Fusion transcript
Wild type JST185
9.0
7.0
4.0
2.4
1.4
0.2
A
B
5' JST185
Lac Z
EtBr
EtBr

secretory trap insertion site, indicating that an additional 1.1kb of 5'UTR sequences have yet to be identified. In total, 3.5kb of JST185 cDNA sequence has been cloned, considering the pA tail is 150-200bp, approximately 2.7-2.75kb of 3'UTR sequences have yet to be identified.

4.5 Expression during mouse development

The developmental expression pattern of JST185 was analysed between gastrulation and birth by staining embryos from heterozygous backcrosses with X-gal. The appearance of the primitive streak at around 6.5dpc marks the beginning of gastrulation, when the bilaminar egg cylinder is transformed into a trilaminar, three-chambered conceptus. JST185 is expressed in the cells which have delaminated from the streak and formed the mesodermal layer between the epiblast and visceral endoderm (Figure 4.8A). Lower levels of expression are seen in the epiblast. At around 8.5dpc JST185 is expressed at high levels in two strips of lateral mesoderm which extend around the lateral edges of the posterior half of the embryo (Figure 4.8B). The lateral mesoderm provides much of the mesenchyme for the developing viscera. Around 9.0dpc, JST185 is expressed at high levels throughout the mesenchymal component of regions undergoing epithelial-mesenchymal interactions (Figure 4.8C). Dissecting through the embryo at the level of the hind limb reveals expression restricted to the splanchnic mesenchyme and excluded from the closely associated gut which is derived from the endoderm. Low levels of expression are seen in the ventral midline of the neural plate (Figure 4.8D).

At 12.5dpc expression is seen at high levels in the developing lung, the gastro-intestinal tract, the brain and in some cranio-facial regions (Figure 4.9A). Higher magnification of the lung and gut reveals JST185 is expressed at very high levels in mesenchyme surrounding the endoderm (Figure 4.9B &C). By 16.5dpc expression levels are down regulated (Figure 4.10D). Expression remains in a sub-population of

Figure 4.8 JST185 expression at early gestation

Heterozygous embryos were stained with X-gal during early gestation. **(A)** Lateral view of 8.0dpc embryo. **(B)** Dorsal view of early 8.5dpc embryo showing βgal activity in the lateral mesoderm (lm). Lower levels are also seen in the head folds (hf). **(C)** Lateral view of late 8.5dpc embryo showing βgal activity in the mesoderm. **(D)** Crude section cut through the posterior third of the embryo shown in (C) confirms expression restricted to the splanchnic mesenchyme (sm) and the somitic primordia (sp). No expression is seen in the gut (g) but low levels are seen along the ventral midline of the neural plate which in the section shown is folded into the neural groove (ng).

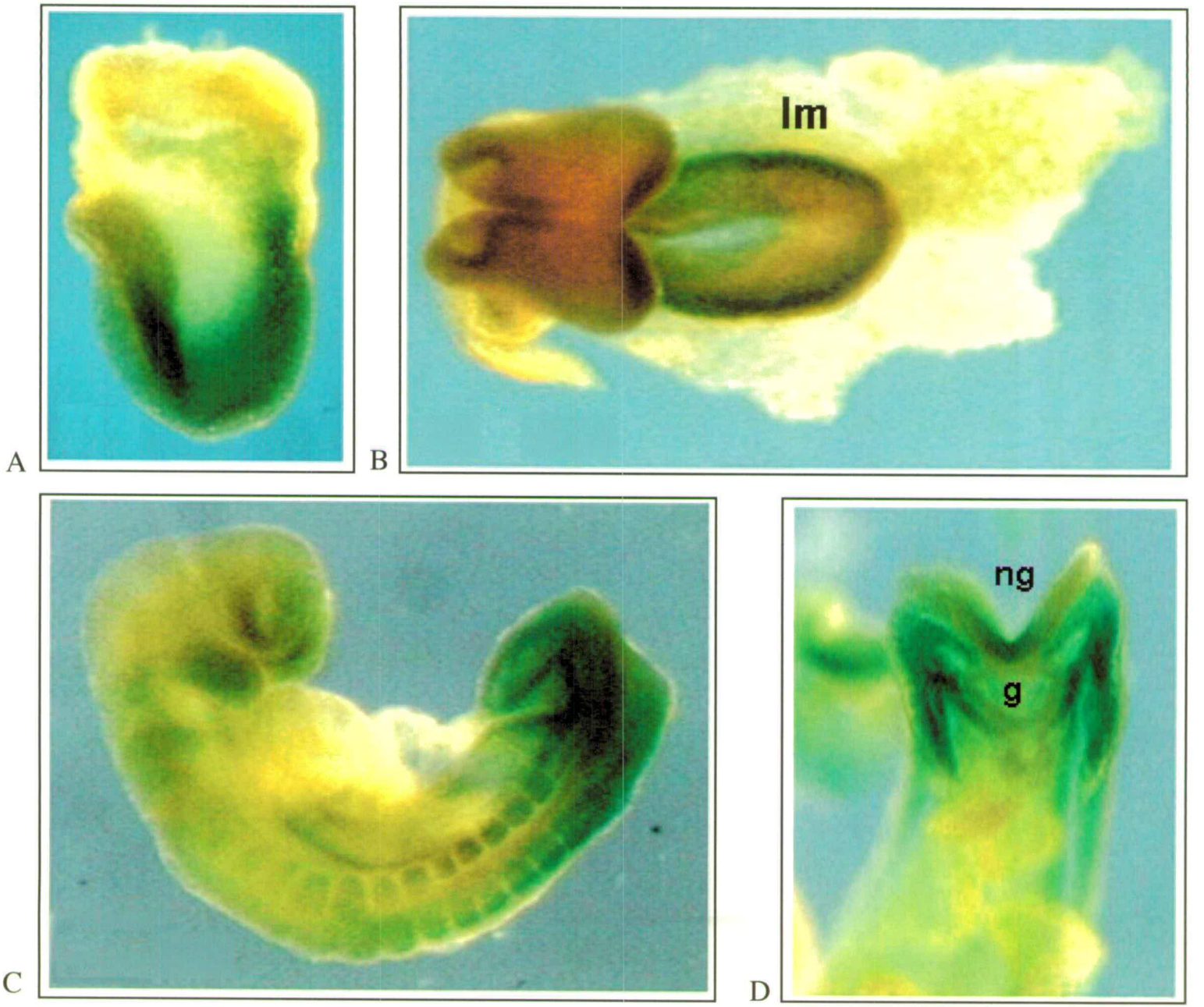
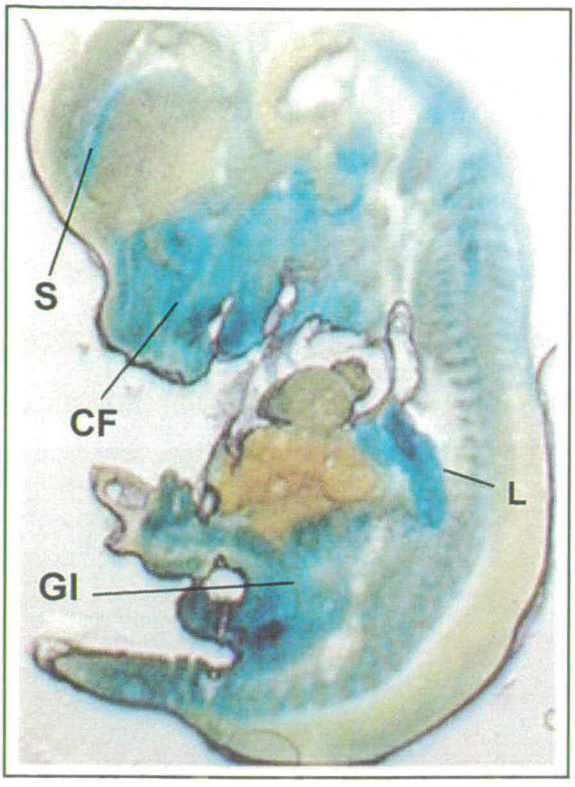
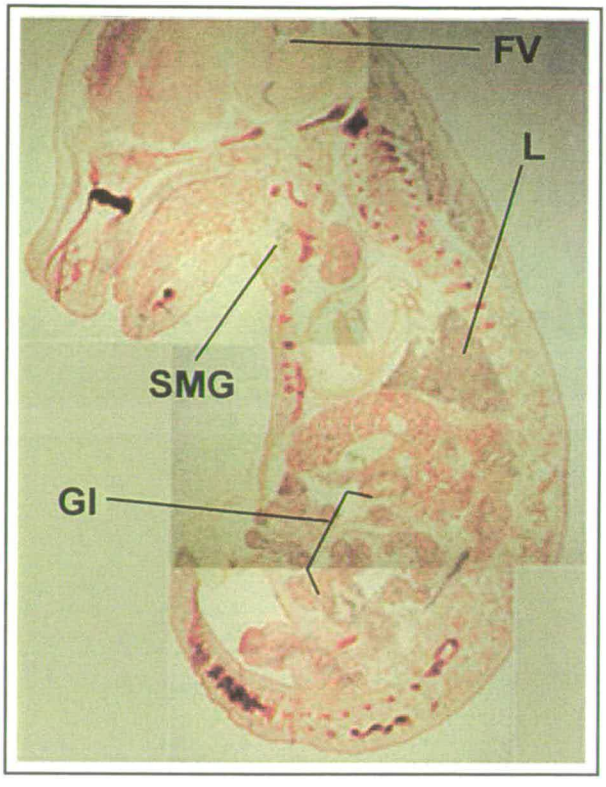
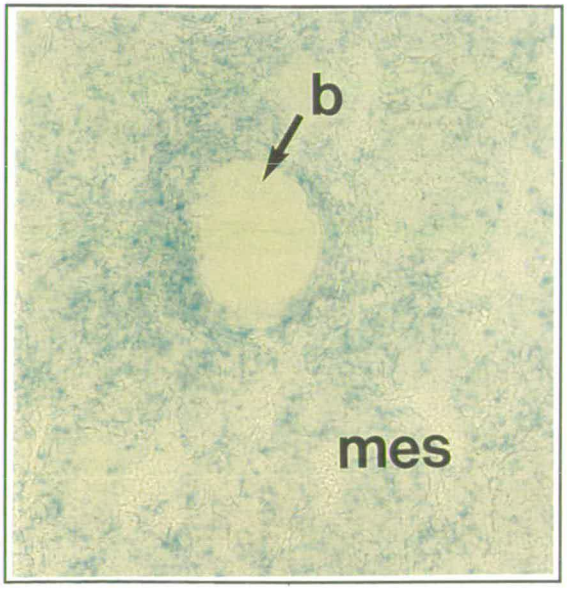

A
B
lm
C
D
ng
g

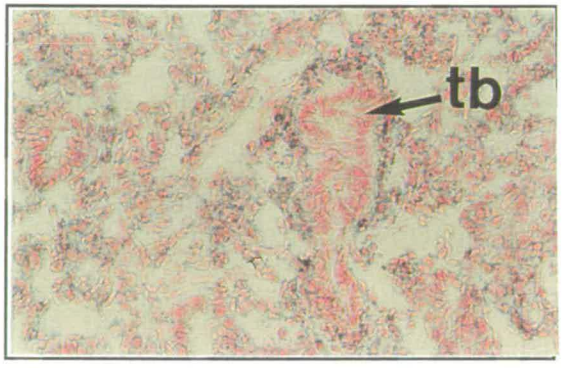
Figure 4.9 JST185 expression during mid to late gestation

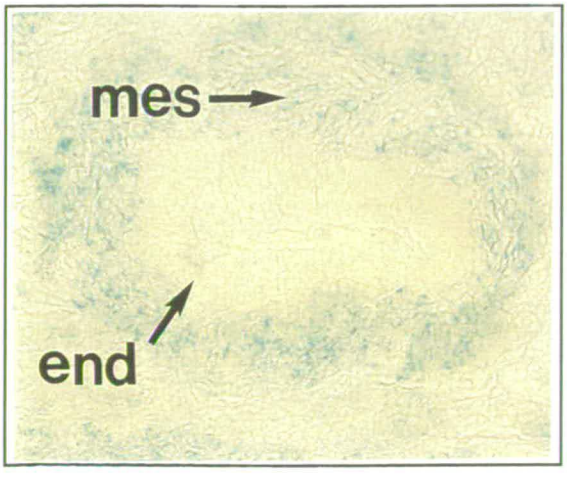
(**A**) Sagital section through 12.5dpc embryo. High levels of βgal activity are seen in the lung (L), throughout the gastrointestinal tract (GI), cranio facial regions (CF) and in the brain at the edge of the striatum (s). (**B**) High magnification view of a section through the lung showing high levels of expression in the mesenchyme, particularly surrounding the branching epithelium. Expression is excluded from the airway epithelium in proximal terminal bronchioles (b) but is seen in the epithelium of more distal respiratory bronchioles. (**C**) Part of the gastro-intestinal tract showing high levels of expression in the mesenchyme surrounding the endodermal derived epithelium. (**D**) Sagital section through 16.5dpc embryo. Levels of expression are down-regulated. Expressing regions include the lung, gastro-intestinal tract, submandibular gland (SMG) and the forth ventricle of the brain (FV). (**E**) High magnification view of a section through lung. Highest levels of expression are seen surounding the bronchioles. Lower levels are seen in subsets of cells which could be either mesenchymal or epithelial. (**F**) High magnification view through a region of the gastrointestinal tract showing high levels of expression in a ring of cells surrounding the epithelium.


S
CF
L
GI


FV
L
SMG
GI
D


b
mes


tb
E


mes
end


F

mesenchymal cells in the lung (Figure 4.9E). Expression is also seen in a ring of cells surrounding the developing gastrointestinal tract (Figure 4.9F). The morphology and location of these cells suggests they may represent smooth muscle cells. Expression is also seen in the submandibular glands and in a small region of the brain.

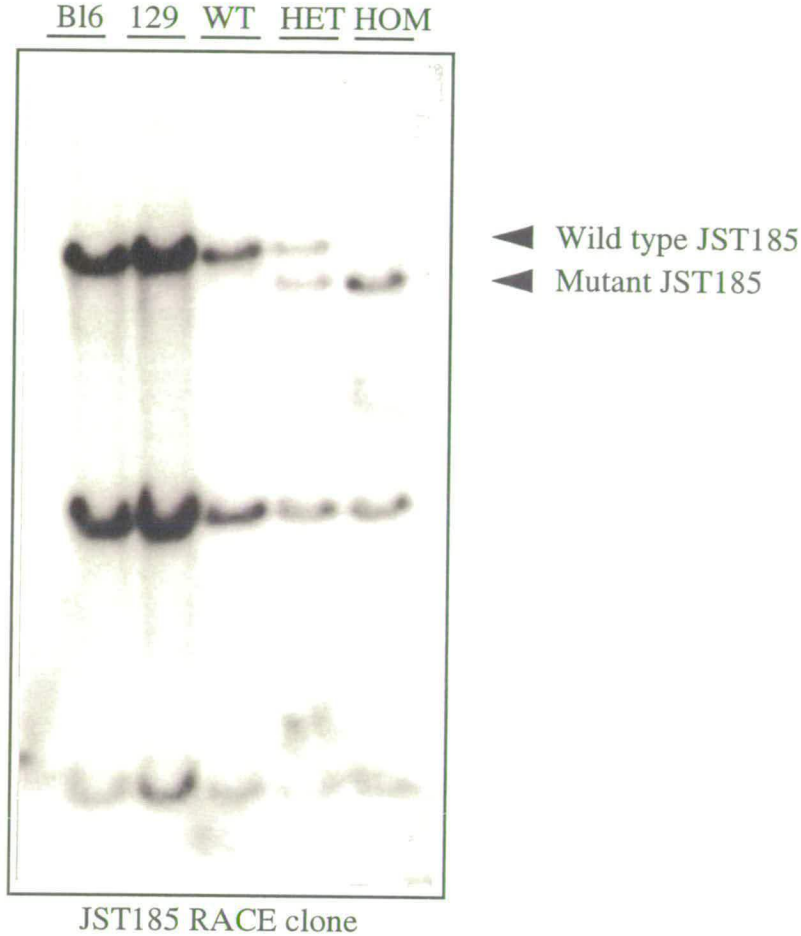
JST185 is expressed in a restricted pattern throughout lung development. At 9.5d it is expressed in the splanchnic mesoderm. It is into this region that the pharyngeal pouch from the foregut endoderm first protrudes. Sections through pseudoglandular stage embryonic lungs reveal JST185 expression at high levels in the mesenchyme surrounding the airway epithelium which is undergoing branching morphogenesis. Lower levels of expression are seen in other regions of mesenchyme. During the canalicular stage of development expression is downregulated. Expression remains highest in mesenchyme surrounding the airway epithelia. In more distal branches of the bronchial tree JST185 expression is seen in the epithelial layer as well as the mesenchyme (data not shown). During the saccular stage JST185 is expressed in isolated cells throughout the lung parenchyma and at birth there is no detectable expression in either heterozygotes or homozygotes.

4.6 Homozygous lethality

Genomic DNA prepared from JST185 heterozygotes and wild type offspring was digested with a panel of enzymes and analysed by Southern hybridisation using a probe derived from the JST185 RACE clone. An RFLP was detected in DNA digested with Pvu II which resulted in a shift in size from a 14 kb wild type band to a 12 kb mutant band (Figure 4.10). Pure C57Bl/6 and 129/OLa DNA show only the 14kb band, therefore the RFLP is a consequence of the vector insertion and not due to a strain difference.

Figure 4.10 RFLP used to genotype intercross offspring

Genomic DNA digested with PvuII analysed by Southern hybridisation with a probe made from the JST185 RACE clone. A restriction fragment length polymorphism (RFLP) is seen as a consequence of the secretory trap insertion in JST185. The polymorphism is not due to a genetic strain difference as the same sized wild type band is seen on a C57Bl/6 and 129/ola background. The probe contains the *En2* exon so background *En2* bands hybridise as well as JST185.



e 4.2 Summary of JST185 heterozygous intercross breeding

	Genetic Strain[a]	Wild Type	Heterozygote	Homozygote	Expected Homozygote[i]
k adult	mixed 129/C57[b]	63 (35%)	115 (63%)	4 (2%)[f]	59
k adult	C57[c]	32 (39%)	50 (61%)	0	27
k adult	129[d]	28 (34%)	52 (63%)	3 (3%)[g]	27
k adult	MF1[e]	28 (37%)	43 (57%)	4 (5%)[h]	24
18.5d embryo	mixed 129/C57	31 (26%)	56 (48%)	30 (26%)	29

quared analysis on the number of 3 week surviving pups indicates the difference between the expected and actual values atistically different (P<0.01)

C57 = C57Bl/6, 129 = 129OLa and MF1 = outbred
F1 [129XC57]; offspring of first backcross to C57
F1 X C57]F4; F1 offspring crossed to C57 four times
F1 X 129]F3/4; F1 offspring crossed to 129 three/four times
F1 X MF1]F4; F1 offspring crossed to MF1 four times
Runted, laboured breathing, died at 4 and 9 weeks (2 culled at 4/5 weeks)
Runted , laboured breathing, died at 6 weeks and 6 months (1 culled at 6 weeks)
Runted , laboured breathing, died at 6 weeks, 16 weeks and 5 months (1 culled 9 weeks)
Calculation based on normal Mendelian segregation. (no. of wild type + no. of heterozygotes) divided by 3

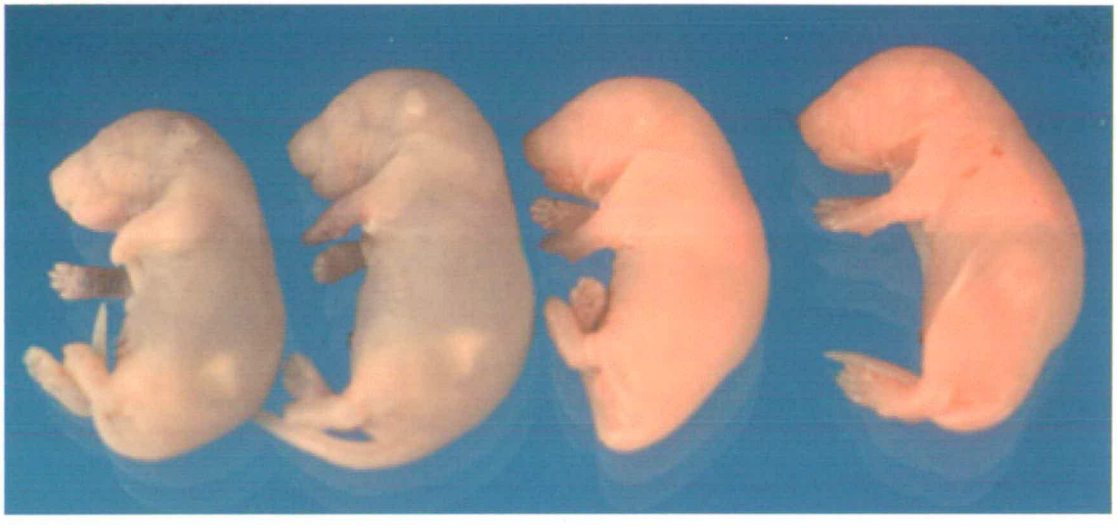
Almost 400 intercross offspring on a variety of genetic backgrounds were genotyped at weaning (Table 4.2). Only 11 homozygous animals survived until this stage indicating that 88% of homozygous pups die before weaning. All the homozygotes which survived were slightly runted and exhibited breathing difficulties which worsened with age. All homozygotes died between 3 weeks and 6 months of age, as compared to heterozygous littermates which typically survive over two years. Preliminary results suggest that genetic background may be influencing the survival rate of homozygotes. In particular the C57Bl/6 genetic background appears to increase the penetrance of the mutation, while 129/OLa and MF1 backgrounds decrease it (Table 4.2).

To investigate at what stage the majority of mutant mice were dying a number of intercross matings were set up, mothers were sacrificed at various stages of pregnancy and embryos were examined for abnormalities or lethalities. The expected Mendelian ratio of genotypes was observed up to 18.5dpc (Table 4.2), suggesting that homozygotes were dying at birth and being cannibalised. Five intercross litters were delivered from their mothers at 18.5dpc by Caesarean section. All pups had intercostal retractions shortly after birth. Most pups established a regular breathing pattern within 15 minutes and turned a healthy pinkish red colour (Figure 4.11). Three newborns remained pale, immobile and sporadically gasping for air for up to 2.5 hrs before they died. All three pups were genotyped as being homozygotes and all the surviving pups were heterozygous or wild type. The lungs from homozygous pups were smaller than controls although there was no difference in weight (Figure 4.11). Homozygous lungs sank when placed in a dish of media, whereas control lungs floated. These observations suggest that JST185 post-natal lethality is due, at least in part, to a failure of newborn pups to inflate their lungs.

The potential airspaces of the lung are filled with liquid during development. At birth, the lung must be rapidly cleared to allow aeration of the lung. To examine

Figure 4.11 Homozygous caesarian section pups fail to inflate their lungs

18.5dpc pups were delivered by Caesarian section. Homozygous pups failed to establish a regular breathing pattern, remained pale compared to littermates and died by 2.5hrs post-delivery. The lungs of homozygotes were slightly smaller and more dense than control litternates. (Hom, homozygote; Wt, wild-type).



Hom 1 Hom 2 Wt 1 Wt 2

Hom 2 Wt 2

whether the JST185 mutant lungs were water saturated at birth, the wet to dry weight ratios was calculated for two intercross litters (data not shown). There was no difference in weight ratios suggesting that water is effectively cleared from mutant lungs.

4.7 Analysis of homozygote lung development

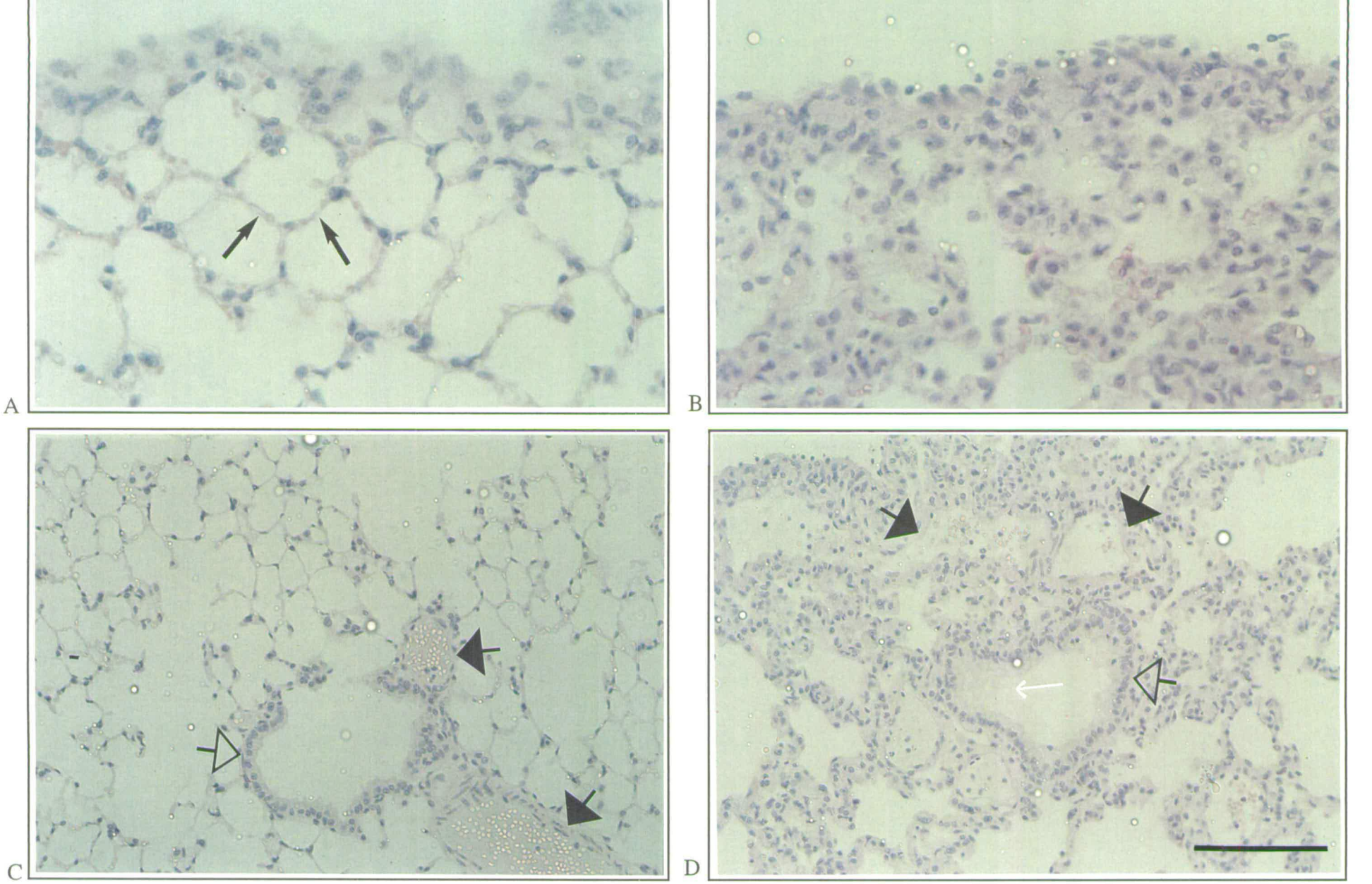
Two homozygotes were culled at 5 and 9 weeks of age, when their breathing difficulties had become very severe. Internal organs were dissected and compared to wild type and heterozygote littermates. All organs appeared normal except for the lungs. The homozygote lungs were reduced in size and, rather than being a uniform pink colour like the control lungs, had patches of brown/yellow tissue. These patches fail to inflate when fluid was injected down the trachea.

