Nucleic Acids Research, 1997, Vol. 25, No. 8 **1531–1536**

Evidence for the stochastic integration of gene trap vectors into the mouse germline

Kamal Chowdhury*, Paolo Bonaldo+, Miguel Torres§, Anastasia Stoykova and Peter Gruss

Department of Molecular Cell Biology, Max Planck Institute of Biophysical Chemistry, Am Fassberg, D-37077 Göttingen, Germany

Received January 17, 1997; Revised and Accepted February 27, 1997

ABSTRACT

A large scale insertional mutagenesis experiment was performed in embryonic stem (ES) cells by introducing two types of gene trap vectors into the genome. These cell lines carrying mutations were introduced into the mouse germline. In order to assess the feasibility of a large scale cloning of the targeted genes from these lines, we have isolated and characterized 55 trapped exons from the corresponding ES cells. Analysis of the data has revealed that vectors containing or lacking an internal ribosome entry site (IRES) can integrate into the ES cell genome stochastically. The targeted genes comprise 30% known genes, 20% expressed sequence tags (ESTs) and 50% novel or unknown genes. The known genes belong to several major classes and represent complete or partial knockouts. Using currently available methods or modifications of them, it should be feasible to do a large scale cloning of trapped genes from the mouse ES cell lines.

INTRODUCTION

The availability of a large number of experimentally generated mutants in *Drosophila melanogaster* and *Caenohrabditis elegans* and isolation of the corresponding genes has greatly facilitated study of the molecular basis underlying the development of these organisms. However, one of the most fundamental and challenging questions addresses the genetic and molecular mechanisms governing mammalian development. The classical genetic routes are hampered by the much larger genome size and development of the embryo *in utero* and thus are an inefficient way to decipher genes involved in controlling mammalian development.

Among the currently available methods, the greatest disadvantage of chemical- or radiation-induced mutagenesis is the extremely lengthy and cumbersome way of identifying and cloning the mutated gene. Furthermore, the expression pattern of the mutated gene cannot be studied before cloning.

The large scale generation of transgenic mice by introduction of exogenous DNA into the mouse genome by retroviral transfer or microinjection of appropriate constructs may lead to recessive phenotypes (1). So far only a few reports about cloning of host transcripts associated with retrovirus-induced mutation have appeared. Only ~5% of retrovirus and 10% of transgenic insertions have been found to cause recessive phenotypes (1). The transgenic insertions generated by microinjection often cause deletion and rearrangement at the locus. Again, cloning of the affected transcriptional unit requires considerable effort and is thus unsuitable for a large scale routine mutagenesis approach.

Another alternative would be the random isolation and sequencing of mammalian cDNAs from a region or cell/tissue-specific subtracted or normal libraries and analysis of their expression patterns by the whole mount *in situ* method. After isolation of the corresponding genomic clones, one could knockout those genes with interesting expression patterns by homologous recombination. Needless to say, although this approach is very precise, it is the least suitable for large scale mutagenesis, solely due to the time factor involved.

In contrast, the gene trap approach in mice (2), adapted and developed from the earlier promoter/enhancer trap protocols used in *Drosophila* (3,4), has very elegantly combined molecular biology and embryonic stem (ES) cell technology to circumvent the difficulties inherent in the procedures mentioned above. In this method, the gene trap vector contains a promotorless *lacZ* (β-galactosidase) reporter gene carrying a splice acceptor at the 5′-end and it is introduced into the ES cell genome by electroporation (5). Such molecularly tagged ES cells are then used to generate transgenic mice. Integration into an actively transcribed gene would thus produce a fusion transcript between the endogenous gene and the *lacZ* gene. This often interferes with normal functioning of the endogenous gene and may lead to a mutant phenotype. Furthermore, due to the transcriptional fusion, expression of the *lacZ* tag faithfully mimics expression of the locus and can be easily detected in embryos or adults by a simple histochemical staining procedure. Most important of all, using the quick 5′-RACE PCR method (6), the mutated exon can be easily cloned from the spliced fusion transcript without having to clone the insertion site from the genomic DNA. Using this strategy, it is conceivable to saturate the ES cell genome with the cloning and expression tag and freeze all these cell lines for future mouse production. According to an estimate, one would need to produce 30 000 cell lines to target all genes expressed in ES cells (7). An additional advantage of the gene trap method is the possibility of pre-selecting the targeted ES cell clones with interesting expression patterns of choice *in vitro* before generating the mutant mice. Another possibility of pre-selecting the targeted ES cells for interesting genes is to test them for their responsiveness to various growth and differentiation factors *in vitro* (8). This also allows a considerable reduction in the number of mice to be studied.