Lung sections were compared between JST185 homozygotes and heterozygous littermates. The epithelial walls of the air spaces in the mutant lungs were much thicker and highly disorganised compared to the heterozygous lungs (Figure 4.12). The architecture of the mutant lungs resemble that of the terminal sac stage of development. The transition from the terminal sac to the alveolar stage occurs post-natally and involves the production of secondary septa from the terminal sacs followed by thinning of the septal walls to form the alveoli, which greatly increase the surface area of the lung. This transition has occurred normally in the heterozygotes where the thin walled alveoli are clearly seen (Figure 4.12 A & C).

Alveogenesis depends on the presence of mesenchymally derived alveolar myofibroblasts which provide the contractile components which keep alveoli expanded (Bostrom *et al.*, 1996). As JST185 is expressed in isolated cells of the saccular stage lung newborn lungs were examined to see whether the myofibroblasts are missing or defective in JST185 mutants. Immunohistochemistry was used to examine the distribution of α-smooth muscle actin (α-sma) which marks myofibroblasts, and

Figure 4.12 Abnormal histology of JST185 homozygote adult lungs

Histological lung sections are shown from a 5 week JST185 heterozygote (A) and (C) and homozygote (B) and (D). Lungs were fixed in Bouin's reagent and counterstained with haematoxylin and eosin. At high magnification the thin walled septa of the alveoli (arrows) are seen in the heterozygote (**A**). There is a complete lack of alveolar structure in the mutant (**B**). In the heterozygote at lower magnification (**C**) a respiratory bronchiole is seen lined with simple cuboidal epithelia with an artery containing blood next to it, in the homozygote (**D**) the walls of both bronchioles and arteries are disorganised . The lumen in many of the respiratory bronchioles in the homozygote are reduced in volume by the presence of either proteinaceous exudate or mucus (white arrow). Bar (A,B) 50μm; (C,D) 100μm.


A
B
C
D

Millers solution was used to stain the elastin fibres produced by the myofibroblasts. In sections through lungs from 18.5dpc pups there is no difference between the levels of α-sma in mutants and controls. As expected the α-sma is located in the smooth muscle layer surrounding the bronchioles and in the arterial walls (Figure 4.13 A & B). Lower levels are seen in the epithelia surrounding the terminal sacs. At lower magnification the histology of the mutant lung is subtly different from the controls. The mesenchyme separating the terminal sacs is thickened in the mutants (Figure 4.13 C & D). There are also fewer distal terminal sacs in the homozygotes suggesting that branching may be retarded (Figure 4.13 E & F). The elastin levels were also similar in both mutant and control newborn lungs (data not shown).

Staining for α-sma and elastin was repeated on sections through homozygous and control lungs taken from 5 week old animals. Unlike the 18.5dpc lungs, there is a slight background of α-sma staining in the adult sections. However low level background staining is clearly distinct from bona fide antibody signal. As expected, in the control lungs α-sma is seen in cells lining the arterial walls and in isolated cells at the tips of alveolar septa which are thought to correspond to myofibroblasts (Figure 4.14 A & C). However, in the homozygous lungs the distribution of α-sma is drastically altered; it is found ectopically in cells throughout the distal airways (Figure 4.14 B & D). In the inflated mutant lung airways are distended but the highly muscularised alveoli are not expanded (Figure 4.14 D). Elastin levels are markedly increased in homozygous lungs compared to controls (Figure 4.15 A & B). Rather than being localised to the matrix and tips of the alveolar septa as in the controls, elastin in the homozygotes completely surrounds the sac-like airway epithelium (Figure 4.15 C & D).

In addition to aberrant marker expression homozygous lungs contain regions containing excessive numbers of cells which occlude entire airways in places (Figure 4.15 B & D). The identity of these cells is unknown although some regions show

Figure 4.13 Smooth muscle actin expression in 18.5dpc lungs

Lungs dissected from pups delivered by Caesarean section were analysed by immunohistochemistry for the presence of α-sma antibody and counterstained with haematoxylin and eosin. Control lungs (A, C and E) were compared to homozygote lungs (B, D and F). Staining is seen in the arterial (a) walls and surrounding the respiratory bronchioles (b) in both control and mutant lungs (**A**) and (**B**). Lower magnification shows that the mesenchyme (arrows) between the airways is thicker in the mutants (**D**) than in the control (**C**). There are many more distal terminal sacs (arrows) in the control (**E**) than the mutant (**F**). Bar (A,B,C,D) 50μm; (E,F) 100μm.

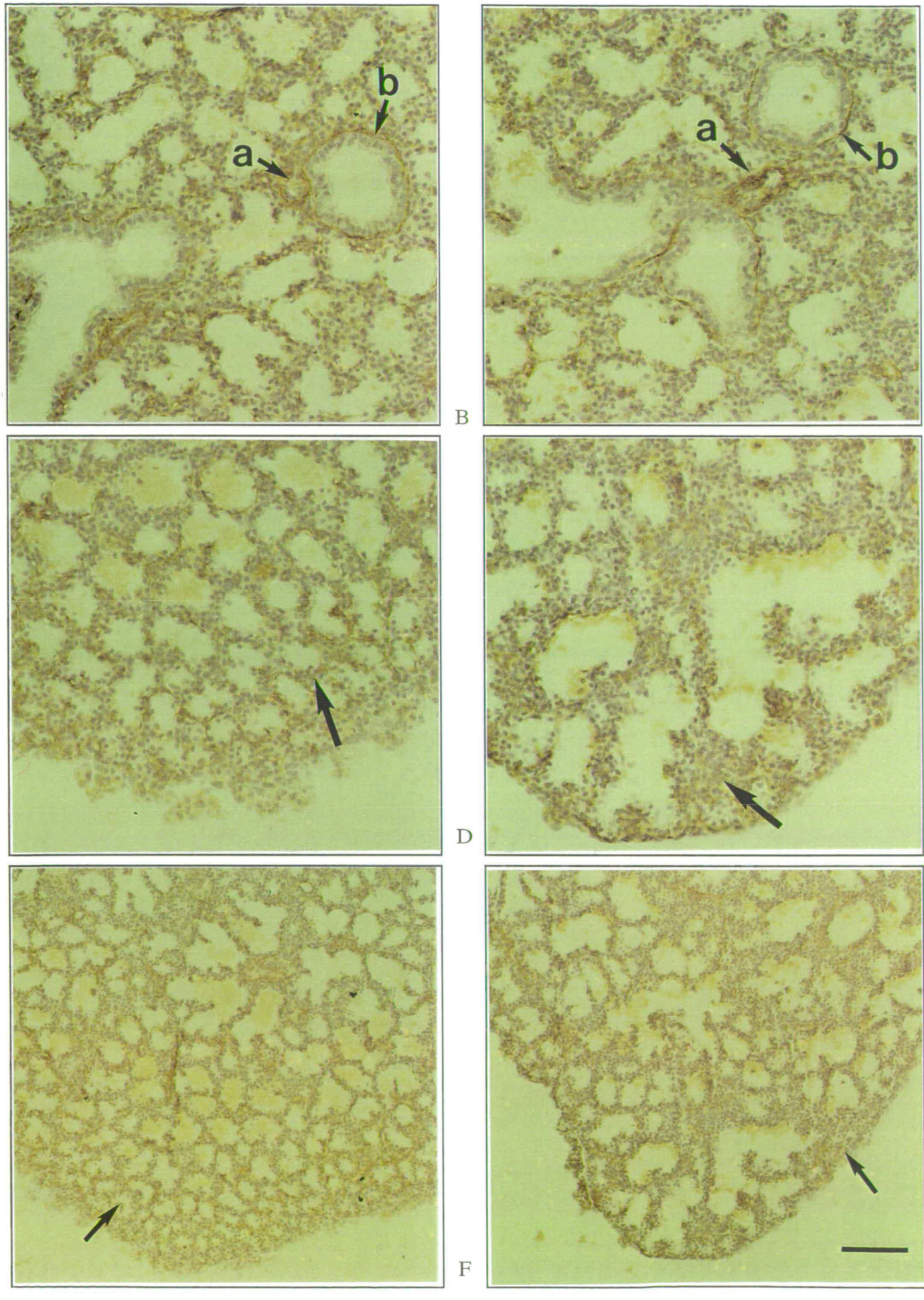

a
b
A
a
b
B
C
D
E
F

Figure 4.14 Smooth muscle actin expression in adult lungs

Lungs taken from 5 week old control mice (A, C & E) and JST185 homozygotes which were suffering with severe breathing difficulties (B, D & F). Lungs were either fixed immediately overnight (A, B, E & F) or inflated with fix prior to overnight fixation (C & D). α-sma (brown) was detected in arterial walls (filled arrows) in both controls and homozygotes (**A**) and (**B**). High levels of α-sma is seen throughout the lung parenchyma in the homozygote lung (open arrowheads, **B** and **D**). Isolated cells expressing α-sma are seen in the control at the tips of the alveolar septa (open arrowhead, **C**). The homozygous lung can be inflated with fix (**D**) however the surface area for gas exchange is greatly reduced compared to the control (**C**). Bar (A,B,C,D) 50μm; (E,F) 100μm.

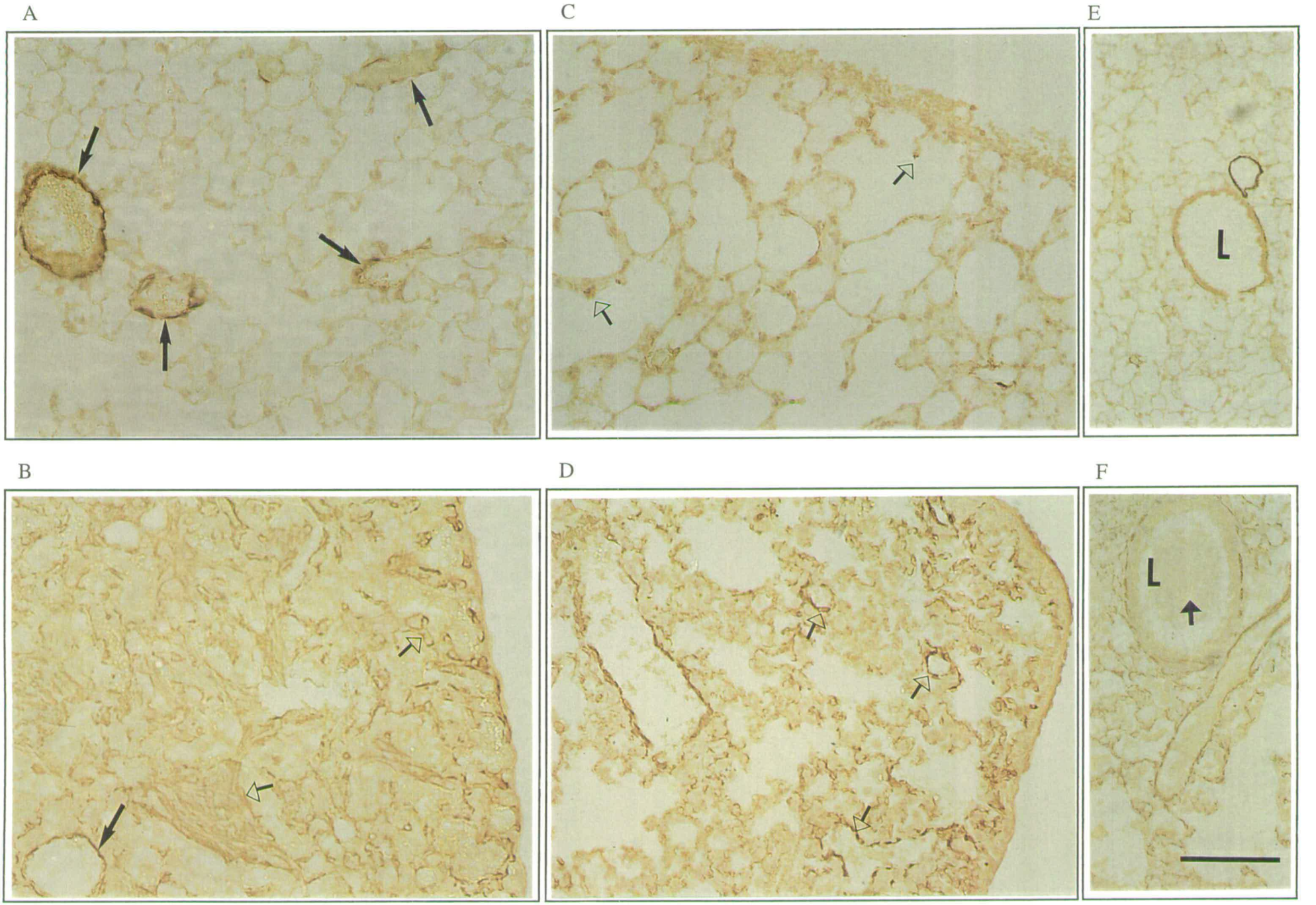
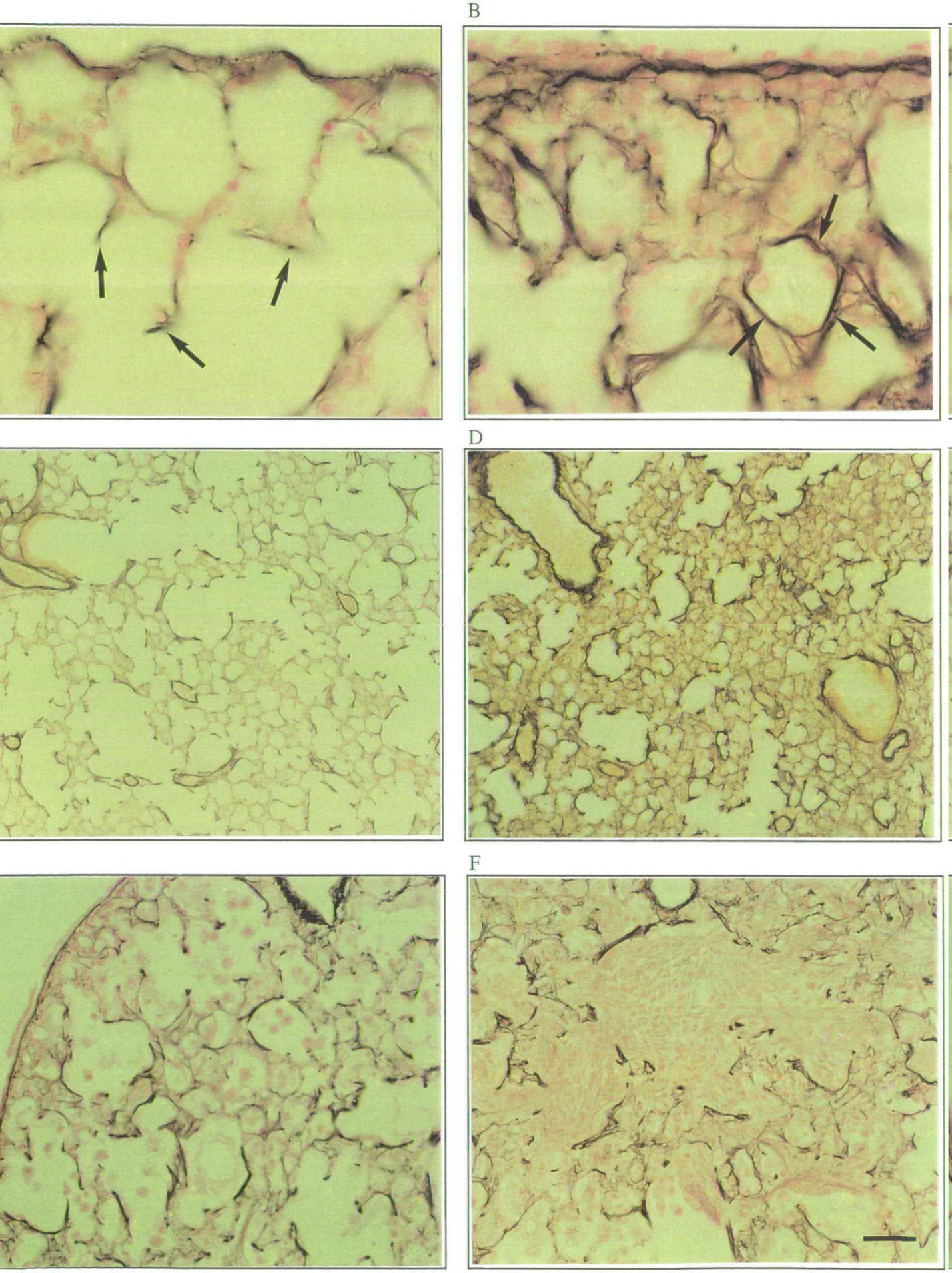

A
C
E
B
D
F
L
L

Figure 4.15 Elastin expression in adult lungs

Lungs were dissected from a 5 week old control mouse (A & C) and a JST185 homozygous littermate (B, D, E & F). Lungs were stained with Miller's and Van Geisons which stains elastin black. In the control lung (**A**) elastin deposits are predominantly seen at the tips of the alveoli septa (arrows), in the mutant lung (**B**) elastin is seen surrounding the alveoli (arrows). At lower power elastin levels are drastically increased in homozygotes (compare **C** and **D**). Regions of the homozygous lungs contain large numbers of cells of unknown identity (**E**), often entire airways are occluded (**F**). Bar (A,B) 10μm; (C,D) 60μm; (E,F) 30μm.


B
D
F

high levels of α-sma expression so may correspond to myofibroblasts (Figure 4.14 B). The adult lungs also show evidence for an inflammatory response in that there are larger numbers of alveolar macrophages in the mutant lung compared with controls (data not shown). There also appears to be an abundance of phagocytic and neutrophil-like cells but confirmation of their identity awaits specific histological stains.

4.8 Identification of a processed pseudogene homologous to JST185

To allow rapid genotyping of animals before processing samples for histology a reverse transcriptase PCR (RT-PCR) strategy was developed. JST185 primers were designed to amplify across the insertion site and downstream of the insertion site (Figure 4.16A). A third set of primers were designed to amplify the insertion site from the fusion transcript (Figure 4.16A). HPRT gene primers were used for each of the RT reactions as positive controls. RT-PCR was performed on wild type, heterozygous and two homozygous RNA samples from 16.5d lung buds and the expected products were observed on an ethidium bromide gel (Figure 4.16B). However, Southern hybridisation with a JST185 probe spanning the insertion site, detected low level wild type bands in the homozygous lanes which were not seen on the ethidium bromide gel (Figure 4.16B). A genomic southern had previously confirmed the genotype of the RNA source and previous data had confirmed splicing around the insertion does not occur.

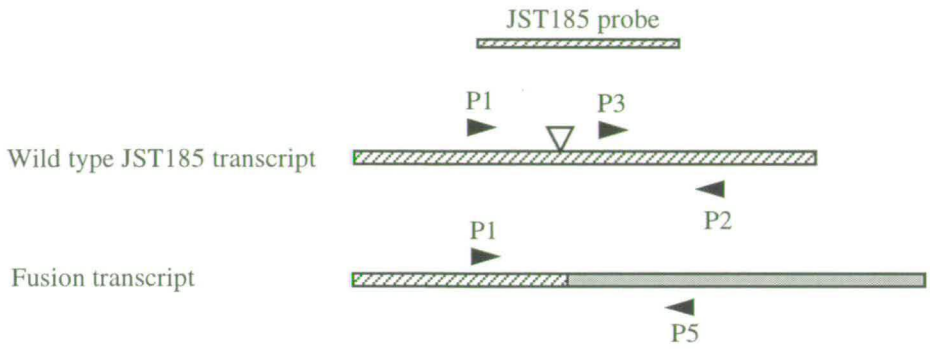
Potentially the faint bands could result from contamination of the PCR reaction with amplified product, or could result from residual genomic DNA in the RNA being used as a template. To distinguish these, DNase I was added to RNA samples before the reverse transcriptase reaction to eliminate gDNA. In the presence of DNAase I, a band failed to amplify in homozygous RNA, showing that gDNA was the source of the previous faint band (Figure 4.16C). As the primer pair (P1 and P2) spans the

Figure 4.16 Identification of JST185 pseudogene

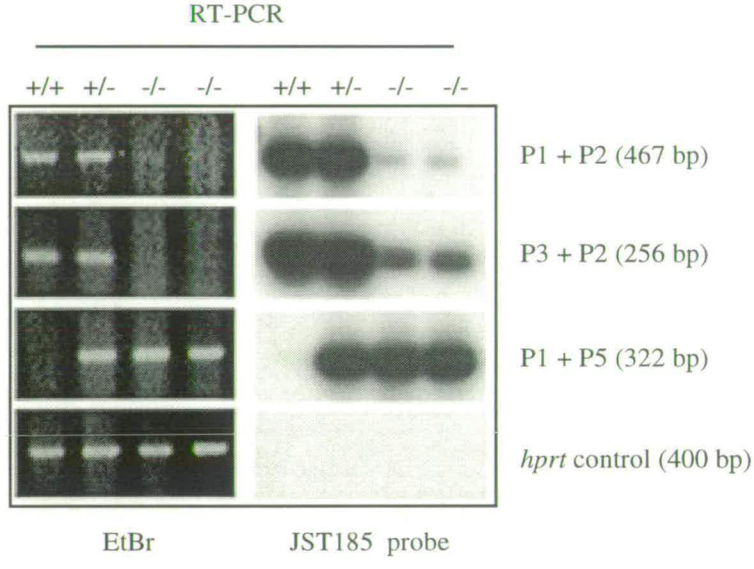
RT-PCR strategy to allow rapid genotyping of JST185 intercross offspring.

(A) Arrangement of primers used to amplify wild type and fusion bands. (Triangle marks position of secretory trap insertion). **(B)** Ethidium bromide gel reveals the expected bands in RNA prepared from wild type, heterozygous and homozygous 16.5dpc lung buds. Southern hybridisation of the RT-PCR samples with a JST185 probe which flanks the insertion site is shown. Bands which were not observed on the ethidium bromide gel are now detected in the homozygous samples. The bands are the size expected to be amplified from wild type transcripts. However, wild type transcripts have previously been shown to be absent from the samples. **(C)** DNase treatment of RNA samples prior to the RT-reaction prevents amplification of the band from homozygous samples. This suggests that the band in the non-DNase treated RNA sample is amplified from gDNA. **(D)** Confirmation that the RT-PCR primers amplify a product from genomic DNA. As the primers reside on exons which are seperated by intron sequences the data suggests there is a processed JST185 pseudogene in the genome.

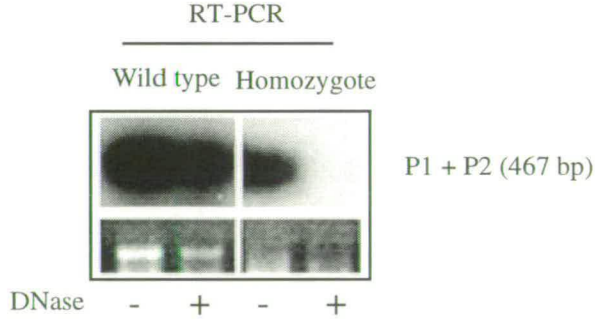
A


JST185 probe
P1
P3
Wild type JST185 transcript
P2
P1
Fusion transcript
P5

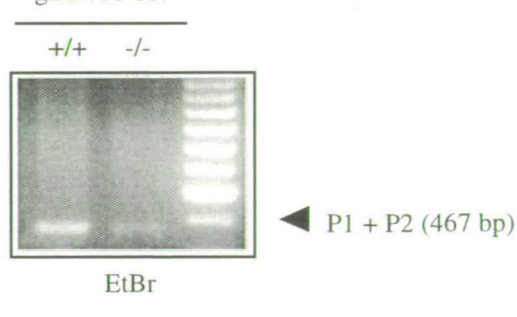
B


RT-PCR
+/+ +/- -/- -/- +/+ +/- -/- -/-
P1 + P2 (467 bp)
P3 + P2 (256 bp)
P1 + P5 (322 bp)
hprt control (400 bp)
EtBr
JST185 probe

C


RT-PCR
Wild type
Homozygote
P1 + P2 (467 bp)
DNase - + - +

D


gDNA PCR
+/+ -/-
P1 + P2 (467 bp)
EtBr

secretory trap insertion site, they are expected to reside on a separate exons. This suggests that amplification from a JST185 processed pseudogene, which lacks intron sequences, may be responsible for the band. Further evidence for the presence of a pseudogene was obtained by performing PCR on genomic DNA using the same primer set and seeing the expected band amplified (Figure 4.16D).

Pseudogenes are DNA sequences that bear significant homology to functional genes, yet they lack promoter sequences for their transcription or contain other mutations that preclude formation of a functional product. They are thought to arise by rare retroposition events and they have no known function.