*To whom correspondence should be addressed. Tel: +49 551 2011507; Fax: +49 551 2011504; Email: kchowdh@gwdg.de

Present addresses: +Institute of Histology and Embryology, University of Padova, Via Trieste 75, I-35121 Padova, Italy and §Departameno de Immunologia y Oncologia, Centro Nacional de Biotecnologia, Universidad Autonoma, Madrid 28049, Spain

Using the gene trap strategy, among others, Chen *et al.* (9), Skarnes *et al.* (5,10), Forrester *et al.* (8), Takeuchi *et al.* (11) and Serafani *et al.* (12) have been able to isolate and study new developmental regulatory genes and produce the corresponding mutant mice. In a large gene trap screen, Wurst *et al.* (13) have analyzed the *lacZ* expression pattern of 279 insertion events in chimeric mouse embryos. Of these, 13% showed restricted, 32% showed widespread and 55% did not show any *lacZ* expression in embryos at day 8.5 post-coitum (pc). One third of these negative clones showed expression at day 12.5 pc. This is an interesting observation because it shows that the gene trap strategy can be used to isolate and study genes with a temporally and spatially restricted expression pattern during murine embryogenesis.

Analysis of expression patterns in chimeric embryos has some disadvantages. Due to the time and work necessary for the production of chimeras, one can only study the expression pattern of a very limited number of embryonic stages and not of adults. Many regulatory genes are also expressed later during development and in specific adult tissues. This is especially true for genes controlling mammalian memory and behavior. Furthermore, after completion of the initial analyses in the chimeras, some ES cell lines may not enter the germline and no mutant mice can be produced from these lines to study the phenotype. To circumvent this problem, we are now attempting a large scale analysis of expression pattern using mice that are heterozygous for the gene trap insertion.

However, thus far no information is available addressing the molecular nature of the insertion sites for a large number of trap events. It has been speculated, but not demonstrated, that the insertion event is random and that genes of many classes can be targeted. This information is a prerequisite for large scale production of targeted ES cells, because it would show that this strategy allows insertion without any apparent bias. Theoretically, there is a remote possibility that due to chromatin- or transcription-related sequence constraints within the ES cell genome, not all sites might be available for vector integration. The number and classes of genes that can be targeted depend also on the vector used for the gene trap. Currently used vectors are dependent on the endogenous initiation codon for translation of the reporter gene and only those trapping events that generate a transcriptional fusion between the reporter and the endogenous gene in the correct frame and orientation can be detected. In an attempt to expand the variety of trapped genes, we have therefore inserted the internal ribosome entry site (IRES) from encephalomyocarditis virus (15) between the splice acceptor and the reporter sequence of the vector. In this case, translation of the reporter gene will be independent of the endogenous initiation codon and start at the ATG present in the IRES sequence. Kim *et al.* (15) showed that *lacZ* was expressed throughout chimeric embryos derived from ES cells transfected with a phosphoglycerate kinase promoter–*neo*–IRES–*lacZ* vector. This result suggests the absence of any tissue specificity for IRES function. The use of the IRES sequence in the study of mammalian transgenesis has been reviewed by Mountford and Smith (16).

Here we present an analysis of 55 different exons isolated and mutated by gene trapping. Examination of the cloning data show that we have trapped 16 known genes, 11 genes with homology to expressed sequence tags (ESTs) and 28 new genes. The nature of the known genes indicate that the trapping event is stochastically distributed within the genome. Several classes of genes were trapped which code for proteins with different functions and found at all major subcellular localizations. The insertions were present at different levels in the individual genes, without any apparent bias, thus leading to complete or partial knockouts. These observations suggest that the IRESβGeo vector can be used successfully to capture all classes of genes in the mouse.

MATERIALS AND METHODS

Gene trap vectors

The SAβgeo vector (kindly provided by W.C.Skarnes) contains the splice acceptor sequence from the mouse En-2 gene (2) attached upstream of promotorless wild-type βgeo (14). The construct contains its own ATG codon and a polyadenylation signal at the 3′-end.

IRESβgeo was obtained by introducing the IRES from the EMC virus (15) (kindly provided by T.Takeuchi and T.Higashinakagawa) between the splice acceptor and the βGeo sequences. We first constructed an IRES–*lacZ* fusion plasmid as described in detail by Kim *et al.* (15). In a second step, the entire IRES and part of the fused *lacZ* sequence was excised with *Eco*RV and introduced into the *Bgl*II (blunt ended)/*Eco*RV-cleaved SAβgeo vector. This reconstructed the entire *lacZ* sequence and introduced the IRES element between the splice acceptor site and the βgeo sequence.

Electroporation and screening of ES cells

Electroporation and screening of ES cells was done essentially as described in detail by Wurst *et al.* (17). R1 ES cells (18) were routinely maintained on a monolayer of mitotically inactived primary embryo fibroblasts in Dulbecco's modified Eagle's medium, 15% fetal calf serum, 1000 U/ml LIF. In a typical experiment, 10^7 ES cells (18) were electroporated with 30 µg linearized vector DNA in 1 ml phosphate-buffered saline, by applying a pulse of 250 V and 500 µF. Cells were plated into 10 cm dishes and allowed to recover for 24 h before adding 250 µg/ml G418 (Gibco-BRL) for selection of neomycin-resistant colonies. After 7–10 days, single neomycin-resistant colonies were picked and expanded for further analyses. *lacZ*-expressing clones were detected by staining the cells for β-galactosidase activity. Generation of mice carrying the gene trap mutation was by morula aggregation as described in detail by Nagy *et al.* (18).

Molecular cloning of the trapped exons by 5′-RACE PCR

This was done essentially as reported by Frohman *et al.* (6) and adapted by Skarnes *et al.* (5) for the isolation of gene trap sequences. The source of reagents was the 5′-RACE kit from Gibco-BRL. Starting material was 1–2 µg total RNA isolated from the corresponding ES cell clones. After nested amplifications in most cases only one visible PCR product was obtained. This was cloned into the pGem-T vector (Promega) and transformed into DH5α bacterial hosts. In order to unequivocally determine the sequences of the trapped exons, 5–10 bacterial colonies were picked for each line and the isolated plasmid DNAs were sequenced by the standard double-strand sequencing protocol using the Pharmacia sequencing kit. Sequences obtained were examined by the GCG sequence analysis program and compared with the GenBank/EMBL and SwissProt sequence databanks.