4.9 Mapping of JST185

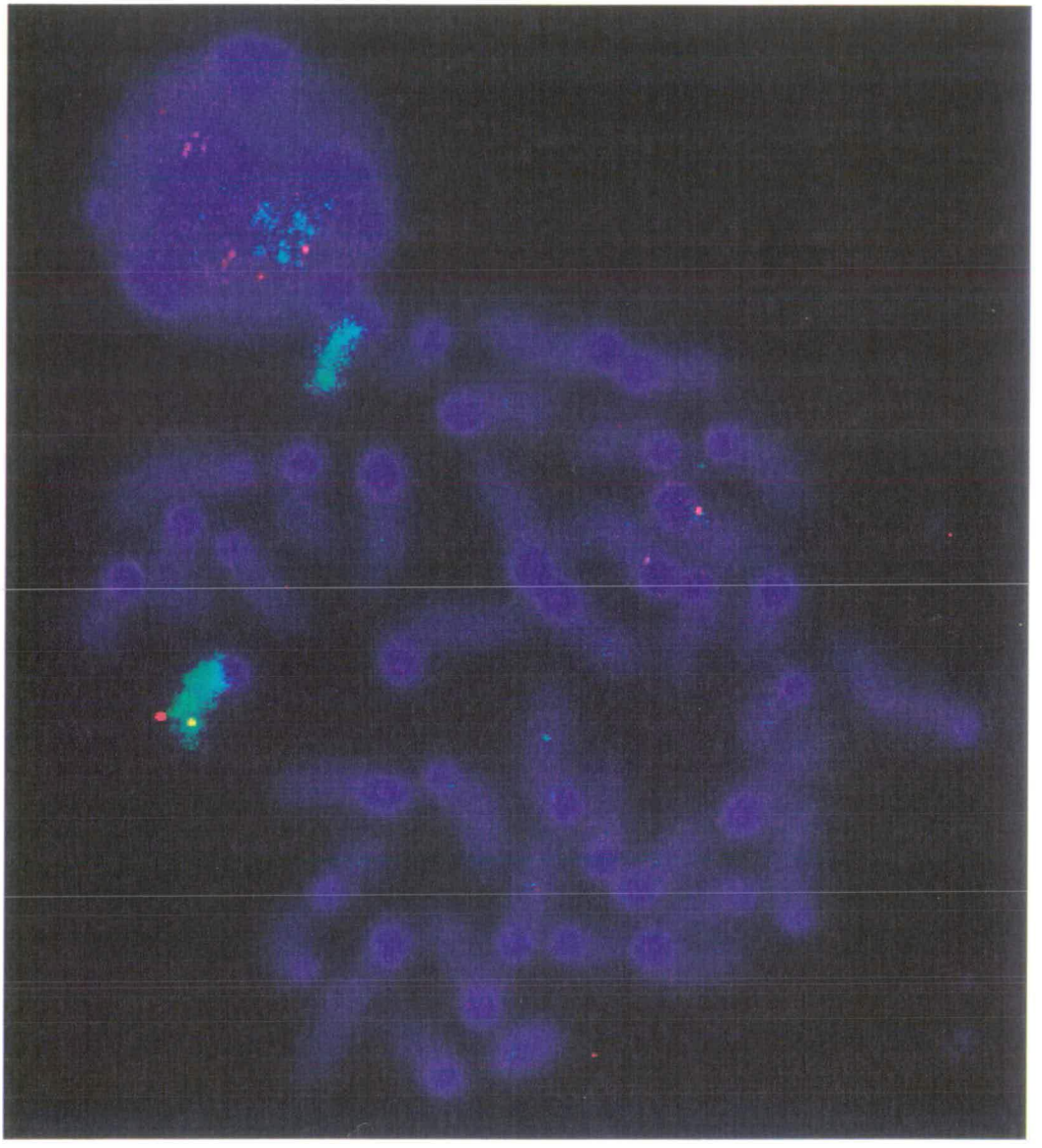
JST185 was mapped by fluorescent *in situ* hybridisation (FISH) by Muriel Lee at the MRC Human Genetics Unit, Edinburgh according to the protocol in Fantes *et al.*, 1992. Metaphase chromosome spreads and interphase nuclei were prepared from JST185 heterozygous spleens. The vector pGT1.8TM was labelled with digoxigenin, hybridised to the samples and visualised using a texas red detection system. A single vector integration site was localised to chromosome 10 by its G-banding pattern. To confirm the chromosomal allocation, paints (using a digoxygenin label and fluoroscein detection system) which are specific for chromosome 10 were hybridised to the spread. The vector and paint were seen to co-localise confirming that JST185 is positioned towards the telomere on chromosome 10 (Figure 4.17).

4.10 Discussion

JST185 carries an insertion in a novel multiple membrane spanning molecule which has been conserved during evolution. It is strongly expressed during the pseudoglandular stage of development, predominantly in the mesenchymal layer, and is down-regulated during the canalicular stage. During the saccula stage, expression is

Figure 4.17 Chromosomal location of JST185

Fluorescent in-situ hybridisation (FISH) to metaphase chromosome spreads localises the secretory trap vector insertion in JST185 to close to the telomere on chromosome 10. The location was confirmed using paints specific to chromosome 10.



observed in isolated cells throughout the lung parenchyma. Most homozygotes die at birth due to respiratory failure. A few survive to adulthood but these have progressively worsening breathing problems which lead to premature lethality.

Structure of JST185

JST185 represents the first report of a membrane spanning molecule containing TPR motifs. The presence of two structurally related C.elegans sequences and a number of human and Drosophila ESTs suggests this structure has been evolutionarily conserved. Furthermore, it is likely that more JST185 family members exist in the mouse. JST185 contains blocks of homology which are conserved across species, degenerate primers can be made to these regions to identify other mouse genes.

It is possible that related molecules functionally compensate for JST185 in the homozygous mutants. The expression pattern of JST185 at mid-gestation suggests it may be involved in signalling between the mesenchyme and epithelium which instructs branching to occur. Histological analysis of new-born lungs at the saccular stage indicates that JST185 homozygotes have fewer terminal branches compared to controls. The role of JST185 in branching morphogenesis may be partially compensated for by a related family member.

The TPR domains may be important in relaying signals from the membrane. To examine this the TPR domains can be used as affinity reagents in a yeast two-hybrid screen to identify interacting proteins. This may allow a convenient entry point to studying any potential signalling pathways mediated by JST185. The TPR domains of JST185 are predicted to be positioned in the intracellular domain. Recently an intracellular SH-2 binding protein was discovered which contains 15 TPR motifs (Malek *et al.*, 1996). The authors speculate that the protein may function as a linker molecule between TPR-containing and SH-2 containing proteins.

RNase protection and Northern analysis confirms that splicing does not occur around the secretory trap insertion to produce wild type transcript in JST185 homozygotes. The secretory trap insertion occurred just before the end of the first TPR domain, thus producing a βgeo reporter protein containing all the JST185 transmembrane domains. Such a structure could act as an interfering mutant by binding ligand at the cell surface but failing to transduce a signal as it contains only one complete TPR domain.

Subcellular localisation studies of several βgal fusion proteins have previously shown that they remain sequestered in the ER and do not reach the surface of the cell to perform their function (Skarnes *et al*, 1995). JST185 antibodies can be raised to its N-terminus to examine the subcellular localisation of the wild type and fusion proteins. The multimembrane spanning nature of JST185 suggests that it might normally be targeted to the ER membrane. Interestingly, a multiple membrane spanning molecule in *Drosophila*, called *porcupine*, has been localised to the ER membrane where it acts as a pore involved in processing or secretion of the *wingless* signalling molecule (Kadowaki *et al.*, 1996). The Wnt gene family are the vertebrate homologues of *wingless* and in the lung at least one family member, *Wnt-2*, is expressed in the distal mesenchyme of the terminal buds (Bellusci *et al.*, 1996).

JST185 phenotype

The majority of JST185 homozygotes fail to make the transition to air breathing at birth. Analysis of the lungs of newborn pups indicate they fail to expand and fill with air. Homozygote pups attempt to breathe as indicated by their costal movements at birth, also, sections through their diaphragms suggest no structural abnormalities. The failure of the lungs to inflate does not appear to be due to an inability to clear the lungs of fluid at birth as there is no difference between the dry and wet weights of mutant and control lungs. Sections through new-born lungs observed

under a light microscope show that the airways are not blocked and that, at a gross morphological level, they are correctly formed. However, there are subtle morphological differences in the airways. Most obvious is their failure to taper into terminal sacs, instead the distal most airways have a diameter similar to the more proximal respiratory bronchioles (Figure 4.13 C & D).

The respiratory problems seen at birth in the JST185 mutants could be due to developmental retardation, over-proliferation of mesenchyme or a block in epithelial cell differentiation. The paucity of distal branches in mutant lungs compared to controls constitutes evidence for retardation. The lung has the appearance of a late canalicular lung rather than a lung at the terminal sac stage. Retardation of lung development would explain the lethality seen at birth, as retarded lungs produce lower levels of pulmonary surfactant. With diminished surfactant, surface tension is elevated, the alveoli are unstable and so collapse.

Alternatively, the abnormal appearance of new born lungs could be due to an excessive production of mesenchymal tissue surrounding terminal airways compared to the controls. Excessive mesenchyme would explain why there is no difference in lung weight between homozygotes and controls although the homozygous lungs are smaller. The abnormal alveogenesis seen in the few homozygous survivors may be a consequence of mesenchymal over-proliferation which prevents the septal walls of the distal airways from thinning out to form the alveoli.

A mesenchymal over-proliferation phenotype is observed in Hoxa-5 mutants of which over 50% show a peri-natal lethality (J. Aubee, pers communication). Similar to JST185, Hoxa-5 is expressed at high levels in the mesenchymal components of the trachea and lung primordia. This may suggest that JST185 and Hoxa-5 are part of the same pathway. It will therefore be interesting to see whether

Hoxa-5 expression is altered in JST185 mutants or vice versa as this may provide insight into which is upstream of the other.

Transgenic mice over-expressing Shh in the entire lung epithelium also show a similar phenotype to the JST185 lungs (Bellusci *et al*, 1997). Shh is normally expressed in the distal tips of the epithelium but in transgenic mice the surfactant protein C promoter drives its expression throughout the distal epithelium of the lung, but not in the proximal airway. Histological defects becomes apparent at the canalicular stage where there is an increase in the relative proportion of mesenchyme to epithelial tubules, however, the organisation of the epithelia looks normal. In the transgenic lung during the saccular stage the architecture of the lung is similar to the JST185 mutants, in that expansion of the airway occurs but the mesenchymal component is much thicker than normal. At birth the transgenic lungs fail to inflate. In the transgenic mice *in vivo* BrdU labelling indicated that Shh overexpression results in increased mesenchymal and epithelial cell proliferation at 16.5 and 17.5dpc. It will therefore be interesting to examine Shh expression in JST185 mutants and vice versa, and use BrdU labelling to demonstrate whether over-proliferation occurs in JST185 mutants.

It is possible that the early expression of JST185 in the mesenchyme is required to promote the differentiation of respiratory cells in the pulmonary epithelium. A failure of type II epithelial cell differentiation would account for the failure of mutant JST185 lungs to inflate at birth, as these cells provide the lung with pulmonary surfactant. During the pseudoglandular stage, when JST185 is expressed at its highest level, the type II cell precursors, which stain for apoprotein A, are present (Thurlbeck, 1996). Loss of surfactant protein B by type II cells has previously been shown to lead to respiratory distress syndrome and peri-natal lethality (Clark *et al.*, 1995).

Further evidence to suggest that JST185 animals might be suffering from surfactant insufficiency comes from the phenotype of the few adult survivors. The pathology of the mutant adult lungs is similar to the chronic lung condition, bronchopulmonary dyplasia (BPD), the most common cause of chronic lung disease in infants (reviewed in Horowitz & Davis, 1997). Most of the pathological hallmarks of this disease are seen in JST185 mutant lungs, including impaired alveolar development, oedema, fibroblast proliferation and inflammatory cell infiltration. There is no single cause for BPD in all infants but it often occurs in prematurely born infants with respiratory distress syndrome which have been treated for surfactant deficiency. Similar to the JST185 mutants, in the early stages of BPD there is increased pulmonary resistance which increases the work of breathing. This increased resistance is thought to be due to the surfactant deficiency causing significant alveolar collapse.

One of the mechanisms proposed to contribute to the pathology of BPD is a protease-antiprotease imbalance which interferes with the layout of elastin leading to growth without proper alveolarisation (Robertson, 1989). Histological staining of JST185 mutant lungs reveals drastic changes in the distribution of elastin. Mechanical ventilation of premature neonates is often associated with BPD, as the high pressures required to open collapsed alveoli cause significant damage to the terminal bronchioles and alveolar ducts. A recent model for ventilation induced BPD in preterm lambs found an excessive production and accumulation of elastin in alveolar walls (Pierce *et al.*, 1997).

Immunohistochemistry reveals marked changes in the level and distribution of α-sma in the homozygous lungs. The increased levels indicate there may be an over-proliferation of mesenchymally derived lung fibroblasts. Abnormal distribution of α-sma has also been observed in the lungs of transgenic mice over-expressing TGF-β1 (Zhou *et al.*, 1996). These animals show severe defects in sacculation. The authors

suggest that TGF-β1 influences myofibroblast differentiation and that the abnormal distribution of myofibroblasts tethers the distal airways and inhibits the expansion of terminal lung buds leading to abnormal sacculation.

To further investigate the role of developmental immaturity in the JST185 homozygotes, pregnant females of heterozygous intercrosses can be treated with dexamethasone and progesterone. Dexamethasone promotes epithelial cell maturation and progesterone delays parturition by one day, prolonging *in utero* development time. In this way we may increase the proportion of JST185 homozygotes which survive. Examination of the levels of surfactant proteins in JST185 new born animals may also indicate whether the lung epithelium is either immature or lacking type II epithelial cells. Ultrastructural studies can also be used to differentiate between these two possibilities as the type II cells are easily recognisable by the presence of large surfactant filled lamellar bodies.

Chapter 5

CONCLUDING REMARKS

The results presented in this thesis demonstrate that the secretory trap approach provides a highly efficient means to screen for novel membrane spanning molecules essential for mouse development and also allows rapid identification of the mutated genes by direct sequencing of 5'RACE products (Townley *et al*, 1997). Using this approach, 58% [7/12] of insertions were shown to occur in secretory genes. Based on the analysis of the remaining insertions, a pre-screen has been developed which should ensure that up to 100% of insertions occur in secretory molecules. This involves only selecting lines which show the secretory pattern of βgal activity in ES cells.

Detailed analysis of secretory trap line JST185 shows that gene function is disrupted since splicing does not occur around the secretory trap insertion. This demonstrates that the vector can act as an effective mutagen. The developmental importance of JST185 is clear, given the recessive lethal phenotype of homozygotes. Recent results from ongoing experiments in this laboratory have shown that 30% [10/28] of secretory trap insertions into secretory molecules cause recessive lethal

phenotypes (S.J. Monkley, J.Brennan, D. Townley, B.J. Avery and W.C. Skarnes, unpublished data).

71% [5/7] of JST insertions which disrupted secretory genes are expressed in a restricted pattern. JST185 exhibits one such pattern in the developing lung and results in aberrant lung development. This indicates that, if used in conjunction with a screen for tissue specific expression patterns, the secretory trap approach can be used to identify novel genes involved in particular developmental pathways. However, other genes with restricted expression patterns, such as JST213, do not result in abnormal phenotypes. Results from the ongoing experiments mentioned above, indicate that disrupting genes with restricted expression patterns gives rise to recessive lethal phenotypes at a frequency similar to those exhibiting widespread expression patterns. This may reflect the fact that in signalling pathways the ligand or receptor can often be widely expressed, whilst its partner exhibits a more restricted expression profile.

In a previously conducted pilot screen using pGT1.8TM, two of the six lines examined had mutations in the same gene (Skarnes *et al*, 1995). This raised the possibility that the secretory trap vector would be unable to detect insertions in more than a subset of secretory genes. None of the 13 JST lines analysed here were repeat events indicating that this may not be a real problem.

Data from our lab and others suggests that electroporated constructs are biased towards inserting into particular genes, as multiple independent insertions have been detected in the same gene using different vectors (Chapter 1). Therefore the best way to achieve a saturation screen would be to use a combination of vector constructs with different integration preferences. The results presented in Chapter 3 confirm the feasibility of an exon trap approach and also indicate ways in which such an approach might be improved. For example, if a random promoter trap screen is to be performed on a large scale, a retroviral based vector would be advisable. Such a vector would

not suffer deletions from its 5' end and would thus reduce the effort required to identify disrupted genes. However, a selective screen for secretory molecules requires insertions to occur in coding regions, so a plasmid based vector must be used. In such a screen it is vital that the selective property of the vector is protected from deletion. This may be achieved by the inclusion of either buffer sequences or a hairpin loop at the 5' end of the vector.

In summary the data described in this thesis has shed new light on the selective mechanism of the secretory trap vector and has indicated ways in which the selective properties may be improved. More importantly this data shows that the secretory trap approach is an efficient way to generate a resource of mutants carrying insertions in novel secretory molecules that are immediately accessible to molecular characterisation and functional analysis.

Chapter 6

MATERIALS AND METHODS

6.1 Materials

Unless stated otherwise, all chemicals were of analytical grade and supplied by BDH Laboratory supplies or Sigma, restriction and modifying enzymes were from Boehringer Mannheim. All bacterial media constituents were supplied by DIFCO laboratories and electrophoresis grade agarose was supplied by GIBCO BRL. Synthetic oligonucleotides were supplied by Oswel, University of Southampton. Radioisotopes were supplied by Amersham and X-ray film was supplied by Kodak.

Computer software packages used for DNA analysis were MacVector (V. 5), University of Winconsin Genetics Computer Group (1994), GCG package (V.7).

6.1.1 DNA vectors

Commercially available:

pBluescript KS II (-) (Stratagene)

pBluescript SK II (-) (Stratagene)

Constructed/provided:

pGT1.8TM (W.C. Skarnes, Figure 6.1)	stock #78
pET1.8TM (J. Brennan, Figure 6.2)	stock #114
p1.8HX (W.C. Skarnes, Figure 6.3)	stock #95
pSA1b (W.C. Skarnes, Figure 6.4)	stock #106
pSAβgeoΔEK (W.C. Skarnes, Figure 6.5)	stock #94
pGTΔTM1 (W.C. Skarnes, Figure 6.6)	stock #90

6.1.3 Probes for hybridisation

JST185 Probes:

Probe	Restriction fragment	Plasmid
RACE probe (genotyping)	520bp XbaI/Asp718	JST185 RACE clone #2
5' cDNA (library screen)	190bp NotI/XbaI	Original JST185 cDNA
3' cDNA (library screen)	170 BamHI	Original JST185 cDNA
200bp RACE (RPA)	200bp HindII/BsshII	JST185 RACE clone #2
480bp cDNA (RPA)	480bp NcoI/BsshII	PvuII/Xba 650 (subcloned from original cDNA)
5' probe (Northern)	965bp HindIII/SacI	JST185 cDNA #10

Figure 6.1

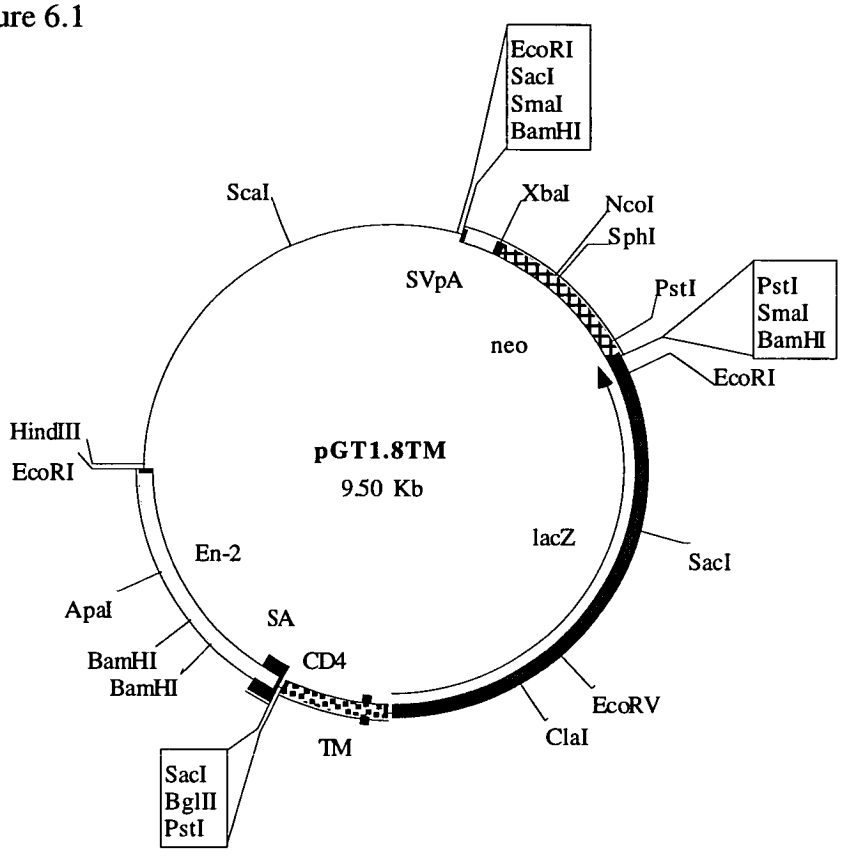


Figure 6.2

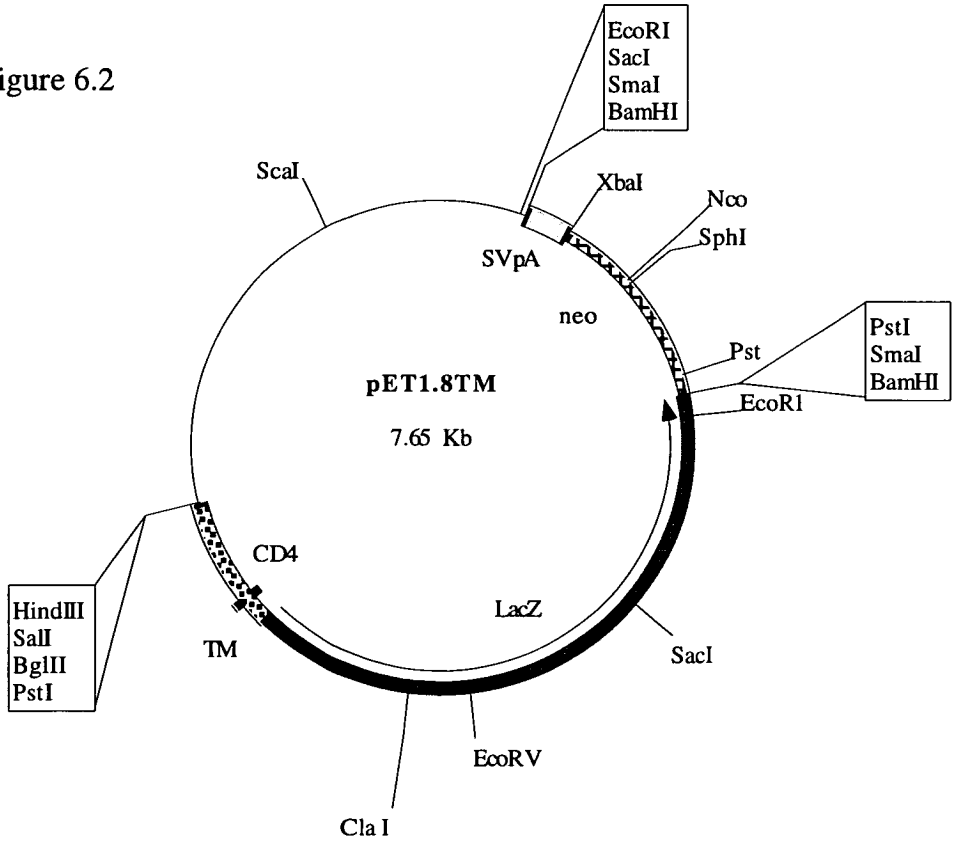


Figure 6.3

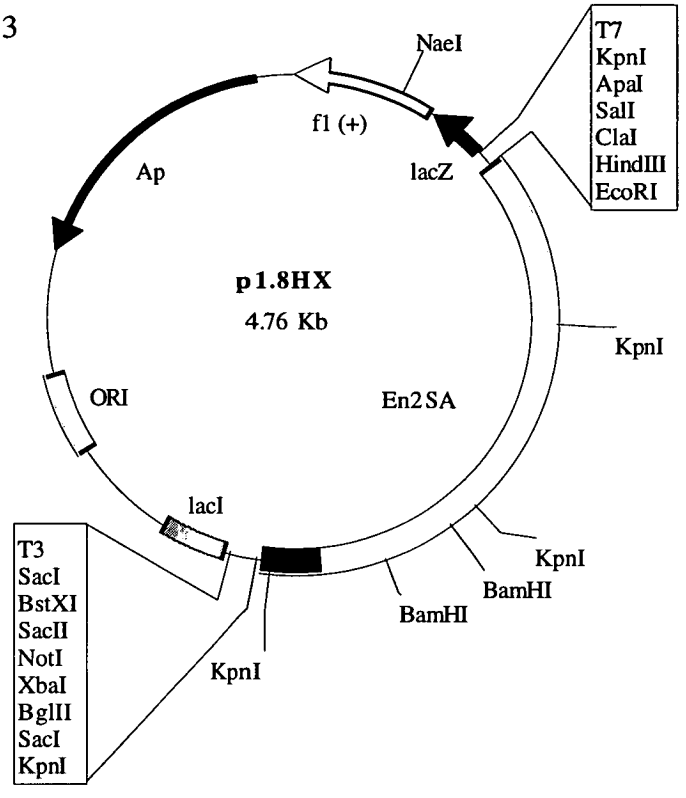


Figure 6.4

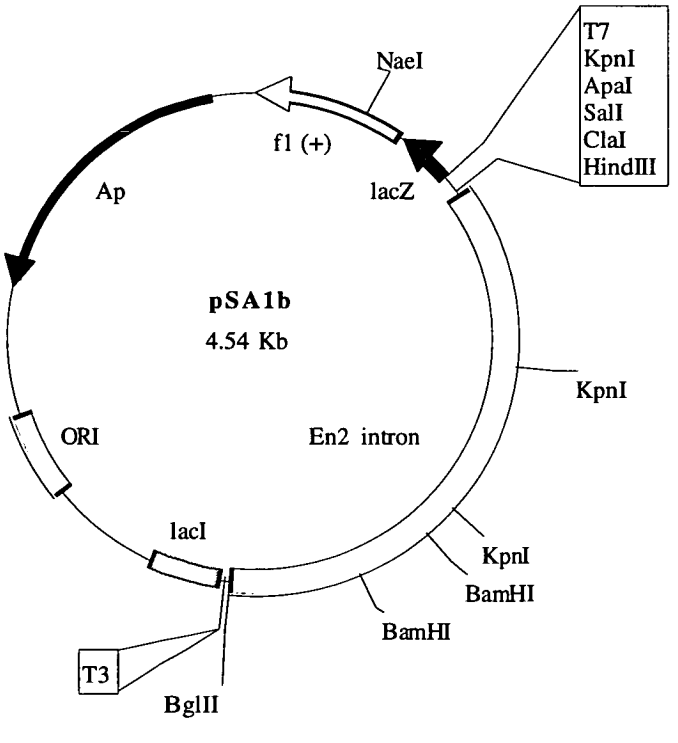


Figure 6.5

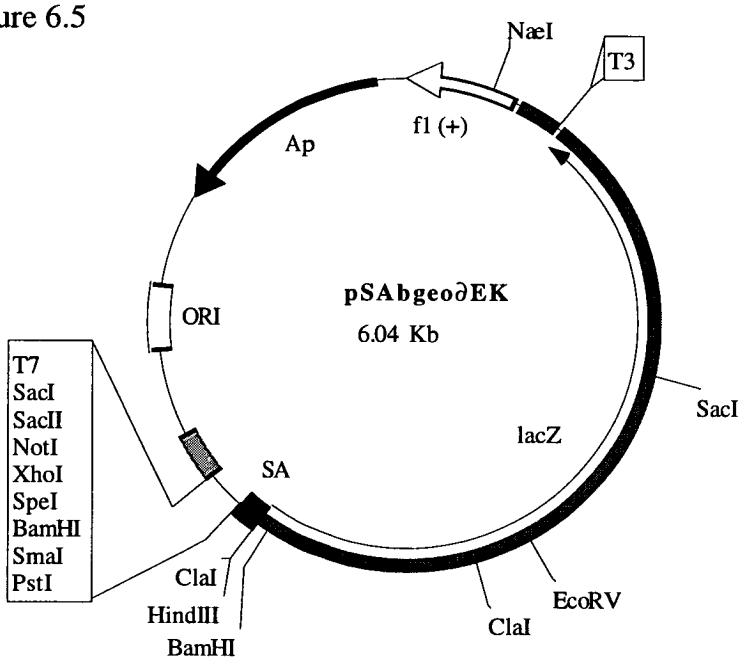
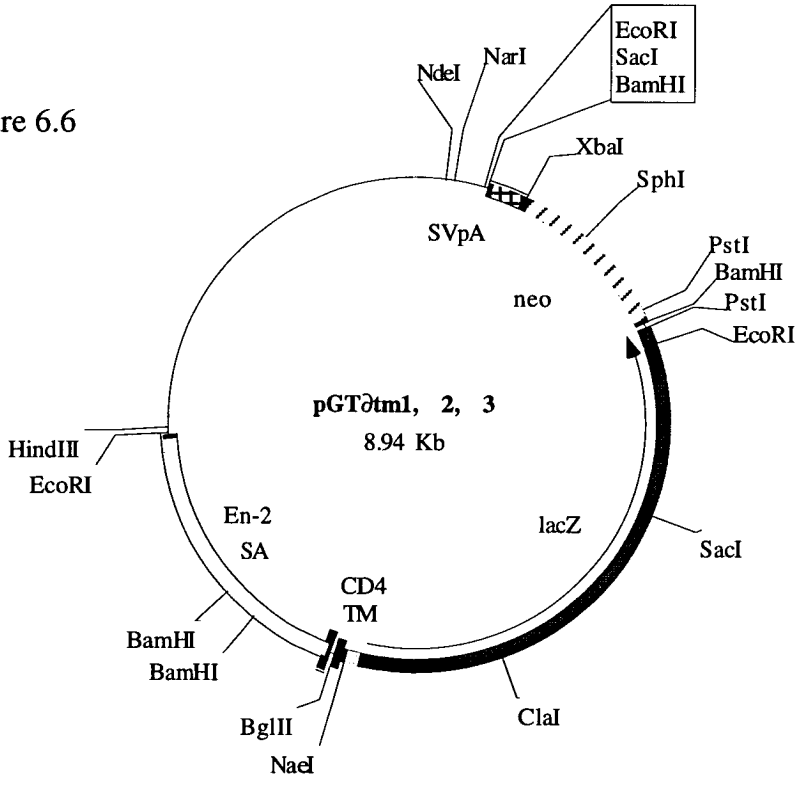


Figure 6.6



Probe	Restriction fragment	Plasmid
Neomycin	1.1kb Xba/BamHI	pGTΔTM1
LacZ	0.8kb ClaI/SacI	pSAβgeoΔEK
En2 exon	0.5kb BamHI/BglII	p1.8HX
En2 intron	1.5kb HindIII/BglII	pSA1b
ET105	330bp XbaI/HindIII	ET105 RACE clone #13

6.1.4 Primers

pGT1.8TM RACE primers (Primer numbering based on Figure 2.6, see also Figure 2.7 for position of primers within vector sequences)

Primer	Stock #	Sequence (5'→3')
P1	92	ccagaaccagcaaactgaaggg
P2	56	ggttgtgagctcttctagatggttttttttttttttttt
P3	67	agtagacttctgcacagacacc
P4	59	ggttgtgagctcttctagatgg
P5	55	tgctctgtcaggtacctgttgg
P6	105	ggttgtgagctcttctagatgg (5'biotinylated)
P7	98	agcagtgaaggctgtgc

pET1.8TM LM-PCR primers (Primer numbering based on Figures 3.4 and 3.5)

Unidirectional adapter was made by ligating the following primers:

#144 5' gctagtctagagggcccagatctgaattc 3'

#145 5' gaattcagatc 3'

Primer	Stock #	Sequence (5'→3')
P4	146	gctagtctagagggcccagat
P5	155	gctagtctagagggcccagat (5'biotinylated)

The positions of the nested primers designed against pET1.8TM sequences are illustrated in Figure 6.7. The primer numbers refer to the following nested primers:

Located within LacZ sequences:

#78 5' taatgggataggttacg 3'

#79 5' agtatcggcctcaggaagatcg 3'

#80 5' attcaggctgcgcaactgctgttgg 3'

#166 5' cgccagggttttcccagtcacgac 3'

Located within CD4 sequences:

#147 5' aataggatgcagagccccgt 3'

#148 5' cagcacgacagccaggaacat 3'

#93 5' ccttcactcagtagacattgcc 3'

#149 5' agacattgccacacccctgct 3'

#150 5' aggtcaaagtgttgctgtcgg 3'

The same nested primers were used to RACE from ET lines.

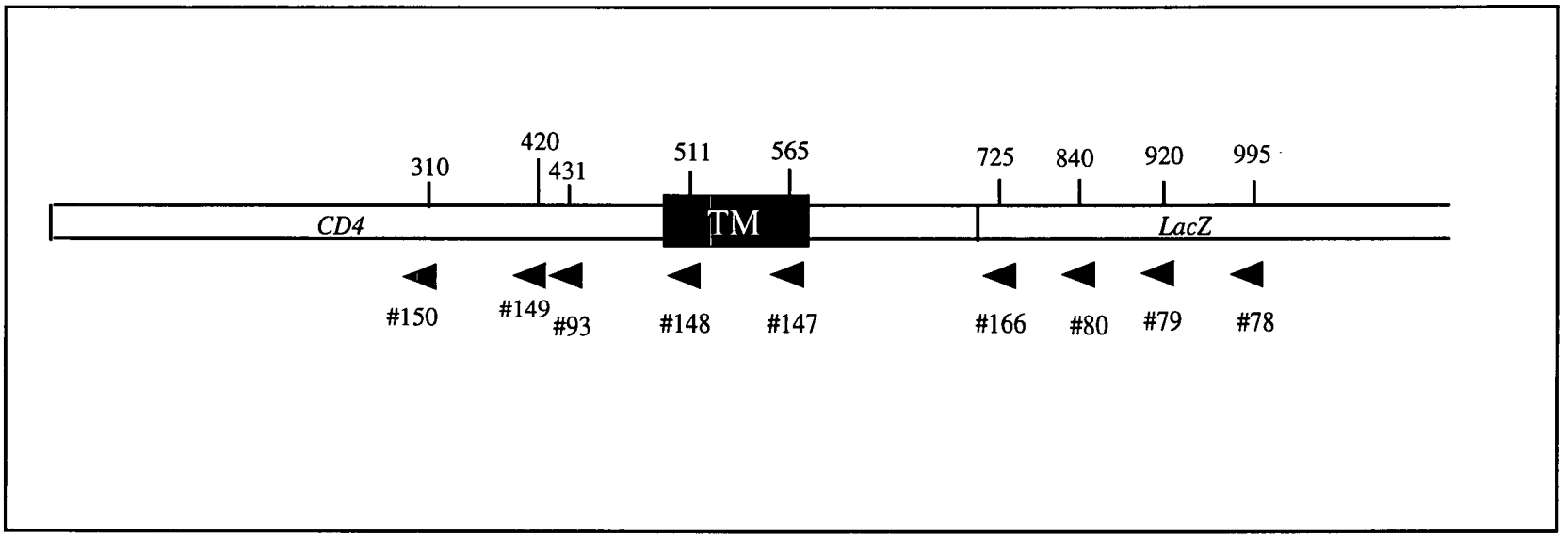


Figure 6.7 Primer sites in pET1.8TM
Primer number is indicated below arrowhead. Position in nucleotides within pET1.8TM vector is indicated above the map.

6.1.5 *Escherichia coli* strains & culture media

Escherichia coli strains

XL1-blue (Stratagene)

DH5α (Stratagene)

SolR (Stratagene)

Culture media

Luria Broth (LB), per litre (in H_20)		LB plates
Tryptone (Difco)	10 g	1 litre LB
Yeast extract (Difco)	5 g	15 g agar (Difco)
NaCl	5 g	

LB Top Agarose

250 mls LB

1.79 g Agarose

SM

100mM NaCl, 10mM $MgSO_4$, 50mM TrisCl (pH7.5), 0.01% gelatin

Antibiotics (Stock solution)

Ampicillin (100mg/ml in H_20)

Kanamycin (10mg/ml in H_20)

Tetracycline (5mg/ml in ethanol)

6.1.6 Embryonic stem cells, culture media & solutions

CGR-8 ES cells (derived from 129/OLa mice by J Nichols)

CGR-8 ES cells are feeder-independent and rely on an exogenous source of differentiation inhibitory activity, DIA (also known as leukocyte inhibitory factor, LIF). These cells are karyotypically male and were derived from the 129/OLa strain of mice as described in (Nichols *et al.*, 1990).

Cell culture medium: (per 400 ml)

10X Glasgow MEM/BHK12 (Gibco), store 4°C	40 ml
7.5% sodium bicarbonate (Gibco), store 4°C	13.2 ml
1X MEM non-essential amino acids (Gibco), store 4°C	4 ml
200 mM glutamine, 100 mM sodium pyruvate (Gibco), store -20°C	8 ml
0.1M 2-mercaptoethanol (Sigma), store 4°C for 1 month	0.4 ml
foetal calf serum (Globepharm, Surrey), batch tested, store -20°C	40 ml
sterile, deionised water	340 ml

(Store up to 1 month at 4°C)

Differentiation inhibiting activity (DIA)

DIA was prepared by D. Colby and D. Rout at the CGR by transient expression of murine or human DIA/LIF expression plasmids in COS-7 cells using the method described in (Smith, 1991). Serial dilutions of the supernatant were tested on ES cells plated in 24-well plates. A 100-fold higher dilution than the minimal dilution required to keep ES cells undifferentiated was typically used. Serum batches were tested for their ability to sustain the growth, differentiation

and viability of ES cells grown at clonal density in the presence and absence of DIA.

Phosphate Buffered Saline (PBS)

One PBS tablet (Dulbecco A; Oxoid BR014G) was dissolved in 100 mls water to give a PBS solution pH 7.3 which was filter sterilised.

Trypsin solution

Dissolve 250 mg of trypsin (Difco) and 372 mg EDTA disodium salt (Sigma) in 1 litre of PBS. Add 10 ml of chicken serum (Flow Labs) and filter sterilise. Store in 20 ml aliquots at -20°C.

1% Gelatin

Add 1 g of gelatin (Sigma,) to 100 ml dH_2O, autoclave and store in 20 ml aliquots at 4°C. For a working solution of 0.1%, add 10 ml of 1% gelatin to 90 ml of PBS.

0.1M 2-mercaptoethanol

Add 100 ml 2-mercaptoethanol (Sigma) to 14.1 ml PBS. Store up to 1 month at 4°C.

Geneticin (G 418; Boehringer-Mannheim)

200 mg/ml dissolved in dH_2O.

Freezing solution

1 ml of DMSO to 9 ml of cell culture medium (made fresh).

6.1.7 Miscellaneous solutions

Standard solutions were made according to (Maniatis et al., 1982).

TE

10 mM Tris HCl, pH 8.0 (20°C); 1 mM EDTA, pH 8.0

DEPC H_2O prepared in a fume hood by adding 1 ml of diethyl pyrocarbonate (Sigma) to 1 litre of dH_2O, shake thoroughly and leave overnight before autoclaving

NaP buffer (1 M sodium phosphate)

Dissolve 70.5 g of anhydrous Na_2HPO_4 in dH_2O, add 4 ml of phosphoric acid and make up to 1 litre with H_2O. Filter sterilise.

TEN

25mM Tris Cl (pH7.4), 24mM EDTA, 75mM NaCl

6.2 Molecular biology methods

Many of the methods described in this section are based on those described in Sambrook et al., 1989.

6.2.1 General cloning techniques

Restriction fragments and PCR products required for further analyses were routinely subcloned into the plasmid Bluescript II KS (-) This typically involved the following steps.

(a) Gel purification.

Restriction digests and PCRs were fractionated by electrophoresis on agarose gels (usually cast in 1x TAE) of the appropriate concentration containing ethidium bromide (0.5 μg/ml). Bands of interest were excised with clean scalpels under long wavelength ultraviolet illumination and DNA recovered using the QIAEX II kit

[QIAGEN] according to the manufacturer's instructions. DNA is bound to a silica resin, subjected to a series of washing steps which eliminate non-nucleic acid impurities and finally eluted in TE or water.

(b) Ligation.

Ligation reactions were usually set up with a vector to insert ratio of 1:3 by molarity. Reactions were performed in 10 µl volumes to include 1 µl each of T4 DNA Ligase (1 Weiss unit) [Boehringer] and a 10 x ligation buffer (containing ATP) [Boehringer]. A control reaction containing vector alone was always set up to determine the background number of colonies due to self-ligation of vector. A successful ligation reaction usually produced an enrichment of more than 100-fold colonies over the vector-alone ligation. Cohesive-end ligations were performed at 16°C for a minimum of 8 hours while blunt-end ligations were allowed to continue overnight at room temperature.

(c) Transformation.

Electrocompetent cells were prepared according to the following procedure:

1. Inoculate a single colony of the required bacterial strain, usually DH5α or XL1-Blue (from a recently streaked plate), into 25 ml LB broth. Incubate at 37°C overnight with vigorous shaking.
2. Inoculate 500 ml LB broth with 5 ml of the overnight culture. Incubate at 37°C with vigorous shaking until OD_{600}=0.3.
3. Chill on ice for 15 minutes. Centrifuge at 4000 x g for 10 minutes at 4°C. In this and all subsequent steps, rotors, centrifuge bottles, tubes and pipettes must be kept cold.
4. Discard supernatant and resuspend bacterial pellet in 10 ml cold water. Bring volume to 500 ml with cold water. Centrifuge at 4000 x g for 10 minutes at 4°C.
5. Repeat step 4 but reduce volume of water to 250 ml.
6. Repeat step 4 but resuspend pellet in 10 ml ice-cold 10% glycerol.

7. Discard supernatant and resuspend the bacterial pellet in a final volume of 1 ml with 10% glycerol.

8. Dispense into 40 µl aliquots. Quick-freeze in liquid nitrogen and store at -80°C.

1-2 µl of the ligation reaction was mixed with 40 µl of electrocompetent cells and incubated on ice for 1 minute prior to electroporation in a prechilled cuvette [Biorad]. The electroporation conditions using Biorad's Gene Pulser were 25 mF, 200 ohm and 2.5 kV (for 0.2 cm electrode gap) and usually resulted in a time constant of 4.5. 600ml of LB was immediately added to the cells following electroporation and were allowed to recover for 1 hour at 37°C with gentle shaking. Appropriate amounts were plated onto LB-Ampicillin plates. Transformation efficiencies of 2 x 10^{10} transformants per mg of supercoiled plasmid are typical with fresh electrocompetent cells. This was found to fall 10-fold upon freezing.

6.2.2 Isolation of nucleic acids

(a) Small scale preparation of plasmid DNA

A single colony was picked with a sterile gilson tip and used to inoculate 5ml of L broth, containing 10mg/ml Ampicillin (Sigma). This culture was shaken at 37°C overnight and DNA prepared according to the following protocol:

1. Cells were harvested by centrifugation for 3 minutes at 13,000 rpm in a 1.5 ml Eppendorf tube.

2. The pellet was resuspended in 200 µl of TGE (24 mM Tris-HCl pH 8.0, 50 mM glucose, 10 mM EDTA).

3. 400 µl of freshly made lysis buffer (0.2 M NaOH/1% SDS) was added, mixed by inversion and stored on ice for 5 minutes.

4. 200 µl of solution III (3 M Potassium Acetate/5 M acetic acid; 100 ml = 60 ml 5M KAcetate, 11.5 ml glacial acetic acid, 28.5 ml water) was added, mixed by vortex and stored on ice for 10 minutes.
5. Th.e resulting precipitate was removed by 10 min centrifugation at 13000 rpm and 700 µl of supernatant was transferred into a clean Eppendorf tube.
6. The DNA was phenol/chloroform extracted and then precipitated by the addition of 360 µl of propanol.
7. After 10 minutes on ice the DNA was pelleted by centrifugation at 13000 rpm for 10 minutes, washed in 70 % ethanol and then air dried.
8. The DNA was resuspended in 50 µl of TE containing 20 µg/ml of RNase A

(b) Large scale preparation of plasmid DNA

Plasmids required for electroporation into cells were prepared according to the following protocol.

1. A 10 ml overnight culture of the appropriate colony was used to seed a 250 ml culture.
2. The cells were harvested by centrifugation at 4000 rpm for 15 minutes at 4°C.
3. The bacterial pellet was then resuspended in 9 ml of lysis buffer (see above).
4. 1 ml of freshly prepared lysozyme (10 mg/ml in 10 mM Tris-HCl, pH 8.0) was added.
5. 20 ml of 0.2 M NaOH/1% SDS was added, mixed by inversion and stored at room temperature for 5 minutes.
6. 10 ml of ice cold 3 M KAcetate/5 M acetic acid was added and mixed by shaking.
7. The bacterial lysate was then centrifuged as before and the rotor allowed to stop without braking.
8. The supernatant was removed and filtered through glass wool.
9. The plasmid DNA was then precipitated with 0.6 volumes of isopropanol.

10. The DNA was recovered by centrifugation at 4000 rpm for 15 minutes at room temperature, washed with 70% ethanol and then resuspended in 10 ml TE buffer.

The plasmid DNA was then purified by CsCl-ethidium bromide gradient.

11. To the above DNA preparation 10.8 g of CsCl was added and dissolved followed by 650 µl of 10 mg/ml ethidium bromide.
12. This mixture was then centrifuged at 4000 rpm for 15 minutes and the supernatant loaded into an ultra centrifuge tube and sealed avoiding air pockets.
13. Centrifugation was performed at 55,000 rpm overnight at 20°C in a Beckman NVT65 rotor.
14. The band corresponding to plasmid DNA (see (Maniatis et al., 1982) was removed and the ethidium bromide removed by extraction with CsCl saturated isopropanol.
15. TE buffer was added to 4 ml and the DNA precipitated with 8 ml ethanol.
16. The pellet was collected by spinning and washed in 70% ethanol, resuspended in 4ml TE buffer, phenol/chloroform extracted twice, chloroform extracted and reprecipitated.
17. The pellet was collected by spinning and washed in 70% ethanol and resuspended at 1 µg/ml in TE buffer.

(c) Preparation of genomic DNA from ES cells

1. 10^7 cells were harvested by trypsinization, washed in PBS and pelleted.
2. The cells were resuspended in 700 µl TEN buffer and transferred to a microcentrifuge tube.
3. SDS was added to 1% and mixed by inversion to lyse cells and nuclei.
4. 0.2 mg of proteinase K (Sigma) was added and the solution incubated at 56^0C for at least 2 hours.
5. DNA was phenol/chloroform extracted.

6. The DNA in the supernatent was precipitated by adding one tenth of a volume of 0.3M sodium acetate and 2 volumes of cold ethanol. The DNA was spooled out with a heat sealed pasteur pipette, rinsed in 70% ethanol and air dried for a few minutes.
7. The DNA was transferred to a fresh microcentrifuge tube containing 200-500 µl TE and dissolved overnight at 4^0C.

(d) Preparation of RNA

RNazol B (Biogenesis Ltd) was used to prepare RNA from cells and tissues. The method used is based on that recommended by the manufacturer.

Preparation from ES cells:

1. A confluent tissue culture well of ES cells was rinsed twice with PBS. A volume of RNazol was added to the cells (1ml RNazol per 6-well; 250µl RNazol per 24-well).
2. The lysate was passed up and down in a gilson pipette several times and transferred to a microcentrifuge tube and placed on ice.
3. One tenth of a volume of chloroform was added and the mixture vortexed for 15 seconds. The sample was left on ice for 5 minutes.
4. The mixture was spun for 15 minutes and the aqueous phase transferred to a new tube.
5. One volume of ice cold isopropanol was added to precipitate the RNA and the sample left on ice for 15 minutes.
6. The RNA was pelleted by spinning for 15 minutes.
7. The pellet was washed with 75% ethanol, air dried and resuspended in DEPC water.
8. RNA was stored at -20^0C.

6.2.3 Dot blot analysis of nucleic acids

(a) Tail biopsy dot blot

A one tube method for preparing genomic DNA from tail biopsies was used. To obtain uniform signals, it was important to apply enough DNA to saturate the membrane. This was determined empirically using a BioRad 96-well dot blot apparatus and Amersham Hybond N^+ membrane.

1. Tail biopsies (1.5 cm in length) were taken at weaning age and digested overnight at 55-65°C in 0.4 ml of tail buffer (recipe below).
2. While tubes are still warm, 0.1 ml 5M NaCl was added and vortexed at high speed for 5-10 sec. 0.5 ml of chloroform was added and vortexed again for 5-10 sec. Samples were spun in a microfuge for 5 min.
3. 50 µl of the top aqueous phase was transferred to a 96-well plate. DNA was denatured by adding 150 µl of 0.53 M NaOH and incubating at 37°C for 30 min.
4. The dot blot apparatus was prepared by cutting a piece of Hybond N^+ membrane and a piece of Whatman paper to fit the apparatus. The membrane was pre-wet in H_2O then soaked in 0.4 M NaOH for 10 min. The Whatman was pre-wet in 0.4 M NaOH and placed underneath the membrane on the apparatus.
5. Samples were applied to the dot blot apparatus and left for 30 min before applying a vacuum.
6. Samples were drawn through with gentle vacuum. Apparatus was disassembled and the membrane washed with 30 mM NaP buffer/0.1% SDS buffer. The membrane was hybridised to a neo probe.

Tail buffer (made up fresh): to 10 mM Tris-HCl (pH 8.0)/100 mM NaCl/ 50 mM EDTA/0.5% SDS, add proteinase K (Sigma P-0390) to a final concentration of 0.5 mg/ml.

(b) RACE dotblot

Second round RACE products were analysed in essentially the same way as the tail dot blot above. 3µl of second round products were denatured in 100µl of 0.4M NaOH for 30 minutes at 37°C. The dot blot was performed as outlined in 4-6 above.

(c) RNA dotblot

20µg of total RNA was resuspend in 40ul DEPC water. The following were added:

80µl formamide

28µl formaldehyde

176µl 1M NaP buffer

Samples were left at 55°C for 15 minutes and then applied to the dot blot as above except for using Hybond N rather than N+.

6.2.4 Southern blot analysis

Southern blots were performed according to the following 'dry blot' protocol.

1. Following electrophoresis, the gel was photographed and shaken for 45 minutes in 0.4M NaOH.
2. Hybond N^+ (Amersham) was pre-wet in distilled water for 5 minutes and then transferred to 0.5M NaP buffer.
3. The transfer sandwich was set up as follows:
   a. Gel was inverted on a glass plate.
   b. Hybond N+ was laid on top, making sure all air bubbles were removed.
   c. 3 sheets of Whatman 3MM paper (pre-wet in 0.5M NaP buffer) were laid on top.
   d. 3 sheets Whatman 3MM paper (dry) was laid on top.
   e. A stack of hand towels was added on top.
4. Next morning the sandwich was disassembled and the filter rinsed briefly in 50 mM NaP buffer followed by UV cross linking the DNA to the filter (1 auto cross link cycle in Stratalinker).

Southern hybridisation of genomic DNA included a 20 minute treatment of the agarose gel in 0.25M HCl prior to the alkali treatment in step 1.

6.2.5 Northern blot analysis

The JST Northern in Figure 2.5 was run by Julie Moss.