Table 1. Identification of the trapped exons from gene trap lines by 5′-RACE PCR

No.	Clone (vector)	GL	5′-RACE (bp)	Gene (accession no., insertion site)
I. Sequences with homology to known genes				
1	gt o-3 (SAβGeo)	–	66	Prp8, p220 (Z24732[a])
2	gt vii-28 (SAβGeo)	+	585	Spnr (X84692, 1149)
3	gt vii-45 (SAβGeo)	+	416	α E-catenin (X59990, 1970)
4	gt xvi-23 (IRES)	+	494	NFκB, p50 subunit (M57999, 1582)
5	gt xvi-30 (IRES)	+	345	α-enolase (X52379, 335)
6	gt xvi-34 (IRES)	+	366	Ubiquitin hydrolase (H06451, Q01477, 326)
7	gt xvi-36 (IRES)	+	580	Nucleolar protein N038 (M33212, 918).
8	gt xvi-46 (IRES)	+	250	Muscle phosphatase PP1M M110 (S74907, 2834)
9	gt xvi-74 (IRES)	?	401	ADPRP (X14206, 2586)
10	gt xvi-78 (IRES)	+	252	MAP-1B (X51396, 339)
11	gt xvi-108 (IRES)	nd	259	Laminin B2 (J02930, 4179)
12	gt xvi-169 (IRES)	+	52	R-PTP-κ (L10106, 1230)
13	gt xviii-72 (IRES)	+	289	TUP1-like enhancer of split (X75296, 2543)
14	gt xviii-79 (IRES)	+	281	Bovine γ-COP (X70019, 1225)
II. Sequences with homology to ESTs and known ORFs				
15	gt x-218 (SAβGeo)	+	303	EST (T31439, D20245)
16	gt xiii-43 (SAβGeo)	+	272	EST (R40887)
17	gt xv-1 (SAβGeo)	+	332	EST (T66211, F12483)
18	gt xvi-43 (IRES)	+	551	EST (S68074)
19	gt xvi-52 (IRES)	+	125	HUMORF (D25304)
20	gt xvi-76 (IRES)	nd	193	EST (R47074)
21	gt xvi-60 (IRES)	+	653	EST (R31173)
22	gt xvi-80 (IRES)	+	570	EST (H19271)
23	gt xvi-109 (IRES)	+	181	EST (GTPase activator, H20358, H18374)
24	gt xvi-136 (IRES)	+	882	EST (Z19131, T74007)
25	gt xvi-178 (IRES)	?	703	HUMORF S53, Alzheimer locus (L40398)
III. Sequences without homology and containing ORF/no ORF				
26	gt x-91 (SAβGeo)	+	365	No homology, ORF
27	gt xiii-45 (SAβGeo)	+	223	No homology, ORF
28	gt xiv-138 (SAβGeo)	+	368	No homology, ORF
29	gt xiv-109 (SAβGeo)	+	189	No homology, ORF
30	gt xv-53 (SAβGeo)	+	348	No homology, ORF
31	gt xvi-21 (IRES)	+	71	No homology, ORF
32	gt xvi56/57 (IRES)	+	218	No homology, ORF
33	gt xvi-73 (IRES)	+	598	No homology, ORF
34	gt xvi-79 (IRES)	+	282	No homology, ORF
35	gtxvi-75 (IRES)	+	269	No homology, ORF
36	gt xvi-85 (IRES)	+	462	No homology, ORF
37	gt xvi-91 (IRES)	+	129	No homology, ORF
38	gt xvi-92 (IRES)	–	562	No homology, ORF
39	gt xvi-94 (IRES)	–	220	No homology, ORF
40	gt xvi-97 (IRES)	+	360	No homology, ORF

Table 1. *continued.*

41	gt xvi-129 (IRES)	+	168	No homology, ORF
42	gt xvi-133 (IRES)	–	343	No homology, ORF
43	gt xvi-152 (IRES)	nd	220	No homology, ORF
44	gt xvi-175 (IRES)	nd	246	No homology, ORF
45	gt xvi-180 (IRES)	–	430	No homology, ORF
46	gt xviii-47 (IRES)	+	298	No homology, ORF
47	gt xviii-73 (IRES)	+	551	No homology, ORF
48	gt iv-3 (SAβGeo)	+	231	No homology, no ORF
49	gt xvi-16 (IRES)	+	155	No homology, no ORF
50	gt xvi-43 (IRES)	+	570	No homology, no ORF
IV. Lines from which multiple RACE products (mr) were isolated				
51	gt x-7 (SAβGeo)	+	109 and mr	No homology, ORF; proteins S8, S12 and L3
52	gt xvi-1 (IRES)	+	173 and 207	Both no homology, no ORF
53	gt xvi-26 (IRES)	nd	170 and 424	No homology, ORF; a new mouse forkhead-containing gene homologous to the human T cell leukemia virus enhancer factor (P32314)
54	gt xvi-56 (IRES)	nd	135	LINE (L1) repeat (X59214)
55	gt xvi-87 (IRES)	nd	218	LINE (L1) repeat (X59224)