Solutions:

10X MOPS Buffer

(store 4°C in dark)

400mM	1 M MOPS*
100mM	Sodium Acetate (pH 5.5)
10mM	EDTA (pH 8.0)

*1M MOPS (acid) pH to 7 with 2M NaOH, filter sterilise and store 4°C in dark

Loading Buffer (10mls)

5ml	formamide
1.58ml	formaldehyde
1ml	10 X MOPS buffer
2.42ml	water

1. RNA was separated on a 1% denaturing agarose gel containing 0.37M formaldehyde.
2. Loading buffer was added to 10-20µg RNA samples (3 parts buffer: 1 part RNA).
3. Samples were incubated at 65°C for 15mins then 1/100th of a volume of 10mg/ml ethidium bromide was added.
4. The gel was electrophoresed in 1X MOPS.
5. Once run, the gel was incubated with shaking under the following conditions:

1-5hrs	water (water changed a couple of times)
45 mins	50 mM NaOH, 10 mM NaCl
30 mins	0.5M NaP buffer

6. RNA was blotted onto Hybond N

RNA was transferred to Hybond N pre-soaked in 0.5M NaP buffer in a similar 'sandwich' as with the Southern hybridisation above. After transfer the filter was baked at 80°C for 2hrs.

RNA molecular weight standards were usually separated on the Northern gel, failing this, transcript size was estimated from the size of 18SrRNA (6333 nucleotides) and 28SrRNA (2366 nucleotides).

6.2.6 Preparation of radiolabelled probes

1. The DNA template was released by digestion of plasmid DNA with an appropriate restriction enzyme. After separation through a low melting point agarose gel the probe template was excised and the concentration was adjusted to 5 ng /µl with TE.
2. 5 µl of melted template (25 ng) was added to 5 µl of water and denatured for 5 minutes at 95-100°C. The DNA sample was collected at the bottom of the eppendorf tube by a brief pulse in a microfuge. 5 µl of DNA labelling mix, 5 µl of ^{32}P dCTP (10 µCi/ml;3000Ci/mmol, Amersham) and 1 µl of Klenow (2 units/µl) were added and the reaction incubated at 37 °C for at least 30 minutes.
3. Unincorporated nucleotides were removed by spinning the reaction through a G-50 sephadex column. Probes were stored for up to 1 week at -20 °C.

6.2.7 Hybridisation conditions

RNA and DNA blots were hybridised using a technique based on that of (Church and Gilbert, 1984).

1. Hybridisations were performed in rotating glass cylinders (Techne oven). 10 ml of hybridisation buffer was routinely used per filter in each glass cylinder.
2. Pre-hybridisation and hybridisation were carried out at 65 °C for DNA blots and 60 °C for RNA blots.
3. Filters were pre-hybridised for 10 minutes during which time radiolabelled double stranded DNA probes were denatured at 95-100 °C (5-10 minutes). The

probe was placed on ice for 2 minutes before being added to fresh buffer at a concentration of approximately 10^6 cpm/ml. Hybridisation proceeded overnight.

4. The first post-hybridisation wash was performed in the glass cylinder for 15 minutes at the same temperature as hybridisation. Two further 15 minute washes were performed in a Tupperware box on a shaker at the same temperature as before.
5. Filters were placed in Saran wrap and exposed to X-ray film between intensifying screens at -70 °C.
6. To re-hybridise filters, the bound probe was stripped by placing the filter in boiling 0.1% SDS and then left to cool.

DNA hybridisation buffer	**RNA hybridisation buffer**
0.05 g Marvel milk powder	0.2 g bovine serum albumin (Sigma)
10 ml 1M NaP buffer	7 ml 1M NaP buffer
7 mls 20 % SDS	6 ml formamide
3 mls water	7 ml 20 % SDS
DNA wash buffer	**RNA wash buffer**
30 mM NaP buffer	150 mM NaP buffer
0.1 % SDS	0.1 % SDS

6.2.8 RNAse Protection Assay

Based on the protocol of King and Melton, 1987).

Probe Templates

Probe templates were made by restriction digest of pBluescript plasmid containing RACE cloned products. In each case, the T7 polymerase site 3' of the sequence of interest was retained. The digests were then run on a 1.2% agarose gel, and the fragments representing the probe templates excised. The DNA was extracted from the agarose by spinning the gel slice through synthetic wool at 40°C and then ethanol precipitated and resuspended in DEPC treated water at approximately 250ng/ml.

Probe synthesis

The following components were mixed together at room temperature:

template	500ng (approximately 2μl)
200mM DTT	0.75μl
2mg/ml BSA	0.75μl
3.3mM ATP/UTP/GTP	2.25μl
RNase inhibitor	0.5μl
[α32P] CTP	6.25μl
10X transcription buffer	1.5μl
polymerase (usually T7)	1μl

For GAPDH loading control probe, 0.25μl of [α32P] CTP was used, together with 6μl of 1mM unlabelled CTP.

The reaction mixtures were then incubated at 37°C (40°C for GAPDH) for 1hr.

1ml of RNase free DNase was then added to each reaction to remove the probe template, and incubation at 37°C continued for 10 minutes. A further 1μl of DNase was added, followed by a further 10 minute incubation to ensure maximum degradation of the probe template.

The volume of each probe was made up to 100μl with sterile water, and each probe extracted once with phenol/ chloroform and spun through a sephadex G-50 column to remove unincorporated nucleotides.

Gel Purification of Probes

12μl of each probe was mixed with 8μl of loading dye, denatured briefly, and run on a 6% polyacrylamide sequencing gel for 1hr at 60W.

The wet gel was wrapped in Saran wrap and orientation marks made. A piece of X-ray film was then placed on the gel, and the same orientation marks made on the film. After developing the film, the two sets of marks were aligned, and the area of the gel containing the probe band carefully excised and removed to a microfuge tube. A further piece of X-ray film was then placed on the gel and developed to ensure that the

correct region of the gel had been removed. 100µl of probe elution buffer (0.5M ammonium acetate; 1mM EDTA; 0.2% SDS) was then added to each gel slice, and the gel crushed with a disposable spatula. The probes were then shaken at 37°C for 2hrs, at the end of which time the gel was spun to the bottom of the tube and the supernatant, containing the probe, removed. To ensure maximum yield of probe from the gel, a further 50µl of elution buffer was added to the gel, vortexed briefly, and the supernatant pooled with the previous 100µl. 1µl of each probe was then added to scintillation fluid and the incorporation of each probe measured.

Hybridisation

10µg samples of target RNAs were precipitated, spun down and resuspended in 2µl DEPC water. A volume of eluted probe containing 350 000cpm was added to each sample (50 000cpm for GAPDH loading control) and the volume made up to 100µl with DEPC water. These mixtures were then precipitated at -70°C using 250µl of ethanol, with 1µl of glycogen as carrier. The precipitated RNA and probe mixtures were then spun down, air dried and resuspended in 3µl of DEPC water, 3µl of 10X hybridisation buffer and 24µl of deionised formamide. The samples were then denatured at 85°C for 15 minutes and hybridised overnight at 55°C.

RNase Digestion

Digestion buffer was made up as follows:

2M Tris pH7.5	30µl
500mM EDTA	60µl
5M NaCl	360µl
dH_2O	5.55µl
25mg/ml RNase A	9.6µl
RNase T1	2.1µl

350µl of digestion buffer was added to each sample. The samples were then put at 30°C for 30 minutes to digest unhybridized RNA.

20µl of 10% SDS and 5µl of 10mg/ml proteinase K were then added to stop the digestion, and the samples incubated at 30°C for a further 10 minutes.

The samples were then extracted once with phenol/chloroform and precipitated at -70°C with 1ml of ethanol using 5µg of tRNA as carrier.

The precipitated RNAs were then spun down, air dried and resuspended in 1ml of DEPC water and 4µl of loading dye, denatured briefly at 95°C and run on a 6% polyacrylamide gel.

6.2.9 DNA sequencing

(a) Dideoxy-termination method

Double-stranded sequencing of plasmids was carried out using the Sequenase Version 2.0 DNA Sequencing Kit [USB-Amersham] and ^{35}S-dATP as label according to the recommendations of the manufacturer with the following modifications:

1. Termination mixes were pre-diluted with DMSO to a final concentration of 10% DMSO. 2.5 µl aliquots were transferred to tubes and kept at room temperature until the labelling reaction, whereupon they were transferred to 37°C.
2. 1-5 µg of plasmid was added to 1.8 pmol of the sequencing primer (17-mer) in a volume of 8.75 µl and denatured for 5 minutes.
3. After chilling samples were chilled on ice for 3 minutes. 1.25 µl of DMSO was added to samples.
4. 2.5 µl of Sequenase 5x reaction buffer was added and the labelling and termination reactions were performed according to the manufacturers recommendations.

(b) Direct sequencing

Preparation of Streptavidin coated beads

20 µl (200 µg) of beads were prepared for each template (Dynabeads M-280). Supernatent was removed from the beads with the aid of a specially designed magnetic tube holder (Dynal).

2X B&W buffer was made (10mM Tris-HCl pH 7.5, 1.0 mM EDTA, 2.0 M NaCl)

1. Beads were resuspended in 20 µl of 1X B&W buffer and mixed gently.
2. Supernatant was removed with the beads on the magnet and they were resuspended in 40 µl of 2X B&W buffer (Beads were now at 5µg/µl).

Immobilisation of PCR products

1. Buffer was removed from beads on magnet and 40 µl of washed beads were added to 40 µl of diluted biotinylated PCR products (in general 5µl of second round PCR products from both RACE and LM-PCR reactions were adequate for sequencing, these were diluted in 35µl of H_20).
2. Samples were incubated for 15 minutes at room temperature with frequent mixing to keep beads suspended.

Denaturation of DNA and removal of unbound strand

1. Supernatant was removed with beads on magnet which were then washed with 40 µl of 1X B&W buffer (samples could be stored at 4°C at this stage for several weeks).
2. Supernatant was removed with beads on magnet and resuspended in 8 µl of fresh 0.1M NaOH and incubated at room temperature for 10 minutes.
3. Supernatant was removed from beads on magnet and the beads were washed twice with 50 µl 0.1M NaOH and once with 40 µl 1X B&W buffer, followed with a final wash with 50 µl TE.
4. TE was removed with the beads on the magnet and they were resuspended in 12.5 µl H_2O.

End labelling sequence primer

The following were made up to a total volume of 20 µl:

10 pmol sequencing primer

2 µl 10X polynucleotide kinase buffer (Boehringer-Mannheim)

50 µCi [γ-^{32}P] ATP (3000-6000 Ci/mM)

10 units polynucleotide kinase (Boehringer-Mannheim)

Samples were incubated at 37°C for 10 mins, heat inactivated at 95°C for 2 mins and then stored at -20°C until ready to use

Sequencing reactions

An Amersham ThermoSequenase Kit (US 78500) was used for sequencing.

1 4µl of each termination mix (ddATP, ddCTP, ddGTP, ddGTP) was added to a strip of 0.2 µl PCR tubes

2 The reaction mix was made by adding the following to 12.5 µl of beads:

2 µl ThermoSequenase reaction buffer

1 µl [^{32}P] - end labelled primer (0.5 pmole)

2 µl ThermoSequenase

3 4 µl of reaction mix was added to each tube containing the termination mix and reactions were overlaid with 20 µl of mineral oil

40 cycles of sequencing were performed using the following conditions

95°C	30 sec
60°C	30 sec
72°C	1 min

4 µl of stop solution was added when cycling was complete and the samples were heated to 80°C for 3 minutes and load on sequencing gel

6.2.10 Polymerase chain reaction (PCR)

(a) Screening by colony PCR

Bacterial colonies were screened for the presence of inserts by colony PCR. PCR primers which flanked the insertion site in pBluescript were used for amplification (normally T3 and T7). If possible one of these primers was replaced with an insert specific primer. A toothpick was used to transfer a small amount of each bacterial colony onto a fresh gridded bacterial plate and into a tube containing the following PCR reaction:

2µl	PCR Buffer (10X)
2µl	$MgCl_2$ (25mM)
0.4µl	dNTPs (10mM)
1µl	Primer 1 (100ng)
1µl	Primer 2 (100ng)
0.2µl	Taq polymerase
13.4µl	water

The following cycles were performed:

1 cycle	94°C	5min
25 cycles	94°C	30sec
	55°C	30sec
	72°C	45sec

PCR reactions were separated on an agarose gel and positive colonies were picked from the gridded master plate and grown for plasmid preparation.

(b) Ligation Mediated PCR

The following protocol is based on a method of (Garrity and Wold, 1992). Solutions were made on ice.

5 X 1st strand buffer

200 mM NaCl

50 mM Tris-HCl (pH 8.9)

25 mM $MgSO_4$

0.05 % gelatin

(store -20oC)

Vent dilution solution

2.2 µl Tris-HCl (pH 7.5)

0.35 µl 1M $MgCl_2$

1 µl 1M DTT

0.25 µl 10 mg/ml DNase-free BSA

16.2 µl H_2O

Precipitation solution

8.4 µl 3M Na Acetate (pH 7)

1 µl 10 mg/ml yeast tRNA

Amplification Mix

20 µl 5 X amplification buffer

1 µl (10 pmol) Linker Primer

1 µl (10 pmol) CD4 Primer *

0.8 µl 25 mM dNTPs

7.2 µl H_2O

1st strand synthesis mix

6 µl 5 X 1st strand buffer

0.3 µl primer 1 (0.3 pmol) [-40 primer]

0.25 µl 25mM dNTPs

18.2 µl H_2O

0.25 µl Vent DNA polymerase (0.5 u)

Ligation solution

0.25 µl 1M $MgCl_2$

0.5 µl 1M DTT

0.75 µl rATP

0.13 µl DNase-free BSA

17.37 µl H_2O (chill, then add)

5 µl (100 pmol) unidirectional linker

1 µl (3u) T4 DNA ligase

5 X Amplification Buffer

200 mM NaCl

100 mM Tris-HCl (pH 8.9)

25 mM $MgSO_4$

0.05 % gelatin

0.05 % Triton-X100

Diluted Vent Polymerase

0.6 µl 5 X amplification buffer

1.9 µl H_2O

0.5 µl (1u) Vent DNA polymerase

* 1st round PCR use CD4 Primer 1, 2nd round PCR use CD4 Primer 5

Preparation of unidirectional linker.

1. Linker oligonucleotides were denatured for 5 min at 95°C. (20mM of each oligo in 250mM Tris pH7.7).
2. After transfer to a 70°C heat block they were gradually cooled to room temperature over a 1 hour period to allow annealing and then left for a further hour. To complete the annealing the oligonucleotides were left overnight at 4°C.
3. Aliquots of linker were stored at -20°C. In future operations linker was always thawed and handled on ice to prevent denaturation.

Synthesis of first strand and ligation of unidirectional linker.

1. 20 µg of genomic DNA was digested overnight with an appropriate enzyme in a volume of 60 µl. The reaction was phenol/chloroform extracted and precipitated by adding 6 µl of 3M sodium acetate and 120 µl ethanol and incubated at -20°C for 30 minutes. The precipitated DNA was pelleted, washed in 70 % ethanol and resuspended in 8 µl of TE. The concentration of DNA was estimated by taking a OD_{260nm} reading. The DNA was diluted to 0.4 µg/µl.
2. 5 µl of cleaved genomic DNA (2 µg) was chilled on ice in a PCR tube. 25 µl of ice-cold 1st strand synthesis mix was added and the tube was placed into a thermocycling block. The mixture was denatured for 5 mins at 95°C, annealed for 30 mins at 60°C and finally extended for 10 mins at 76°C.
3. The mix was cooled on ice and briefly pulsed in a microfuge before addition of 20 µl of ice-cold Vent dilution solution followed by 25 µl of ice-cold ligation solution. The mixture was then incubated at 17 °C for 12-16 hrs.
4. 9.4 µl of ice-cold precipitation solution was added to the sample and 220 µl of ice-cold 100% ethanol. After mixing by inversion the DNA was precipitated at -20°C for at least 2 hours.

PCR amplification.

1. The precipitated DNA was pelleted by spinning for 15 minutes, washed in 70 % ethanol, air-dried, resuspended in 70 µl of H_2O and stored on ice.
2. 30 µl of ice-cold amplification mix and 3 µl of diluted Vent polymerase was added and the reaction was overlain with 90 µl of mineral oil. PCR was performed in an Omnigene thermocycler as follows:

Cycle 1		
Denature	95°C	4 mins
Anneal	60°C	2 mins
Extend	76°C	3 mins
Cycles 2-17		
Denature	95°C	1 min
Anneal	60°C	2 mins
Extend	76°C	3 mins (+5 secs)*
Cycle 18		
Denature	95°C	1 min
Anneal	60°C	2 mins
Extend	76°C	10 mins

*increased by 5 second increments at each cycle

3. A second round of PCR was performed using a nested vector primer and biotinylated adapter primer. To 5 µl of 1st round products, 65 µl of water, 30 µl of amplification mix and 3 µl of diluted Vent polymerase was added. The reaction was overlain with 90 µl of mineral oil and PCR was performed using the same conditions as above.
4. Second round products were purified using a prep-a-gene kit (BioRad) according to the manufacturers instructions and sequenced directly.

Cloning of products

Second round products were cloned for line ET105, as a probe was required to identify a genotyping RFLP. Products were digested with XbaI, which cuts in the unidirectional adapter, and cloned into pBluescript digested with XbaI and EcoRV (the latter enzyme generates a blunt ended fragment which is compatible with the blunt end of the PCR products).

(c) Reverse Transcriptase PCR

1-3μg of RNA was transcribed in the same buffer used to amplify cDNAs.

Reverse transcription reaction

The following components were mixed in a 1.5ml eppendorf tube:

10μl RNA

3μl PCR buffer (10X)

2.5μl 10mM dNTPs

6μl 25mM $MgCl_2$

1μl random primers (1.8mg/ml)

0.5μl MMLV reverse transcriptase (200u/μl)

16μl H_2O

Samples were left at 25^0C for 10 minutes, followed by 1 hour at 42^0C.

cDNAs were denatured at 95^0C and placed on ice.

PCR

The following components were mixed in a 0.5ml PCR tube:

6μl cDNA

1.5μl PCR buffer (10X)

0.2μl Taq polymerase

1μl primer 1 (20μM)

1μl primer 1 (20μM)

10.3μl H_2O

Products were amplified for 30 cycles using the following conditions:

95^0C. 30 sec

60^0C. 45 sec

72^0C. 60sec

6.2.11 Rapid amplification of cDNA ends (5'RACE)

The gene disrupted in JST197 was identified by Anne Camus, a visiting scientist to the laboratory from the Pasteur Institute, Paris.

The following protocol has been streamlined and contains a number of modifications to the original method described by (Frohman et al., 1988). The most important of these are 1) alkaline hydrolysis of the RNA (required for efficient T-tailing with terminal deoxytransferase), 2) synthesis of second strand cDNA with Klenow instead of Taq polymerase and 3) the use of microdialysis filters, which serves a dual role to remove primers/buffers between steps and to size-select informative cDNA fragments above 300 bp in length. Primer numbers refer to those on the RACE schematic in Figure 2.6.

1st strand cDNA Synthesis

The following were mixed on ice:

1 µl 10ng /µl primer 1

1 µl RNA (5-10 µg)

Sterile distilled H_2O to 11µl

After heating to 70^oC for 5 mins samples were chilled on ice and collected at the bottom of the sample tube by a brief pulse in a microcentrifuge.

The following were added on ice:

4µl 5x 1st strand buffer (supplied with Superscript)

2µl 0.1M DTT

1µl 10mM dNTP

After 2 minutes at 37°C 1µl of Superscript II was added and the samples incubated at 37°C for 1 hr.

2.2µl of 1M NaOH was added to each sample followed by incubation at 65°C for 20 minutes. Samples were neutralised by adding 2.2µl of 1M HCl .

Microdialysis

The 1st strand reaction was loaded onto a 0.025µm microdialysis filter (Millipore) floating in a petri dish (~50ml) of T. E. and left for 4hrs to dialyse. Samples were then removed in a volume of 20µl H_20.

Poly A addition

To 20µl of 1st strand reaction, 6µl 5X TdT buffer (supplied with TdT) and 2µl of 2mM dATP were added and samples left at 37°C for 2 minutes. 2µl of TdT (500u/33.7ml) were added and reactions incubated for 5mins at 37°C followed by 2mins at 70°C. Samples were spun briefly in microfuge.

2nd Strand Synthesis

To 15µl of tailed cDNA the following were added:

2µl 10x restriction buffer M (Boehringer)

1µl 10mM dNTPs

1µl primer 2 (10ng/ml)

1µl Klenow (2U/ml)

A 30 minute incubation at room temperature was followed by 30 mins at 37°C and 5 mins at 70°C.

Microdialysis

cDNAs were microdialysed on 0.1 µm microdialysis filters as above and samples were collected in 37µl of H_20.

1st Round PCR amplification

37µl cDNA

5µl 10xPCR buffer (500mM KCl/100mM Tris, pH8.3)

4µl 25mM $MgCl_2$

1µl 10 mM dNTPs

1µl Primer 3 (100ng)

1µl Primer 4 (100ng)

1µl Amplitaq (Perkin Elmer)

Samples were overlaid with 50µl of mineral oil and amplified using the following conditions,

30 cycles of: 94°C 1min 30sec

60°C 1min 30sec

72°C 3min

Size selection of 1st Round PCR products

PCR products were dialysed on 0.1 µm filters as above and collected in 40 µl H_20.

2nd Round PCR amplification

The following reactions were set up. If 2nd round products were to be sequenced directly then biotinylated primer 6 was used, if samples were to be cloned then primer 4 was used.

5µl dialysed 1st round products

5µl 10x PCR buffer (500mM KCl/100mM Tris (pH8.3)

4µl 25mM $MgCl_2$

1µl 10mM dNTP's

1µl Primer 5 (100ng)

1µl Primer 6 (100ng) or Primer 4 (100ng)

1µl Taq polymerase

32µl dH_2O

Samples were amplified using the same conditions as in the first round of PCR.

Size Selection 2nd round PCR Products

Selection was performed as for 1st round products.

Products were either cloned (see below) or directly sequenced (see Section 6.2.9b).

Cloning of Amplified cDNA

cDNAs amplified from JST lines were digested with XbaI and Asp718 and phenol/chloroform extracted. Samples were precipitated with ammonium acetate (0.2M final concentration) and 2.5 volumes of ethanol. cDNA fragments were cloned into KpnI/XbaI digested Bluescript II SK+

6.2.12 λ Zap II cDNA Library Screening

Bacterial strains

The various *E. coli* strains used for plating phage cDNA libraries and conditions for their propagation are summarised in the table below.

Bacterial strains	Agar plate for bacterial streak	Medium for plating cultures
XL1 Blue F'	LB-tet[a]	LB + 0.2% maltose + 10mM $MgSO_4$
XL1 Blue MRF'	LB-tet[a]	as above
SOLR	LB-kan[b]	no supplements

XL1 Blue F' (*sup*E44 *hsd*R17 *rec*A1 *end*A1 *gyr*A46 *thi rel*A1 *lac*$^-$ [F' *pro*AB *lac*I^q Z ΔM15 Tn10 (Tet^r)] was used as a host strain for plating the D3 ES cell cDNA library. Plating cultures were always started from a fresh LB plate (supplemented with the appropriate antibiotic) of cells using a single colony and grown overnight with vigorous shaking in 50 ml LB (with or without supplements as appropriate). The next morning, the cells were spun at 3560 rpm for 15 minutes in a Beckman-GPR benchtop centrifuge and gently resuspended in ~20 ml 10 mM $MgSO_4$. For in vivo excision of insert-containing Bluescript plasmid from λZAP II, the strains XL1 Blue MRF' and SOLR were used.

Titration of lambda phage

Titers of phage libraries and clones were determined by serial dilution as follows:

1. Make appropriate dilution series in SM.

2. Add 100 µl of the diluted phage to 100 µl plating culture (section 4.3.1) and allow the phage to adsorb for 20 minutes at 37°C.
3. Add 3 ml molten (45°C) LB top agarose, invert 1-2 times to mix and immediately pour onto pre-warmed (to 42°C) LB plates.
4. Allow top agarose to set and incubate inverted at 37°C overnight.
5. Count the number of plaques and calculate the titre taking into account the dilution factor. For example, if there are 20 plaques on a plate made from a 1/10,000 dilution, that means there are 20 x 10,000 x 10 = 2 x 10^6 plaque forming units (pfu) per ml of the original phage stock. (The last 10 of the calculation converts the 100 µl of the diluted phage which was plated to a per ml basis.)

Plating and transferring phage libraries

Bacteriophage cDNA libraries were plated at high density onto agar plates and then transferred to nylon filters according to the following protocol:

1. Pour LB plates several days in advance and allow them to dry at room temperature prior to plating.
2. Melt top agarose, allow to cool to 45°C and supplement with maltose to 0.2% and $MgSO_4$ to 10 mM.
3. Mix phage with plating cultures according to Table 2.

TABLE 2 - Proportion of phage to plating cells and volume of top agarose required for different plate sizes

	Plate size		
	10 cm	15 cm	24.5 x 24.5 cm
bacteria/ml	0.2	0.5	2
phage/pfu	5,000	20,000-30,000	150,000
top agarose/ml	3	7	30

4. Incubate at 37°C for 20 minutes.

5. Add top agarose, invert to mix and quickly pour onto plate. Disperse bacteria and agarose by tilting plates back and forth.

6. Incubate plates at 37°C until plaques cover the plate but not confluent (6 - 12 hours).

7. Leave plates at 4°C for 1 hour at least to allow agarose to harden before performing plaque lifts.

8. Leave nylon filters (Hybond N from Amersham) on plates for 3 - 5 minutes to allow phage particles to transfer. Make orientation marks by stabbing a 23-G needle through the filter into the agar at several asymmetric points. For second, duplicate filter (replica filter), allow 5 - 7 minutes for transfer.

9. Dry filters on the benchtop for at least 10 minutes. The drying process binds the plaques to the filter.

10. Place filters on a sheet of Whatman 3 MM paper in a tray saturated with 0.2M NaOH, 1.5M NaCl. The 3MM paper should be wet enough to allow immediate saturation of filters but not so wet that the solution pools on the surface. Leave for 5 minutes.

11. Transfer filters to 3MM paper saturated with 0.4M Tris pH7.6, 2X SSC for 5 minutes and then to 3MM paper saturated with 2X SSC for 2 minutes.

12. Allow filters to dry briefly on bench and then UV irradiate (one auto cross link cycle in a Stratalinker). Prehybridize immediately or, if storing filters, dry at 80°C.

Screening of ES cell cDNA libraries

A random-primed D3 ES cell cDNA library was cloned in the vector λZAP II (Hitoshi Niwa, unpublished). Approximately 1 x 10^6 recombinants of the library were screened by plating 2 x 10^5 plaques per 245 x 245 mm plate (Nunc) as described above.

The filters (no more than four 245 x 245 mm filters per stack) were transferred to hybridisation bottles (Hybaid) and prehybridised for at least 2-4 hours at 65°C in 25ml of pre-hybridisation buffer (see below). Hybridisation was performed overnight at 65°C with not more than 2 x 10^6 cpm per ml of hybridisation buffer (see below) of a probe labelled to a specific activity in excess of 1 x 10^9 dpm/μg. On the following day, filters were washed to a final stringency of 2X SSC, 0.1% SDS at 65°C. Exposure at -80°C in the presence of two intensifying screens was usually for 5 days.

Positive plaques (those demonstrating radioactivity on replica filters) were each cored with the wide bore end of a glass pasteur pipette and transferred to 0.5 ml SM overnight at 4°C. A secondary screen was then performed on 2 or 3 plates (<100 plaques/plate) of each primary plaque. A single, well isolated hybridising plaque for each clone was picked and the process repeated until pure clones were obtained.

For 100mls of pre-hybridisation buffer		For 100mls of hybridisation buffer	
0.5 ml	100X Denhardts	0.5 ml	100X Denhardts
30 ml	20XSSC	30 ml	20XSSC
5 ml	10% Sarkosyl	5 ml	10% Sarkosyl
64.5 ml	water	64.5 ml	water
1 ml	herring sperm (10mg/ml)	10g	Dextran Sulphate
		1 ml	herring sperm

***In vivo* excision of Bluescript plasmid from λZAPII clones**

The Bluescript SK- plasmid portion carrying the cloned insert of λZAPII clones was excised using the Ex Assist helper phage (Stratagene). The helper phage gene II protein recognises two signals in the λZAPII vector arms, initiator (I) and terminator (T), where f1(+)- strand DNA synthesis will initiate and terminate, respectively. The newly synthesised DNA strand is circularised and can be recovered by infecting an F' strain and growing in the presence of Ampicillin. In addition, the helper phage carries

an amber mutation that prevents its replication in a non suppressing *E.coli* strain such as SOLR. This allows only the excised plasmid (actually phagemid) to replicate in the host. The ExAssist helper phage was used for *in vivo* excision as follows:

1. Prepare plating cultures of XL-1 Blue MRF' and SOLR as described in Table 1. Dilute cells to an OD_{600} of 1.0 in 10mM $MgSO_4$.
2. In a 50 ml conical tube combine the following:
   200µl of XL1 Blue MRF'
   250µl of λZAPII phage supernatent (from cored plaque of interest)
   1µl of ExAssist helper phage (~10^{10} pfu/ml)
3. Incubate at 37°C for 15 minutes
4. Add 3 ml LB and incubate at 37°C for 2-2.5 hours with shaking
5. Heat at 70°C for 15 mins and pellet cells by centrifugation at 4000xg for 15 mins. Save supernatent (may be stored at 4°C for up to 2 months).
6. To recover the excised phagemids, add 200µl of a plating culture of SOLR cells to 1µl, 10 and 100 of the supernatent in separate tubes.
7. Incubate at 37°C for 15 minutes then plate 200µl from each tube on LB/Amp plates. Incubate plates at 37°C overnight.

The excised plasmid possesses a pBR322 ori and is thus maintained at a low copy number per cell. A starter culture was prepared by inoculating a single bacterial colony into 10 ml LB/Amp and shaken vigorously at 37°C. At the end of the day, 2.5 ml of this culture was inoculated into 250 ml LB/Amp and, following vigorous shaking overnight at 37°C, 200 ml was used for plasmid preparation (see chapter 2). Under these conditions, the yield of plasmid DNA was 50 - 100 µg.

6.3 ES Cell Culture

ES cells were maintained at 7.5 % CO_2, at 37 °C in a humidified incubator (Heraeus). Methods for routine culture of CGR-8 ES cells are based on those described in (Smith, 1991). All ES cell manipulations were carried out in a laminar flow sterile hood (ICN,

flow). All solutions were warmed in a 37°C water bath prior to use. Tissue culture grade flasks and dishes (Corning) were gelatin coated (0.1 % in PBS, Sigma) for at least 15 minutes and aspirated before use.

6.3.1 Passage and expansion of ES Cells

Careful maintenance of ES cells was crucial for successful germline transmission of gene trap cell lines. Medium was changed every day and cells were not allowed to grow past confluence (approximately 10^7 cells per 25cm^2 tissue culture flask).

1. ES cells were passaged once they had nearly reached confluence. Medium was aspirated off and cells were rinsed twice with 5 to 10 ml of PBS.
2. Enough trypsin solution was added to cover over the cells (1 ml / 25 cm^2) and the flask was incubated at 37°C for 1 to 2 minutes until a uniform suspension was formed. Flasks were tapped to ensure complete dislodging of cells from the surface.
3. 9 ml of ES cell medium was added to stop trypsinization.
4. Cells were counted and 10^6 (approximately 1/10 of a nearly confluent 25 cm^2 flask) were seeded in a freshly gelatinised flask containing 10 ml of medium supplemented with 100 Units/ml of DIA/LIF (see section 6.1.4).

6.3.3 Freezing ES Cells

1. Cells were harvested as in steps 1-3 above. After transfer to a 20 ml Universal tube they were pelleted by centrifugation (Denley BS400 Benchtop, 1200 rpm, 5 minutes, RT).
2. 5 % of DMSO (Analar, BDH) was added to the cell suspension and the cells again pelleted by centrifugation.
3. The cell pellet was resuspended in 0.5 ml of freezing medium (culture medium + 10 % DMSO) per 5 X 10^6 cells, quickly aliquotted into cryotubes (Nunc) and placed in a -80 °C freezer for 24 hours before being transferred to a liquid nitrogen cell bank for long term storage.

The cryoprotector dimethyl sulfoxide (DMSO) is toxic to cells and also an inducer of ES cell differentiation so exposure time was minimised before freezing.

6.3.4 Thawing ES Cells

1. Frozen vials were retrieved from storage and quickly thawed in a 37°C water bath. The contents were transferred into a 20 ml centrifuge tube containing 10 ml of pre-warmed medium.
2. Cells were pelleted at 1200 rpm for 5 minutes, the supernatant aspirated and the cells were gently resuspended in 10 ml of pre-warmed DIA/LIF conditioned medium.
3. Medium was changed after about 8 hrs growth to remove dead cells and any residual DMSO (an inducer of ES cell differentiation).

6.3.5 Electroporation

Gene trap vectors were introduced into ES cells by electroporation. The procedure involves exposure of a suspension of ES cells and vector DNA to a high voltage electric discharge. This leads to the formation of membrane pores that are large enough to allow macromolecules such as DNA to enter.

(a) Vector preparation

1. 150 µg of vector DNA was linearised with the appropriate restriction enzyme in a volume of 0.3 ml. (Plasmid DNA for electroporation was prepared by alkaline hydrolysis and banded on a caesium chloride gradient as described in 6.2.2b).
2. Digested DNA was precipitated in 2 volumes of absolute ethanol on ice for 5 minutes. After spinning in the microfuge the pellet was washed several times with 70% ethanol and as much of the ethanol as possible was drained off.
3. The remaining ethanol was allowed to evaporate in the hood by keeping the lid of the tube open (for approximately 1 hour). The DNA pellet was resuspended

in 100µl of sterile PBS and vortexed occasionally over a period of 4 hours to ensure that the DNA was completely dissolved.

(b) Electroporation

1. A total of 10^8 cells were required for each electroporation. 3 x 10^6 cells were seeded in a 80 cm^2 flask containing 30 ml of medium. Once the cells reached confluence two 175 cm^2 flasks were seeded with 5 x 10^6 cells each containing 50 ml of medium. The next day the flasks were topped up with 40 ml of fresh media. The morning of the electroporation the media was replaced with 50 ml of fresh media. Each confluent flask should yield about 8 x 10^7 cells.
2. Each 175 cm^2 flask was trypsinised using 5 ml trypsin as described above.
3. Once the cells were sufficiently disaggregated and freed from the flask the trypsin was inactivated with 15 ml media. Cells were pelleted at 1200 rpm for 3 minutes and resuspended in 10 ml of cool PBS which had been taken from the fridge 15 minutes prior to use.
4. The pooled cells were counted and after re-pelleting were resuspended at a concentration of 10^8 cells in 0.7 ml of cool PBS.
5. The cells were added to the eppendorf tube containing the 0.1 ml (150µg) of linearised vector and then immediately transferred to a 0.4 cm electroporation cuvette. The cuvette was pulsed in a Bio-Rad Gene Pulsar unit set at 3µF, 0.8kV which gave a time constant of 0.1 msec.
4. Twenty 8.5 cm^2 plates were gelatin coated whilst the cells were left to recover in the cuvette for 20 min.
5. Cells were diluted in 200 ml of medium (without selection) and 10 ml (5 x 10^6 cells) of the cell suspension was plated onto each 8.5 cm^2 plate and incubated over-night.
6. On the following day the media was replaced with medium containing 200 µg/ml G418. For the first five days the medium was changed daily until most of the

cells had died and G418-resistant colonies had started to appear. Then the medium was changed every other day.

(c) Picking G418 resistant colonies

1. After about 10-12 days colonies were about 1 mm in diameter and ready to be picked. Colonies were circled with a marker pen on the bottom of the dish.
2. Enough 24-well plates for the number of colonies to be picked were gelatin coated (one colony per well).
3. Each 8.5 cm^2 plate was rinsed with PBS and then 10 ml of PBS was added. Colonies were picked under a dissecting microscope on the bench using flame pulled glass pipettes which had been snapped to produce a tapered end of approximately 1-2 mm in diameter. Each colony was removed from the dish by gentle scraping and suction by mouth and transfer to a 24-well dish in a volume of about 100 µl. Once 24 colonies had been picked, the plate was transferred to the laminar flow hood and 100 µl of trypsin was added to each well and incubated at 37°C for 10 minutes. The dish was tapped several times to disperse the cells and then 2 ml of medium was added to each well.
4. To identify ßgal-positive colonies, 4/5 of a nearly confluent well were split into a new 24-well plate (Experimental) for staining with X-gal and the remaining 1/5 split into a second 24-well plate (Master) for maintenance and subsequent expansion of selected cell lines. Since the colonies grow at different rates, the cells are split in batches over a period of several days. Cells in the Experimental plates are stained with X-gal the following day (see below) and ßgal activity is scored after an overnight incubation. By this time the cells on the Master plate should be ready to expand for further analysis. (see Section 6.6.1 for X-gal protocol)

6.4 Blastocyst injections

Most of the injections of JST lines were performed by W.C. Skarnes.

The procedure used for blastocyst injection was essentially the same as that outlined in (Robertson, 1987). Most of the blastocysts used for the production of chimaeras were flushed from mice of the C57BL/6 inbred strain on the fourth day of pregnancy using PB1 medium supplemented with 10% FCS (see below). Occasionally, mice of the MF1 Swiss albino outbred strain were used. Unexpanded embryos were transferred to ES cell medium and placed in a humidified incubator at 37°C with 6% CO_2, to allow expansion to occur. When required for injection, blastocysts were transferred to small hanging drops of PB1 medium with FCS on a siliconised (Repelcote; BDH) coverslip suspended over a manipulation chamber (Custom-made by G. Robertson) flooded with liquid paraffin (Boots). Trypsinised cells, generally single cells or pairs, were placed in the same hanging drop, and the chamber refrigerated for at least 10 minutes. Injections were performed by means of a rounded holding pipette of internal diameter 20μm, and a heat-polished, flat-ended injection pipette of internal diameter 15μm, both constructed from glass capillary tubing (0.1mm I. D.; Drummond). For these, suitable needles were pulled using an electrode puller (Camden Instruments), and cut to size and heat-polished by means of a microforge (DeFonbrune).

Holding and injection pipettes were each attached to instrument holders (Leitz) and operated by Leitz manipulators. The suction of the holding and injection pipettes were controlled by specialised injectors (Narashige for the holding pipette, DeFonbrune for the injection pipette). An IMT2 image-corrected microscope (Olympus) was integrated into the micro manipulation assembly. Between 10 and 20 cells were injected into each blastocyst. Operated embryos were then allowed to recover for a few hours in ES cell medium in the incubator before being transferred to pseudopregnant recipients. The recipient mothers selected were of an F_1 stock from matings between C57BL/6 and CBA strains. F_1 females had been mated with

vasectomised DBA males 2.5 days previously. Unilateral transfers to the uterus were performed; each inoculum contained 6 to 12 embryos.

Contribution of ES cells to resulting pups was estimated by coat colour. When blastocysts from C57Bl/6 mice were used the appearance of sandy coloured, agouti or white hairs amongst the host black hairs indicated successful incorporation of the ES cells into the resulting individual. Germline transmission was ascertained by test breeding. Generation of agouti offspring indicated successful transmission through the germline of the ES cells, whereas black offspring were indicative of transmission from the host blastocyst.

PB1 solution (pH7.0-7.2):	822 mg	NaCl
	21 mg	KCl
	300 mg	Na_2HPO4
	20 mg	KH_2PO_4
	104 mg	Glucose
	4.5 mg	Sodium pyruvate
	6.2 mg	Penicillin
	1.4 mg	$CaCl_2$
	1 mg	$MgCl_2$
	1 ml	phenol red (1%)

Made up to 100 ml with UHP H_20

6.5 Sample collection

6.5.1 Embryo collection (8.5dpc-10.5dpc)

Females were mated overnight. Stages of embryos are represented as number of days post coitum (dpc), assuming fertilisation to have occurred in the middle of the dark period preceding the discovery of the vaginal copulation plug. In PB1 supplemented

with 10% FCS, the embryos were dissected free from the decidual tissue and Reichert's membrane was removed.

6.5.2 Caesarean sections (18.5dpc)

Caesarean sections were performed on 18.5dpc pregnant females essentially according to the method in (Nagy and Rossant, 1993). Following dissection of neonates from the uterus and removal of foetal membranes the umbilicus was cut. The neonates were then placed under a heat lamp on a damp tissue. They were prodded gently at intervals until they began to breath normally.

6.5.3 Lung inflation

Protocol provided by Gillian Morrison.

A lethal injection of Euthatol (0.1ml) was administered to an adult mouse. The chest cavity was opened and the heart removed to expose the lungs. A stretch of ligature was wrapped around the trachea and tied in a loose knot. A small incision was made in the trachea and a 21gauge needle attached to a syringe was inserted. The ligature was secured around the needle to keep it in place and 1ml of Bouin's fixative was injected into the lungs. The needle was quickly removed and the ligature was tightened around the trachea so the lungs remained inflated with fixative. The lungs were completely dissected from the chest cavity and fixed in Bouin's overnight.

6.6 Histology

6.6.1 Staining cells and embryos for βgal activity

Based on the protocol in Beddington *et al*, 1989

Solutions for X-gal staining:

X-gal (5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside)

Stock solution, 50 mg/ml, dissolved in dimethylformamide (Sigma). Store in the dark at -20°C

Phosphate Buffer (0.1M)

Made by adding 21 parts of 0.1M Na_2HPO_4 (Sigma) to 4 parts 0.1M NaH_2PO_4 (Sigma) to give a final pH 7.5.

Phosphate buffer stock (5X) store at room temp, do not autoclave

298.1 g Disodium hydrogen phosphate anhydrous

55.2 g Sodium dihydrogen phsphate.$1H_2O$

10.2 g Magnesium chloride.$6H_2O$

Bring to 5 litres in dH_2O

Fix store at 4°C for up to 1 month

Dilute stock 1 to 5 (200ml 5X stock to 1 litre final)

Add to 1 litre, 8 ml 25% gluteraldehyde

20 ml 0.25M EGTA

Wash store at 4°C.

Minus detergent: Dilute stock 1 to 5 (200ml 5X stock to 1 litre final)

Plus detergent: Dilute stock 1 to 5 (200ml 5X stock to 1 litre final)

10 ml 10 % deoxycholate stock solution

2 ml 10 % NP40

0.5 ml 0.05 % BSA

X-gal staining solution store in the dark at 4°C. Filter before use to remove crystals. To 50 ml of X-gal wash buffer add:

Potassium ferrocyanide (Sigma)	0.106g
Potassium ferricyanide (Sigma)	0.082g
X-gal (50 mg/ml in Dimethyl formamide)	1 ml

1. Cells and embryos were washed in PBS and fixed for 5 minutes (cells), 15 minutes (8.5dpc embryos) or 30 minutes (9.5dpc embryos) at room temperature.
2. Cells were washed 3 times for 5 minutes (wash buffer - detergent), embryos were washed 3 times 20 minutes (wash buffer + detergent)
3. Cells and embryos were left to stain in X-gal solution overnight in a humidified chamber at 37°C.
4. After staining, cells and embryos were rinsed in wash solution and stored in 4% paraformaldehyde as this treatment was found to intensify the stain.

6.6.2 Cryostat sections

(a) Treatment of slides

1. Glass slides (Chance Popper) were treated with 2% TESPA (3-aminopropyltritheoxysilane; Sigma, A3648) in acetone for 2 minutes.
2. The slides were then rinsed twice in acetone for 2 minutes and stored desiccated at 4°C for up to one month.

(b) Treatment of tissue

1. Freshly dissected embryos were embedded in OCT compound and frozen at -20C overnight.
2. The next day samples were transferred to -70°C.
3. Before sectioning, the frozen embryos were transferred to the cryostat and allowed to equilibrate to the cutting temperature for about 1 hour .
4. Embryos were sectioned at -14°C and sections were immediately transferred to TESPA coated slides.
5. Sections were then stained for βgal activity using the same conditions as for cells in Section 6.6.1.

6. After overnight staining sections were either counter stained with neural red (see below) or post-fixed for 1 hour and mounted under coverslips using Moviol (Hoechst).

(c) Neutral red counter staining

Neutral red was used to stain the nuclei of X-gal stained sections

1. After overnight staining in X-gal solution (Section 6.6.1) sections were washed three times in X-gal wash buffer and then stained for 30 minutes in Neutral red (1%, pH3.3).
2. Sections were dehydrated as follows:

3x 5 minutes	30% Ethanol
2x 5 minutes	70% Ethanol
5 minutes	95% Ethanol
5 minutes	100% Ethanol
5 minutes	propanol

3 After 5 minutes in histoclear sections were mounted under coverslips using DPX mountant (BDH, 36125)

6.6.3 Wax sections

(a) Fixing, dehydrating and embedding

Bouin's reagent (BDH) was used to fix samples which were going on to be analysed for histology and for elastin staining. Buffered formalin was used to fix samples which were going on to be analysed by immunohistochemistry. Freshly dissected tissues were fixed overnight and then processed for paraffin wax embedding in an automatic tissue processor (Shandon, Citadel 2000).

Tissues were placed through the following solutions as described below:

70%	ethanol	1 hour
80%	ethanol	1 hour
90%	ethanol	1 hour
100%	ethanol	1 hour
100%	ethanol	2 hours
100%	ethanol	2 hours
	histoclear	1 hour
	histoclear	1 hour
	histoclear	1 hour
	wax	3 hours
	wax	4 hours

Sections were embedded in wax using an embedder (Blockmaster III, Raymond Lamb). 7-10μm wax sections were deposited on polylysine coated microscope slides (BDH).

(b) Haematoxylin and Eosin staining

Haematoxylin (1:3 in water) was used to stain nuclei and eosin (made in 70% EtOH) to stain cytoplasm. Harris haematoxylin solution (Sigma) was diluted 1:3 in water. Eosin was made by mixing 3 parts 1% eosin to 1 part 1% ethanol and adding acetic acid to give 0.05% final concentration. Sections were treated in histoclear for 5 minutes, rehydrated through an ethanol series of 100%, 96%, 90% and 70% for 5 minutes each and then stained in haematoxylin for 30 seconds. Sections were rinsed briefly in tap water and stained in eosin for 90 seconds. After another brief wash in water, sections were dehydrated through the same ethanol series as before, but in the reverse order. Finally, sections were treated in histoclear for 5 minutes before mounting in DPX.

(c) Elastin and Van Geison stain

Miller's elastin and Van Geison's stain were from Raymond A. Lamb. The protocol was provided by Pauline Lamb. Sections were de-waxed in xylene for 5 minutes and then placed in 100% ethanol for 5 minutes prior to staining in Miller's elastin for 2 hours. After a 100% ethanol and a water rinse, sections were stained in Van Geison's for 5 minutes. Sections were dehydrated through an ethanol series (see above), cleared in xylene for 5 minutes and mounted under coverslips using DPX mountant. The sections stain elastin a blue/black colour, muscle is yellow and collagen red.

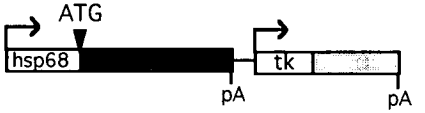
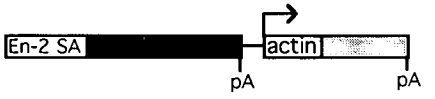
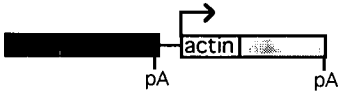
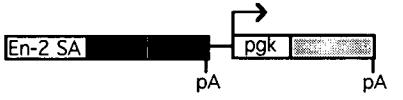
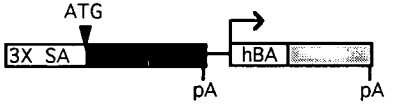
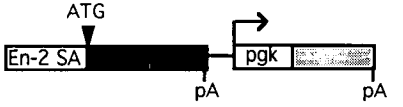
(d) Antibody staining

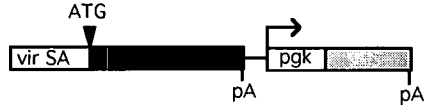
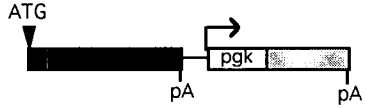
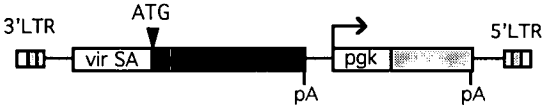
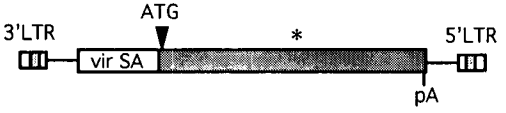
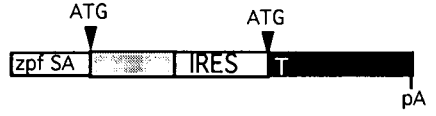
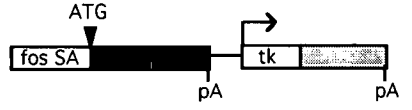
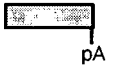
Immunohistochemistry was performed using the Vector Elite ABC system (Vector Labs). The protocol used was adapted from the manufacturers recommendations by Euan Slorach.

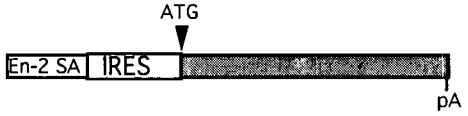
1. Formalin fixed sections were de-waxed in Xylene and rehydrated through an ethanol series of 100%, 90%, 70%, 50% and 30% for 5 minutes each. After 5 minutes in water, the endogenous peroxidase activity was blocked by treating with hydrogen peroxide (3% in methanol) for 10 mins.
2. Sections were rinsed gently with TBS (Tris Buffered Saline, 0.05M Tris HCl (pH 7.6), 0.15M NaCl), using a wide bore plastic pastette and washed for 5 minutes in TBS.
3. TBS was removed and blocking serum was added to sections for 20 minutes.
4. Serum was removed and, without rinsing, the primary mouse antibody in blocking serum solution was added to sections and left for 30 minutes at room temperature. (Monoclonal anti-α-smooth muscle actin, Sigma A-2547 was used at 1:400 dilution).
5. Antibody was removed and sections rinsed with TBS, followed by a 5 minute wash in TBS.

6. Biotinylated antibody (made up in TBS with blocking serum) was added for 30 minutes. During this incubation step the ABC solution was mixed and left to stand for 30 minutes.
7. Biotinylated antibody was removed and sections rinsed with TBS, followed by a 5 minute wash in TBS.
8. TBS was removed and the ABC solution was added to sections for 30 minutes
9. Sections were rinsed in TBS and then washed for 5 minutes in TBS
10. TBS was removed and peroxidase activity was visulaised using a DAB substrate kit (Vector labs).
11. Sections were counter-stained with haematoxylin and eosin as in section b above.

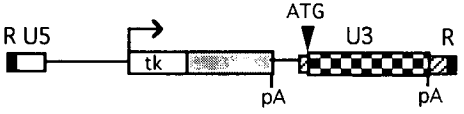
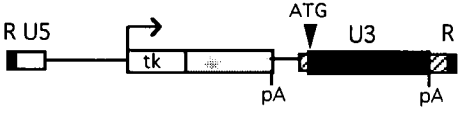
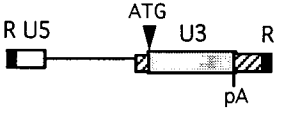
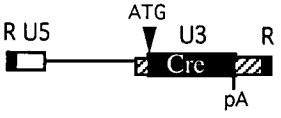
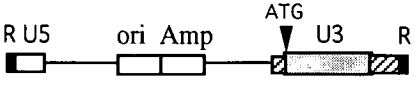
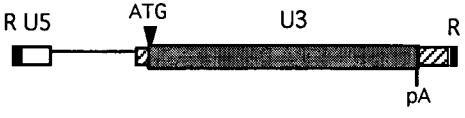
Appendix I

		Structure	% of neo^R colonies showing βgal activity	Phenotype Frequency	Comments
Enhancer trap					
1.	p3LSN (Gossler *et al*, 1989, Korn *et al*, 1992, Soininen *et al*, 1992, Neuhaus *et al*, 1994)		14 - 20 % [6/42, 13/59]	0/2	27/59 (46%) restricted expression at 9.5dpc. In the 2 cases analysed expression of endogenous genes was more widespread than reporter activity.
Electroporated Promoter traps					
2.	pGT4.5 (Gossler *et al*, 1989, Skarnes *et al*, 1992, Gasca *et al*, 1995)		2 % [10/600]	1/3	3/7 ubiquitous, 2/7 widespread, 2/7 restricted expression in embryos
3.	pGT4.5ΔSA (Gossler *et al*, 1989),		0.16 %	N.D.	12.5 fold fewer colonies express lacZ in the absence of a splice acceptor (ie compared to pGT4.5)
4.	pPT-1 (Wurst *et al*, 1995)		1 % [393/38 730]	N.D.	In 8.5dpc embryos 13% of lines show restricted patterns, 32% widespread expression and 55% show no expression
5.	pGTi (Schuster-Gossler *et al*, 1994)		6 % [80/1393]	N.D.	Between 8.5-12.5dpc 28 % of embryos show restricted expression patterns
6.	pT-1 ATG (Forrester *et al*, 1996)		N.D.	1/5	In an induction screen 0.6% lines respond to retinoic acid treatment *in vitro*, of these 95% show restricted expression in embryos between 8.5-11.5dpc
7.	pGTfu (Schuster-Gossler et al, 1994)		100 % [58/58]	N.D.	Between 8.5-12.5dpc 77 % of embryos show restricted expression patterns
8.	pGT1.8βgeo (Skarnes *et al*, 1995 V.Wilson & W.C. Skarnes, unpublished results)		60 % [312/520]	27% [4/15]	Broad range of βgal staining intensities and lower number of βgal negative colonies

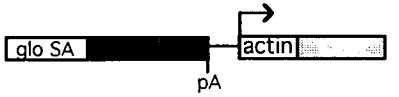
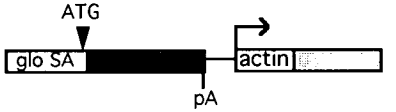
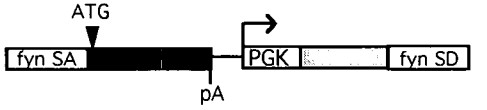
9.	pSAβgeo (Friedrich & Soriano, 1991)		95 % [19/20]	1/2	
10.	pGT1.8TM (Skarnes *et al*, 1995, Townley *et al*l, 1997, W.C. Skarnes unpublished)		20 % [72/360]	45 % [13/29]	
11.	pSAβgal (Friedrich & Soriano, 1991)		4.6 % [16/350]	1/3	
12.	pβgal (Friedrich & Soriano, 1991)		0.1 % [1/1300]	N.D.	
13.	ROSAβgal (Friedrich & Soriano, 1991)		11.6 % [31/268]	N.D.	
14.	ROSAβgeo (Friedrich & Soriano, 1991, Chen *et al*, 1994, Deng *et al*, 1995, Baker *et al*, 1997)		95.6 % [196/205]	37 % [7/19]	
15.	TV2 (Takeuchi *et al,* 1995)		93.2 % [41/44]	N.D.	
16.	pT10 (Rijkers & Ruther, 1996)		10 % [14/139]	N.D.	
17.	NASTI (Macleod *et al*, 1991)		N/A	N.D.	6/7 lines inserted close to or within a CpG island

18.	PHLEAL (Camus *et al*, 1996)		77 % [49/64]	38% [6/15]	71% of lines show similar reporter expression in heart and liver. Postulated preferred integration site. 3/3 integrations map to same region of same chromosome.
19.	pSAβgeo (Chowdhury *et al*, 1997)		30 %	N.D.	
20.	pSAIRESβgeo (Chowdhury *et al*, 1997)		75 %	N.D.	Three fold more G418 colonies are obtained when an IRES is added to pSAβgeo. No embryo expression patterns reported

Retroviral Promoter traps

21.	U3his (Von Melchner & Ruley, 1989)		N/A	1/1	
22.	U3LacZ (Reddy *et all*, 1991, Jönsson *et al*, 1996)		0.4%	N.D.	
23.	U3neo (Reddy *et al*, 1991)		N/A	1/3	
24.	U3Cre (Russ *et al*, 1996))		N/A	N.D.	
25.	U3neoSV1 (Hicks *et al*, 1997)		N/A	38% [6/16]	
26.	U3βgeo (Chen *et al*, 1996, Scherer *et al*, 1996)		65 %	N.D.	

Poly A trap vectors

27.	U1 (Niwa *et al*, 1993)		16.1 % [11/68]	N.D.	
28.	U2 (Niwa *et al*, 1993)		7.7 % [9/117]	N.D.	4/4 insertions into promoter regions 2/2 insertions asociated with >10kb genomic deletion 2-20 bp of vector (inc SA) deleted before insertion
29.	pPAT (Yoshida *et al*, 1995)		8 % [4/49]	0/1	3' RACE succesful in 25 % of G418 lines 4/6 lines analysed represent proper gene trap events which pick-up a pA signal

Key

	Lac Z
	neomycin
	promoter
	βgeo (fusion between *lacZ* and *neomycin phosphotransferase)*
TM	transmembrane domain from CD4
	histididinol dehydrogenase
T	nuclear transport signal
	phleomycin
hsp68	*mouse heat shock protein 68*
ATG	initiation methionine
tk	*thymidine kinase*
En-2 SA	*engrailed-2* splice acceptor
Cre	*Cre* recombinase
pgk	*phosphoglycerate kinase-1*
3X SA	splice acceptor of MoMuLV *env* gene allowing splicing products in all three reading frames
hBA	human *β-actin*
LTR	long terminal repeat
vir SA	adenovirus major late transcript splice acceptor
*	point mutation known to reduce neomycin activity (Skarnes *et al*, 1995)
zpf SA	*zinc finger protein 40*
IRES	internal ribosomal entry site from encephalomyocarditis virus
fos SA	*c-fos* gene splice acceptor
glo SA	rabbit *β-globin* splice acceptor
fyn SA	*fyn* splice acceptor
fyn SD	*fyn* splice donor

Appendix II

5'RACE sequence for JST lines

JST51 (LDL receptor)

```
  1 ccgtgtcagccgatgcattcctgactcctggagatgtgatgganaggtagactgtgaaaa 60
    R  V  S  R  C  I  P  D  S  W  R  C  D  G  ?  V  D  C  E  N

                            SA
                            ↓
 61 tgactcagacgaanaaggctgtcgtcccagg 91
    D  S  D  E  ?  G  C  R  P  R
```

JST62 (novel, myeloblast cDNA)

```
  1 accattctccagcagctggattcccaggcctgcacctcagaaaaccagaagtctgaggag 60
    T  I  L  Q  Q  L  D  S  Q  A  C  T  S  E  N  Q  K  S  E  E

                                                  ↓
 61 aatgaacagacagaggaggggaagccaagcgccattgaaggtcccaggtccc 112
    N  E  Q  T  E  E  G  K  P  S  A  I  E  G  P  R  S
```

JST71 (hsp 86)

```
  1 gaagaagctttcagagctgttgcggtactacacatctgcttctggggacgagatggtttc 60
    E  E  A  F  R  A  V  A  V  L  H  I  C  F  W  G  R  D  G  F

                                                                  ↓
 61 tctgaaggactactgtaccagaatgaaggaaaannagaacnnatctattttatcacaggt 120
    S  E  G  L  L  Y  Q  N  E  G  K  ?  E  ?  I  Y  F  I  T  G


121 cccaggtcc 129
    P  R  S
```

JST85 (novel, myeloid cDNA, no open reading frame)

```
  1 aatggtagtgtctagaaagggtatgtcccttcaagagagaggtgccaatgtccaaccggc 60
    N  G  S  V  *  K  G  Y  V  P  S  R  E  R  C  Q  C  P  T  G
      W  *  C  L  E  R  V  C  P  F  K  R  E  V  P  M  S  N  R  P


 61 ctaataacaatccaggggggtcactgcgacgttcacagaggaacactgccggggcccaac 120
    L  I  T  I  Q  G  G  H  C  D  V  H  R  G  T  L  P  G  P  N
      N  N  N  P  G  G  S  L  R  R  S  Q  R  N  T  A  G  A  Q  P

                            ↓
121 cacaagacgactcaataggaggaaggtcccaggtcc 156
    H  K  T  T  Q  *  E  E  G  P  R  S        ← βgal ORF
      Q  D  D  S  I  G  G  R  S  Q  V          ← myeloid cDNA ORF
```

JST136 (p23, 5'UTR)

```
                                                        ↓
  1 agncccncccacccgtttgtctgnccctctgccccgttcaccatgtcccaggtccc 56
    ?  P  P  T  R  L  S  ?  P  L  P  R  S  P  C  P  R  S
```

JST185 (novel)

```
  1 tttctgggggctttgggaatattcagtctcagataccctggtgactcctcaaagactgtc 60
    F  L  G  A  L  G  I  F  S  L  R  Y  P  G  D  S  S  K  T  V

 61 ctaatggcgctttgtttaatggcgttaccatttatttccgcatcaaacctgttctttcct 120
    L  M  A  L  C  L  M  A  L  P  F  I  S  A  S  N  L  F  F  P

121 gttggatttgtggttgctgagcgagtactatatgttcctagcatggggttctgtatttta 180
    V  G  F  V  V  A  E  R  V  L  Y  V  P  S  M  G  F  C  I  L

181 gtagcccatggatggcaaaagatttcaaacaaaagtgtgctgaaaaagctctcgtgggtt 240
    V  A  H  G  W  Q  K  I  S  N  K  S  V  L  K  K  L  S  W  V

241 tgtctgtccatggtgatactaacccatgccttgaaaacacttcatagaaattgggactgg 300
    C  L  S  M  V  I  L  T  H  A  L  K  T  L  H  R  N  W  D  W

301 gagtcagnatatacattgtttatgtcagccctaaaggtgaataaaaacaatgccaaatta 360
    E  S  ?  Y  T  L  F  M  S  A  L  K  V  N  K  N  N  A  K  L

361 tggaataatgtgggtcatgctctggagaatgagaagaactttgagaaagctttgaaatac 420
    W  N  N  V  G  H  A  L  E  N  E  K  N  F  E  K  A  L  K  Y

421 ttcttgcaggctacccatgttcagccaggtcccaggtccc 460
    F  L  Q  A  T  H  V  Q  P  G  P  R  S
```

JST191 (novel)

```
  1 gcgggagccaccgtaccgccgccgcctcccgggtgggcgcccttctccttggacgccggc 60
    A  G  A  T  V  P  P  P  P  P  G  W  A  P  F  S  L  D  A  G

                  ↓
 61 gacccaggacgaggtcccaggtccc 85
    D  P  G  R  G  P  R  S
```

JST197 (non-muscle myosin, 5'UTR)

```
  1 cccgagcgctcgagaaagtccactcggaagaaccagcgcctggttccccgggcagaccca 60

                  ↓
 61 ggttcaggtcccaggtccc 79
```

JST203 (β 1-4, galactosyltransferase)

```
  1 aactgtcttcgttgccagtgcccaccaccactggactgttgtcgctgccagcttgccctg 60
      L  S  S  L  P  V  P  T  T  T  G  L  L  S  L  P  A  C  P  E

                     ↓
 61 aggagtccccgctgctcggtcccaggtccc 90
      E  S  P  L  L  G  P  R  S
```

JST213 (murine DNAJ-2)

```
  1 tatgatcagncagctggcnccaggaatggtgcagcagatgcagtccgtgtgctccgactg 60
    M  I  ?  Q  L  A  P  G  M  V  Q  Q  M  Q  S  V  C  S  D  C

                 ↓
 61 taatggagaaggtcccaggtccc 83
     N  G  E  G  P  R  S
```

JST219 (murine fat)

```
  1 ggccgttgacgcgaatgagaacctacacactccggtgttctccagctttnaagagaaggg 60
    A  V  D  A  N  E  N  L  H  T  P  V  F  S  S  F  ?  E  K  G

 61 tatagtgaaagaagacgtccctactggctcatcagtcatgaccgtgtcagctcacgacga 120
    I  V  K  E  D  V  P  T  G  S  S  V  M  T  V  S  A  H  D  E

121 agacacaggcagagacngagagatccggtattccattagagatggctctggaattggtgt 180
    D  T  G  R  D  ?  E  I  R  Y  S  I  R  D  G  S  G  I  G  V

181 cttcaggatagatgaagaaacaggtgtnatagagacttcagaccgactggatagagagtc 240
    F  R  I  D  E  E  T  G  V  I  E  T  S  D  R  L  D  R  E  S

241 gacttcccactactggctcacggtctatgccacagatcagggcgtggtgcctctgtcgtc 300
    T  S  H  Y  W  L  T  V  Y  A  T  D  Q  G  V  V  P  L  S  S

301 cttnatagaggtctacatagaggtggagggtgtcaatgacaacgcgccacagacatcaga 360
    ?  I  E  V  Y  I  E  V  E  G  V  N  D  N  A  P  Q  T  S  E

361 acccgtgtactatccagaaatcatggaaaactcacccaaggatgtatctgtggtccagat 420
    P  V  Y  Y  P  E  I  M  E  N  S  P  K  D  V  S  V  V  Q  I

421 tgaggcttttgacccggattctagctccaatgacaaactgacatacagaattacaagcgg 480
    E  A  F  D  P  D  S  S  S  N  D  K  L  T  Y  R  I  T  S  G

                                              ↓
481 gaaccctcaaggattcttctcaatacatcctaagacaggtcccaggtccc
    N  P  Q  G  F  F  S  I  H  P  K  T  G  P  R  S
```

JST232 (RBP-Jκ pseudogene)

```
                                                        ↓
  1 ccctgactctaggcgtctgcttaacccggcttctcaaatgttggnctataggtcccaggt 60
     L  T  L  G  V  C  L  T  R  L  L  K  C  W  ?  I  G  P  R  S
```

Appendix III

LM-PCR and 5'RACE sequence for ET lines

ET2 (5'ETS) LM-PCR sequence

```
  1 tagtctagaggccagatctgaattccggtttgtcctccgcgagctctccgccgccgccgc 60

                                                    Fusion to vector
CD4
                                                                  ↓
 61 ctcctcctcctctctcgcgctctcngtcccgcctggtcctgtcccacccnacgggtgctt
```

ET50 (novel) LM-PCR sequence

```
  1 ctagatctgattcgcccagcgccgcctccgcccgcccgcccgagcaaagtttgatgcgaa 60

                                                                  ↓
 61 cacggcgtgctctaagagctgggtaaactttaggtgcacgaggggcacttaaaaaaaaag 120
```

ET80 (novel) 5'RACE sequence

```
  1 ccgnacagacaaagggacagngagacagacagacagagtcagagagacagagaggtcaat 60
     P  ?  R  Q  R  D  ?  E  T  D  R  Q  S  Q  R  D  R  E  V  N
     ↓
 61 aagacctgtgag
     K  T  C  E
```

ET94 (28S rRNA) LM-PCR sequence

```
  1 gcggagtaactatgactctcttaaggtagccaaatgcctcgtcatctaattagtgacgcg 60

                                  ↓
 61 catgaatggatgaacgagattcccacctggctgtcgtgctg
```

ET105 (novel) LM-PCR (identical to 5'RACE sequence but longer)

```
  1 ggagttcaagagttgtagaggacagagcatttcagtagattcacaagcagtgatacagaa 60

 61 cctagagaatacctgctatnctgagggaaaggtacttgaattagctggagatgagattgg 120

                           ↓
121 aaaggaaactagaaggctggaggcaggagaaagtgat
```

ET124 (28S rRNA) 5'RACE sequence

```
  1 cgctgtccttccctattaacactaaggacactagaagagacccttttcgatttaaggctgt 60
                ↓
 61 tttgcttgtccagcctattctttgtgtcaggtgccgg
```

ET138 (novel) LM-PCR (identical to 5'RACE sequence but longer)

```
  1 aagacaccgtgctcgcgagacttgctgttttccccaattattcttatnttagttgagttt 60
 61 ttatttttagtttctttttaggttttagagaagggtttctctaagctcactctgtagag 120
121 caggcaggcctcggactcagagatccacctgacttgatctctggagggtaggtagggatt 180
                          ↓
181 agaggcctggtggctttgggacagactccaagatccaggttttatccaaa
```

ET143 (novel) 5' RACE sequence

```
  1 atgatggctgattcccagttgactgatgtcaagtcggcttcccagtgcttatattataga 60
 61 agtggaggttgagcctggggttccatgcatgcacagcaaacactcttatcagctaagcta 120
                                                              ↓
121 tgtcctagcccctaactgattttgttcatttgtttttgagacaagatttttctgtctggc
```

Bibliography

Alescio, T. and Cassini, A. (1962). Induction *in vitro* of tracheal buds by pulmonary mesenchyme grafted on tracheal epithelium. *J Exp Zool* **150**, 83-94.

Allen, N. D., Cran, D. G., Barton, S. C., S, H., Reik, W. and Surani, M. A. (1988). Transgenes as probes for active chromosomal domains in mouse development. *Nature* **333**, 852-855.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Bio* **215**, 403-410.

Baker, R. K., Haendel, M. A., Swanson, B. J., Shambaugh, J. C., Micales, B. K. and Lyons, G. E. (1997). *In vitro* pre-selection of gene-trapped embryonic stem cell clones for characterizing novel developmentally regulated genes in the mouse. *Dev Biol* **185**, 201-214.

Bayner, E. M., Shaper, J. H. and Shur, D. (1988). Temporally specific involvement of β 1,4-galactosyltransferase during mouse embryo morula compaction. *Cell* **53**, 145-157.

Beddington, R., Rashbass,P. and Wilson,V. (1992). *Brachyury*-a gene affecting mouse gastrulation and early organogenesis. *Dev Supp **1992***, 157-165.

Bellen, H. J., O'Kane, C. J., Wilson, C., Grossniklaus, U., Pearson, R. K. and Gehring, W. J. (1989). P-element-mediated enhancer detection : a versatile method to study development in *Drosophila*. *Genes & Dev* **3**, 1288-1300.

Bellofatto, V., Shapiro, L. and Hodgson, D. A. (1984). Generation of a Tn5 promoter probe and its use in the study of gene expression in *Caulobacter crescentus*. *Proc Natl Acad Scii USA* **81**, 1035-1039.

Bellusci, S., Henderson, R., Winnier, G., Oikawa, T. and Hogan, B. L. M. (1996). Evidence from normal expression and targeted misexpression that Bone Morphogenetic Protein-4 (BMP-4) plays a role in mouse embryonic lung morphogenesis. *Development* **122**, 1693-1702.

Beltzer, J. P., Fiedler, K., Fuhrer, C., Geffen, I., Handschin, C., Wessels, H. P. and Spiess, M. (1991). Charged residues are major determinants of the transmembrane orientation of a signal-anchor sequence. *J Biol Chem* **266**, 973-978.

Bier, E., Vaessin, H., Shepherd, S., Lee, K., McCall, K., Barbel, S., Ackerman, L., Carretto, R., Uemura, T., Grell, E., Jan, L. Y. and Jan, Y. N. (1989). Searching for pattern and mutation in the *Drosophila* genome with a P-LacZ vector. *Genes & Dev* **3**, 1273-1287.

Bostrom, H., Willetts, k., Pekney, M., Laveen, P., Lindahl, H., Hedstrand, H., Pekna, M., Hellstrom, M., Gebre-Medhin, S., Schalling, N., Nilsson, M., Kurland, S., Tornell, J., Heath, J. and Betsholtz, C. (1996). PDGF-A signalling is a critical event in lung alveolar myofibroblast development and alveogenesis. *Cell* **85**, 863-873.

Bradley, A., Evans, M., Kaufmann, M. H. and Robertson, E. J. (1984). Formation of germ-line chimeras from embryo-derived teratocarcinoma cell lines. *Nature* **309**, 255-256.

Brandon, E. P., Idzerda, R. L. and McKnight, G. S. (1995). Targeting the mouse genome: a compendium of knock-outs (Parts I-III). *Curr Biol* **5**, 625-634, 758-765, 873-1073.

Brown, A., Bernier, G., Mathieu, M., Rossant, J. and Kothary, R. (1995). The mouse *dystonia musculorum* gene is a neural isoform of *bulbous pemphigoid antigen 1*. *Nat Genet* **10**, 301-306.

Camus, A., Kress, C., Babinet, C. and Barra, J. (1996). Unexpected behavior of a gene trap vector comprising a fusion between the Sh *ble* and the *lacZ* genes. *Mol Repro Dev* **45**, 255-263.

Capecchi, M. R. (1989). Altering the genome by homologous recombination. *Science* **244**, 1288-1292.

Casabadan, M. J., Chou, J. and Cohen, S. N. (1980). *In vitro* gene fusion that joins an enzymatically active βgal segment to amino terminal fragments of exogenous proteins: *Escherichia coli* plasmid vectors for the detection and cloning of translational initiation signals. *J Bacteriol* **143**, 971-980.

Casabadan, M. J. and Cohen, S. N. (1979). Lactose genes fused to exogenous promoters in one step using a Mu-lac bacteriophage: *In vivo* control probe for transcriptional control sequences. *Proc Natl Acad Sci USA* **76**, 4530-4533.

Chabot, B., Stephenson, D. A., Chapman, W. M., Besmer, P. and Bernstein, A. (1988). The proto-oncogene *c-kit* encoding a transmembrane tyrosine kinase receptor maps to the mouse W locus. *Nature* **335**, 88-89.

Cheetham, M. E., Brion, J.-P. and Anderton, B. H. (1992). Human homologues of the bacterial heat-shock protein DnaJ are preferentially expressed in neurons. *Biochem J* **284**, 469-476.

Chen, J., Nachabah, A., Scherer, C., Ganju, P., Reith, A., Bronson, R. and Ruley, H. E. (1996). Germ-line inactivation of the murine Eck receptor tyrosine kinase by gene trap retroviral insertion. *Oncogene* **12**, 979-988.

Chen, Z., Friedrich, G. A. and Soriano, P. (1994). Transcriptional enhancer factor -1 disruption by a retroviral gene trap leads to heart- defects and embryonic lethality in mice. *Genes & Dev* **8**, 2293-2301.

Chowdhury, K., Bonaldo, P., Torres, M., Stoykova, A. and Gruss, P. (1997). Evidence for the stochastic integration of gene trap vectors into the mouse germline. *Nucl Acids Res* **25**, 1531-1536.

Church, G. M. and Gilbert, W. (1984). Genomic sequencing. *Proc Natl Acad Sci USA* **81**, 1991-1995.

Clark, J. C., Wert, S. E., Bachurski, C. J., Stahlman, M. T., Stripp, B., Weaver, T. E. and Whitsett, J. A. (1995). Targeted disruption of the surfactant protein B gene disrupts surfactant homeostasis, causing respiratory failure in newborn mice. *Proc Natl Acad Sci USA* **92**, 7794-7798.

Cooke, J., Nowak, M. A., Boerlijst, M. and Maynard-Smith, J. (1997). Evolutionary origins and maintenance of redundant gene expression during metazoan development. *Trends Genetics* **13**, 360-364.

Cooley, L., Kelly, R. and Spradling, A. (1988). Insertion mutagenesis of the Drosophila genome with single P-elements. *Science* **239**, 1121-1128.

Deng, J. M. and Behringer, R. R. (1995). An insertional mutation in the BTF3 transcription factor gene leads to an early postimplantation lethality in mice. *Trans Res* **4**, 264-269.

Denzer, A. J., Nabholz, C. E. and Speiss, M. (1995). Transmembrane orientation of signal-anchor proteins is affected by the folding state but not the size of the N-terminal domain. *EMBO J* **14**, 6311-6317.

Dietrich, W. F., Copeland, N. G., Gilbert, D. J., Miller, J. C., Jenkins, N. A. and Lander, E. S. (1995). Mapping the mouse genome: Current status and future prospects. *Proc Natl Acad Sci USA* **92**, 10849-10853.

Driever, W., SolnicaKrezel, L., Schier, A. F., Neuhauss, S., Malicki, J., Stemple, D. L., Stainier, D., Zwartkruis, F., Abdelilah, S., Rangini, Z., Belak, J. and Boggs, C. (1996). A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* **123**, 37-46.

Drutel, G, Arrang, J.M, Diaz, J, Wisnewsky, C, Schwartz, K, and Schwartz, J.C. (1995). Cloning of *oll*, a putative olfactory receptor and its expression in the developing rat heart. *Recept Channels* **3**, 33-40.

Dunne, J., Hanby, A. M., Poulsom, R., Jones, T. A., Sheer, D., Chin, W. G., Da, S. M., Zhao, Q., Beverley, P. C. L. and Owen, M. J. (1995). Molecular cloning and tissue expression of *FAT*, the human homologue of the *Drosophila fat* gene that is located on chromosome 4q34-q35 and encodes a putative adhesion molecule. *Genomics* **30**, 207-223.

Engels, W. R. (1989). P elements in *Drosophila melanogaster*. In *Mobile DNA*. (ed. Berg, E. M. and Howe, M. M), pp. 437-484. American Society for Microbiology, Washington D.C.

Evans, M. J. and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**, 154-156.

Fantes, J.A., Bickmore, W.A., Fletcher, J.M., Ballesta, F., Hanson, I.M., van Heyningen, V. (1992). Submicroscopic deletions at the WAGR locus, revealed by nonradioactive *in situ* hybridization. *Am J Hum Genet* **6**, 1286-1294.

Forrester, L. M., Nagy, A., Sam, M., Watt, A., Stevenson, L., Bernstein, A., Joyner, A. L. and Wurst, W. (1996). An induction gene trap screen in embryonic stem cells - identification of genes that respond to retinoic acid *in vitro*. *Proc Natl Acad Sci, USA* **93**, 1677-1682.

Freeman, W. H. and Bracegirdle, B. (1966). An Atlas of Histology. London, Heinemann Educational Books, Ltd.

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.

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. USA* **85**, 8998-9002.

Garrity, P. A. and Wold, B. (1992). Effects of different DNA polymerases in ligation-mediated PCR: Enhanced genomic sequencing and *in vivo* footprinting. *Proc Natl Acad Sci, USA***89**, 1021-1025.

Gasca, S., Hill, D. P., Klingensmith, J. and Rossant, J. (1995). Characterization of a gene trap insertion into a novel gene, *cordon-bleu*, expressed in axial structures of the gastrulating mouse embryo. *Dev Genetics* **17**, 141-154.

Goebl, M and Yanagida, M. (1991). The TPR snap helix: a novel repeat motif from mitosis to transcription. *Trends Biochem Sci* **16**, 173-177.

Gossler, A., Joyner, A. L., Rossant, J. and Skarnes, W. C. (1989). Mouse embryonic stem-cells and reporter constructs to detect developmentally regulated genes. *Science* **244**, 463-465.

Gossler, A. and Zachgo, J (1993). Gene and enhancer trap screens in ES cell chimeras. In *Gene Targeting*. (ed. Joyner, A.), pp. 181-227. IRL Press, Oxford.

Gunnery, S. and Matthews, M. B. (1995). Functional mRNA can be generated by RNA polymerase III. *Mol Cell Biol* **15**, 3597-3607.

Hackett, B. P. and Gitlin, J., D (1997). Role of transcription factors in the development of the pulmonary epithelium. In *Lung growth and development*. (ed. McDonald, J. A.), pp. 55-80. Marcel Dekker, Inc, New York.

Haffter, P., Granato, M., Brand, M., Mullins, M. C., Hammerschmidt, M., Kane, D. A., Odenthal, J., vanEeden, F., Jiang, Y. J., Heisenberg, C. P., Kelsh, R. N., FurutaniSeiki, M., Vogelsang, E., Beuchle, D., Schach, U., Fabian, C. and NusseinVolhard, C. (1996). The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Development* **123**, 1-36.

Hammerschmidt, M., Brook, A. and McMahon, A. P. (1997). The world according to *hedgehog*. *Trends Genetics* **13**, 14-21.

Hartenstein, V., Younossihartenstein, A. and Lekven, A. (1994). Delamination and division in the drosophila neurectoderm - spatiotemporal pattern, cytoskeletal dynamics, and common control by neurogenic and segment polarity genes. *Dev Biology* **165** 480-499.

Hartmann, E., Rapoport, T.A. and Lodish, H.F (1989). Predicting the orientation of eukaryotic membrane-spanning proteins. *Proc Natl Acad Sci, USA* **86**, 5786-5790.

Hathaway, H. and Shur, B. D. (1992). Cell surface β 1,4-galactosyltransferase functions during neural crest migration and neurulation *in vivo*. *J Cell Biol* **117**, 369-382.

Hathaway, H. J. and Shur, B. D. (1996). Mammary gland morphogenesis is inhibited in transgenic mice that overexpress cell surface β1,4-galactosyltransferase. *Development* **122**, 2859-2872.

Hawkins, J. D. (1988). A survey on intron and exon lengths. *Nucl Acids Res* **16**, 9893-9908.

Hermann, B. G., Labeit, S, Poustka, A King, T and Lehrach, H (1990). Cloning of the T gene required in mesoderm formation in the mouse. *Nature* **343**, 617-622.

Hicks, G. G., Shi, E.-g., Li, X.-M., Li, C.-H., Pawlak, M. and Ruley, H. E. (1997). Functional genomics in mice by tagged sequence mutagenesis. *Nat Gen* **16**, 338-344.

High, S. (1992). Membrane protein insertion into the Endoplasmic Reticulum - Another Channel Tunnel? *Bioessays* **14**, 535-540.

Hilfer, S. R. (1996). Morphogenesis of the lung: Control of embryonic and fetal branching. *Annu Rev Physiol* **58**, 93-113.

Hislop, A. and Reid, L. (1974). Development of the acinus in the human lung. *Thorax* **26**, 195-205.

Hodgkinson, C. A., Moore, K., Nakayama, A., Steingrimsson, E., Copeland, N. G., Jenkins, N. A. and Arnheiter, H. (1993). Mutation at the mouse micropthalmia locus are associated with defects in a gene encoding a novel helix-loop-helix-zipper protein. *Cell* **74**, 395-404.

Hogan, B., Beddington, R., Costanini, F. and Lacy, E. (1994). *Manipulating the Mouse Embryo. A Laboratory Manual.* Cold Spring Harbor Laboratory Press. 2nd, ed.

Hogan, B. L. M. (1996). Bone morphogenetic proteins: multifunctional regulators of vertebrate development. *Genes & Dev* **10**, 1580-1594.

Horowitz, S. and Davis, J. M. (1997). Lung injury when development is interupted by premature birth. In *Lung growth and development.* (ed. McDonald, J. A.), pp. 577-610. Marcel Dekker Inc, New York.

Huang, E., Nocka, D. R., Beier, D., Chu, T., Y, Buck, J., Lahm, H., W, Wellner, D., Leder, P. and Besmer, P. (1990). The haematopoietic growth factor KL is encoded at the Steel locus and is the ligand of the c-kit receptor, the gene product of the W locus. *Cell* **63**, 225-233.

Hurst, L. D., McVean, G. and Moore, T. (1996). Imprinted genes have few and small introns. *Nat Gen* **12**, 234-237.

Jaenisch, R. (1981). Chromosomal position and activation of retroviral genomes inserted into the germ line of mice. *Cell* **24**, 510-529.

Jaenisch, R. (1988). Transgenic Animals. *Science* **240**, 1468-1474.

Jahner, D. and Jaenisch, R. (1985). Retrovirus-induced *de novo* methylation of flanking host sequences correlates with gene inactivity. *Nature* **315**, 594-597.

Jeanotte, L. J., Rutz, J. C. and Robertson, E. J. (1991). Low level of Hox1.3 expression does not preclude the use of promoterless vectors to generate a targeted gene disruption. *Mol Cell Biol* **11**, 5578-5585.

Jonsson J-I, Wu, Q., Nilsson, K. and Phillips, R. A. (1996). Use of a promoter-trap retrovirus to identify and isolate genes involved in differentiation of a myeloid progenitor cell line *in vitro*. *Blood* **87**, 1771-1779.

Kadowaki, T., Wilder, E., Klingensmith, J., Zachery, K. and Perrimon, N. (1996). The segment polarity gene *porcupine* encodes a putative multitransmembrane protein involved in *wingless* processing. *Genes & Dev* **10**, 3116-3128.

Keller, G. M. (1995). *In vitro* differentiation of embryonic stem cells. *Curr Op Cell Biol* **7**, 862-869.

Kemphues, K. (1988). Genetic analysis of *Caenorhabditis elegans*. In *Developmental Analysis of Higher Organisms*. (ed. Malacinski, G. M.), pp. 193-219. Macmillan, New York.

Kessel, M. and Gruss, P. (1990). Murine developmental control genes. *Science* **249**, 374-379.

King, A. and Melton, D. W. (1987). Characterisation of cDNA clones for *hypoxanthine-guanine phosphoribosyltransferase* from the human malarial parasite, *Plasmodium faciparum*: comparison to the mammalian gene and protein. *Nucl Acid Res* **15**, 10469-10481.

Kingsley, D. M., Bland, A. E., Grubber, J. M., Marker, P. C., Russel, L. B., Copeland, N. G. and Jenkins, N. A. (1992). The mouse *short ear* skeletal morphogenesis locus is associated with defects in a bone morphogenetic member of the TGFβ superfamily. *Cell* **71**, 399-410.

Kirby, M. L. (1993). Cellular and molecular contributions of the cardiac neural crest to cardiovascular development. *Trends Cardiovasc Med* **3**, 18-23.

Klein, R. D., Gu, Q., Goddard, A. and Rosenthal, A. (1996). Selection for genes encoding secreted proteins and receptors. *Proc Natl Acad Sci USA* **93**, 7108-7113.

Koncz, C., Martini, N., Mayerhofer, R., Knocz-kalman, Z., Korber, H., Redei, G. P. and Schell, J. (1989). High frequency T-mediated gene tagging in plants. *Proc Natl Acad Sci USA* **86**, 8467-8471.

Korn, R., Schoor, M., Neuhaus, H., Henseling, U., Soininen, R., Zachgo, J. and Gossler, A. (1992). Enhancer trap integrations in mouse embryonic stem cells give rise to staining patterns in chimeric embryos with a high frequency and detect endogenous genes. *Mech Dev* **39**, 95-109.

Kothary, R., Clapoff, S., Brown, A., Campbell, R., Peterson, A. and Rossant, J. (1988). A transgene containing lacZ inserted into the *dystonia* locus is expressed in the neural tube. *Nature* **335**, 435-437.

Kozak, M. (1987). An analysis of 5' non-coding sequences from 699 vertebrate messenger RNAs. *Nucl Acids Res* **15**, 8125-8148.

Kozak, M. (1987). At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *J Mol Biol* **196**, 947-950.

Krakowsky, J. M., Boissy, R. E., Neumann, J. C. and Lingrel, J. B. (1993). A DNA insertional mutation results in *micropthalmia* in transgenic mice. *Transgenic Res* **2**, 14-20.

Kumar, S. and Simons, J. P. (1993). The effects of terminal heterologies on gene targeting by insertion vectors in embryonic stem cells. *Nucl Acids Res* **21**, 1541-1548.

Kyte, J. and Doolittle, R. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**, 105-132.

Lamb, J. R., Michaud, W. A., Sikorski, R. S. and Hieter, P. (1994). Cdc16p, Cdc23p and Cdc27p form a complex essential for mitosis. *EMBO J* **13**, 4321-4328.

Lamb, J. R., Tugendreich, S. and Hieter, P. (1995). Tetratricopeptide repeat interactions: to TPR or not to TPR? *Trends Biochem Sci* **20**, 257-259.

Li, X. C., Giot, J. F., Kuhl, D., Hen, R. and Kandel, E. R. (1995). Cloning and characterization of 2 related serotonergic receptors from the brain and the reproductive-system of *Aplysia* that activate phospholipase-C. *J Neuroscience* **15**, 7585-7591.

Lu, Q., Hasty, P. and Shur, B. (1997). Targeted mutation in beta1,4-galactosyltransferase leads to pituitary insufficiency and neonatal lethality. *Dev Biol* **181**, 257-267.

Lyon, M. F., Rastan, S. and Brown, S. D. M. (1996). Genetic Variations and Strains of the Laboratory Mouse. Oxford University Press. 3rd, ed.

Macleod, D., Lovell-Badge, R., Jones, S. and Jackson, I. (1991). A promoter trap in embryonic stem (ES) cells selects for integration of DNA into CpG islands. *Nucl Acid Res* **19**, 17-23.

Mahoney, P. A., Weber, U., Onofrechuk, P., Biessmann, H., Bryant, P. J. and Goodman, C. S. (1991). The *fat* tumor supressor gene in *Drosophila* encodes a novel member of the cadherin gene superfamily. *Cell* **67**, 853-868.

Malek, S. N., Yang, C. H., Earnshaw, W. C., Kozac, C. A. and Desidero, S. (1996). p150TSP, a conserved nuclear phosphoprotein that contains multiple tetratricopeptide repeats and binds specifically to SH2 domains. *J Biol Chem* **271**, 6952-6962.

Maniatis, T., Fritsch, E. F. and Sambrook, J. (1982). *Molecular Cloning, A Laboratory Manual.* Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Mansour, S.L, Thomas, K.R and Capecchi, M.R, . (1988). Disruption of the proto-oncogene *int-2* in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* **336**, 348-352.

Martin, G. M. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci USA* **78**, 7634-7638.

McMahon, A. P. and Bradley, A. (1990). The *Wnt-1 (int-1)* proto-oncogene is required for development of a large region of the mouse brain. *Cell* **62**, 1073-1085.

Meisler, M. H. (1992). Insertional mutation of 'classical' and novel genes in transgenic mice. *Trends Genetics* **8**, 341-344.

Miklos, G. and Rubin, G. M. (1996). The role of the genome project in determining gene-function - insights from model organisms. *Cell* **86**, 521-529.

Minoo, P. and King, R. J. (1994). Epithelial-mesenchymal interactions in lung development. *Annu. Rev. Physiol.* **56**, 13-45.

Moon R.T, B. J., Torres M (1997). WNTs modulate cell fate and behavior during vertebrate development. *Trends Genetics* 157-162.

Mooslehner, K., Karls, U. and Harbers, K. (1990). Retroviral integration sites in transgenic *Mov* mice frequently map in the vicinity of transcribed DNA regions. *J Virol* **64**, 3056-3058.

Moyer, J. H., Lee-Tischler, M. J., Kwon, H.-Y., Schrick, J. J., Wilkinson, J. E., Avner, E. D., Sweeney, W. E., Godfrey, V. L., Cacherio, N. L. A. and Woychik, R. P. (1994). Candidate gene associated with a mutation causing recessive polycystic kidney disease in mice. *Science* **264**, 1329-1333.

Nagy, A. and Rossant, J. (1993). Production of completely ES cell-derived fetuses. In *Gene targeting: a practical approach.* (ed. Joyner, A.), pp. IRL Press, Oxford.

Nehls, M., Kyewski, B., Messerle, M., Waldschutz, R., Schuddekopf, K., Smith, A. J. and Boehm, T. (1996). Two genetically separable steps in the differentiation of thymic epithelium. *Science* **272**, 886-889.

Neuhaus, H., Bettenhausen, B., Bilinski, P., Simon-Chazottes, D., Guénet, J.-L. and Gossler, A. (1994). Etl2, A novel putative type-I cytokine receptor expressed during mouse embryogenesis at high levels in skin and cells with skeletogenic potential. *Dev Biol* **166**, 531-542.

Nichols, J., Evans, E. P. and Smith, A. G. (1990). Establishment of germline-competent embryonic stem (ES) cells using differentiation inhibiting activity. *Development* **110**, 1341-1348.

Nichols, J., Davidson, D., Taga, T., Yoshida, K., Chambers, I. and Smith, A. (1996). Complementary tissue-specific expression of Lif and Lif-receptor messenger-RNAs in early mouse embryogenesis. *Mech Dev* **57**, 123-131.

Niwa, H., Araki, K., Kimura, S., Taniguchi, S., Wakasugi, S. and Yamamura, K. (1993). An efficient gene-trap method using poly-A trap vectors and characterization of gene-trap events. *J. Biochem* **113**, 343-349.

Noguchi, A., Reddy, R., Kursar, J. D., Parks, W. C. and Mecham, R. P. (1989). Smooth muscle isoactin and elastin in fetal bovine lung. *Exp Lung Res* **15**, 537-552.

Nüsslein-Volhard, C., Wiescaus, E. and Kluding, H. (1984). Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster* I. Zygotic loci on the second chromosome. *Roux's Arch Dev Biol* **193**, 267-282.

O'Kane, C. and Gehring, W. (1987). Detection *in situ* of genomic regulatory elements in *Drosophila*. *Proc Natl Acad Sci USA* **84**, 9123-9127.

Oka, C., Nakano, T., Wakeham, A., de la Pompa, J., Mori, C., Sakai, T., Okazaki, S., Kawaichi, M., Shiota, K., Mak, T. and Honjo, T. (1995). Disruption of the mouse *RBP-J kappa* gene results in early embryonic death. *Development* **121**, 3291-3301.

Palmer, T. D., Miller, A. D., Reeder, R. H. and McStay, B. (1993). Efficient expression of a protein coding gene under control of an RNA polymerase I promoter. *Nucl Acids Res* **21**, 3451-3457.

Patel, G.C., Greenwood, M.T., Panetta, R., Demchyshyn. L., Niznik, H. and Srikant, C.B. (1995). Mini review - the somatostatin receptor family. *Life Sci* **57**, 1249-1265.

Pierce, R. A., Albertine, K. H., Starcher, B. C., Bohnsack, J. F., Carlton, D. P. and Bland, R. D. (1997). Chronic lung injury in preterm lambs: disordered pulmonary elastin deposition. *Am J Physiol* **272**, L452-L460.

Ramirez-Solis, R., Liu, P. and Bradley, A. (1995). Chromosome engineering in mice. *Nature* **378**, 720-724.

Reith, A. D., Rottapel, R., Giddens, E., Brady, C., Forrester, L. and Bernstein, A. (1990). W mutant mice with mild or severe developmental defects contain distinct point mutations in the kinase domain of the c-kit receptor. *Genes & Dev* **4**, 390-400.

Rijkers, T., Peetz, A. and Ruther, U. (1994). Insertional mutagenesis in transgenic mice. *Trans Res* **3**, 203-215.

Rinchik, E. M. (1991). Chemical mutagenesis and fine-structure functional analysis of the mouse genome. *Trends Genetics* **7**, 15-21.

Robertson, B. (1989). The evolution of neonatal respiratory distress syndrome into chronic lung disease. *Eur Respir J Suppl* **3**, 33S-37S.

Robertson, E., Bradley, A., Kuehn, M. and Evans, M. (1986). Germ-line transmission of genes introduced into cultured pluripotential cells by a retroviral vector. *Nature* **309**, 255-256.

Robertson, E. J., Ed. (1987). Teratocarcinomas and embryoinc stem cells. Practical Approach Series. IRL Press, Oxford.

Robey, E. (1997). Notch in vertebrates. *Curr Opp Gen Dev* **7**, 551-557.

Rohdewohld, H., Weiher, H., Reik, W., Jaenisch, R. and Breindl, M. (1987). Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. *J Virol* **61**, 336-343.

Rudnick, D. (1933). Developmental capacities of the chick lung in chorioallantoic grafts. *J Exp Zool* **66**, 125-153.

Russ, A. P., Friedel, C., Ballas, K., Kalina, U., Zahn, D., Strebhardt, K. and vonMelchner, H. (1996). Identification of genes induced by factor deprivation in hematopoietic cells undergoing apoptosis using gene-trap mutagenesis and site-specific recombination. *Proc Natl Acad Sci* **93**, 15279-15284.

Russell, W. L. (1951). X-ray induced mutations in mice. *Cold Spring Harbor Symp Quant Biol* **16**, 327-336.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). Molecular cloning, a laboratory manual. Cold Spring Harbor, CSH laboratory press.

Scherdin, U., Rhodes, K. and Breindl, M. (1990). Transcriptionally active genome regions are preferred targets for retrovirus integration. *J Virol* **64**, 907-912.

Scherer, C. A., Chen, J., Nachabeh, A., Hopkins, N. and Ruley, H. E. (1996). Transcriptional specificity of the pluripotent embryonic stem cell. *Cell Growth Diff* **7**, 1393-1401.

Schuster-Gossler, K., Zachago, J., Soininen, R., Schoor, M., Korn, R. and Gossler, A. (1994). Gene trap integrations in genes active in mouse embryonic stem cells efficiently detect developmentally regulated gene expression. *Transgenics* **1**, 281-291.

Serafini, T., Colamarino, S. A., Leonardo, E. D., Wang, H., Beddington, R., Skarnes, W. C. and Tessier-Lavigne, M. (1996). *Netrin-1* is required for commisural axon guidance in the developing vertebrate nervous system. *Cell* **87**, 1001-1014.

Shedlovsky, A., King, T. R. and Dove, W. F. (1988). Saturation germ line mutagenesis of the murine *t* region including a lethal allele at the *quaking* locus. *Proc Nat Acad Sci USA* **85**, 180-184.

Shih, C.-C., Stoye, J. P. and Coffin, J. M. (1988). Highly preferred targets for retroviral integration. *Cell* **53**, 531-537.

Skarnes, W. C. (1990). Entrapment vectors : a new tool for mammalian genetics. *Biotech* **8**, 827-831.

Skarnes, W. C., Auerbach, B. A. and Joyner, A. L. (1992). 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.

Skarnes, W. C., Moss, J. E., Hurtley, S. M. and Beddington, R. S. P. (1995). Capturing genes encoding membrane and secreted proteins important for mouse development. *Proc. Natl. Acad. Sci. USA* **92**, 6592-6596.

Smith, A. G. (1991). Culture and differentiation of embryonic stem cells. *J Tiss Cult Meth* **13**, 89-94.

Smith, D., Wohlgemuth, J., Calvi, B. R., Franklin, I. and Gelbart, W. M. (1993). *hobo* enhancer trapping mutagenesis in *Drosophila* reveals an insertion specificity differenet from P elements. *Genetics* **135**, 1063-1076.

Soininen, R., Schor, M., Henseling, U., Tepe, C., Kisters-Woike, B., Rossant, J. and Gossler, A. (1992). The mouse enhancer trap locus 1 (Etl-1): a novel mammalian gene related to *Drosophila* and yeast transcriptional regulator genes. *Mech Dev* **39**, 111-123.

Spradling, A. C., Stern, D. M., Kiss, I., Roote, J., Laverty, T. and Rubin, G. M. (1995). Gene disruptions using P transposable elements: An integral component of the *Drosophila* genome project. *Proc Natl Acad Sci USA* **92**, 10824-10830.

Springer, P. S., Mccombie, W. R., Sundaresan, V. and Martienssen, R. A. (1995). Gene trap tagging of *prolifera*, an essential mcm2-3-5-like gene in *Arabidopsis*. *Science* **268**, 877-880.

Sundaresan, V., Springer, P., Volpe, T., Haward, S., Jones, J. D. G., Dean, C., Ma, H. and Martienssen, R. (1995). Patterns of gene action in development revealed by enhancer trap and gene trap transposable elements. *Genes & Dev* **9**, 1797-1810.

Szczesna-Skorupa, E., Browne, N., Mead, D. A. and Kemper, B. (1995). Positive charges at the NH2 terminus convert the membrane-anchor signal peptide of cytochrome P-450 to a secretory peptide. *Proc Natl Acad Sci USA* **85**, 738-742.

Takeuchi, T., Yamazaki, Y., Katohfukui, Y., Tsuchiya, R., Kondo, S., Motoyama, J. and Higashinakagawa, T. (1995). Gene trap capture of a novel mouse gene, *jumonji*, required for neural-tube formation. *Genes & Dev* **9**, 1211-1222.

Tashiro, K., Tada, H., Heilker, R., Shirozu, M., Nakano, T. and Honjo, T. (1993). Signal sequence trap : A cloning strategy for secreted proteins and type I membrane proteins. *Science* **261**, 600-603.

Ten Have-Opbroek, A. A. W. (1991). Lung development in the mouse embryo. *Exp Lung Res* **17**, 111-130.

Theiler, K. (1972). The House Mouse. Berlin-Heidelberg, Springer-Verlag.

Thompson, S., Clarke, A. R., Pow, A. M., Hooper, M. L. and Melton, D. W. (1989). Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. *Cell* **56**, 313-321.

Thurlbeck, W. M. (1996). Lung Growth and Development. In Spencer's pathology of the lung (ed. P.S. Hasleton) McGraw-Hill, New York pp. 37-78

Torres, M., Stoykova, A., Huber, O., Chowdhury, K., Bonaldo, P., Mansouri, A., Butz, S., Kemler, R. and Gruss, P. (1997). An *α-E-catenin* gene trap mutation defines its function in preimplantation development. *Proc Natl Acad Sci USA* **94**, 901-906.

Townley, D. J., Avery, B. J., Rosen, B. and Skarnes, W. C. (1997). Rapid sequence analysis of gene trap integrations to generate a resource of insertional mutations in mice. *Genome Research* **7**, 293-298.

Tzamarias, D. and Struhl, K. (1995). Distinct TPR motifs of Cyc8 are involved in recruiting the Cyc8-Tup1 corepressor complex to differentially regulated promoters. *Genes & Dev* **9**, 821-831.

Vijaya, S., Steffen, D. L. and Robinson, H. L. (1986). Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites in chromatin. *J Virol* **60**, 683-692.

Vogt, T. F., Jackson-Grusby, L., Wynshaw-Boris, A. J., Chan, D. C. and Leder, P. (1992). The same genomic region is disrupted in two transgene-induced *limb deformity* alleles. *Mamm Genome* **3**, 431-437.

von Melchner, H., deGregori, J. V., Rayburn, H., Reddy, S., Friedel, C. and Ruley, H. E. (1992). Selective disruption of genes expressed in totipotent embryonal stem cells. *Genes & Dev* **6**, 919-927.

von Melchner, H. R, and Ruley, H. E. (1989). Identification of Cellular Promoters by Using a Retrovirus Promoter Trap. *J Virol* 3227-3233.

Wessels, N.K. (1970). Mammalian lung development: interactions in formation and morphogenesis of tracheal buds J Exp Zoology **175**, 455-66.

Wilson, C., Pearson, R. K., Bellen, H. J., O'Kane, C., Grossniklaus, U. and Gehring, W. J. (1989). P-element-mediated enhancer detection: an efficient method for isolating and characterising developmentally regulated genes in *Drosophila. Genes & Dev* **3**, 1301-1313.

Withers-Ward, E. S., Kitamura, Y., Barnes, J. P. and Coffin, J. M. (1994). Distribution of targets for avain retrovirus DNA integration *in vivo. Genes & Dev* **8**, 1473-1487.

Woychik, R. P., Stewart, T. A., Davis, L. G., D'Eustachio, P. and Leder, P. (1990). An inherited *limb deformity* created by insertional mutagenesis in a transgenic mouse. *Nature* **318**, 36-40.

Wurst, W., Rossant, J., Prideaux, V., Kownacka, M., Joyner, A., Hill, D. P., Guillemot, F., Gasca, S., Cado, D., Auerbach, A. and Ang, S.-L. (1995). A large-scale gene trap screen for insertional mutations in developmentally regulated genes in mice. *Genetics* **139**, 889-899.

Yenofsky, R. L., Fine, M. and Pellow, J. W. (1990). A mutant *neomycin phosphotransferase-II* gene reduces the resistance of transformants to antibiotic selection pressure. *Proc Natl Acad Sci USA* **87**, 3435-3439.

Yoshida, M., Yagi, T., Furuta, Y., Takayanagi, K., Kominami, R., Takeda, N., Tokunaga, T., Chiba, J., Ikawa, Y. and Aizawa, S. (1995). A new strategy of gene trapping in ES cells using 3' RACE. *Trans Res* **4**, 277-287.

You, Y., Bergstrom, R., Klemm, M., Lederman, B., Nelson, H., Ticknor, C., Jaenisch, R. and Schimenti, J. (1997). Chromosomal deletion complexes in mice by radiation of embryonic stem cells. *Nat Gen* **15**, 285-288.

Zhou, L., Dey, C. R., Wert, S. E. and Whitsett, J. A. (1996). Arrested lung morphogenesis in transgenic mice bearing an SP-C-TGF-β1 chimeric gene. *Dev Biol* **175**, 227-238.

Ziegelhoffer, T., Johnson, J. L. and Craig, E. A. (1996). Protein folding: chaperones get Hip. *Curr Biology* **6**, 272-275.