GL, germline transmission; +, yes; –, no; nd, not yet determined; ?, in progress.
[a]Exact insertion site could not be determined from the sequence of the RACE product. The homology to yeast Prp8 was detected by comparing the mouse cDNA sequence isolated using the 66 bp RACE fragment.

RESULTS AND DISCUSSION

Comparison of the efficiency of the IRESβgeo and SAβgeo vectors

Two types of vectors were used. One of them, designated SAβgeo, produces a fusion protein containing both the β-galactosidase (β-gal) and the neomycin resistance activities. It therefore serves as both a reporter and a selection marker. Due to the absence of a promoter, a transcriptional fusion mediated by the upstream splice acceptor and a translational fusion to the targeted endogenous protein is necessary for effective reporter and selection activity.

However, due to the necessity for translational fusion, at best only one in six insertions will be fully productive. Furthermore, fusion of βgeo protein to an endogenous protein might have unpredictable effects, since the βgeo sequence might partially or completely lose its reporter and/or selection activity. It has been demonstrated that β-gal activity is lost when βGeo is fused to a signal peptide, therefore βgeo fusions with secreted molecules cannot be detected (10). Because of these constraints, some classes of genes may be absent or under-represented after trapping with the SAβgeo vector. In order to circumvent these problems, we introduced an IRES (15) between the splice acceptor and the βgeo sequences. In this case, the βgeo sequence will be independently translated from the IRES irrespective of the reading frame of the fusion transcript. Therefore, 50% of the gene trap insertions could be productive. Indeed, the number of neomycin-resistant colonies we obtained with IRESβgeo vector was ~3-fold higher than that obtained with SAβgeo. The proportion of *lacZ*-positive colonies was also increased: only 30% of the colonies obtained with SAβgeo expressed β-gal activity, but this proportion rose to 75% when IRESβgeo was used. There was a broad range of distribution and intensity of β-gal staining in ES cells, with several clones showing *lacZ* expression in differentiated cells but not in undifferentiated ES cells. After introduction of selected clones *in vivo*, we observed different temporal and spatial patterns of *lacZ* activation during embryogenesis (data not shown). Taken together, these observations suggest that the IRESβgeo vector is effective for trapping genes at high efficiencies and that it can be used also to detect genes expressed at very low levels in ES cells.

Insertion of the gene trap vectors into the mouse genome is stochastic

One of the main predictions and prerequisites of the gene trap mutagenesis approach is that insertion of the vector into the mouse genome is stochastic or unbiased. To answer this question, we isolated and sequenced 55 trapped exons from targeted ES cell lines. The size of the isolated trapped sequences varies between 52 and 882 bp and they can be classified into four different groups (Table 1).

The first group (I) contains 14 clones whose sequences are known and identical to a mouse gene or contain sequence identity at the amino acid level to a gene from another organism. The GenBank accession number and the transcriptional fusion site is shown for each of them.

The second group (II) contains sequences which are identical at the nucleotide level or at the amino acid level to known ESTs. The GenBank accession numbers are also shown. The finding that our gene trap approach has successfully trapped many ESTs is very interesting, because these gene trap lines will provide mice with *lacZ* expression tags for individual ESTs and possibly mutant phenotype data for the corresponding genes. This approach will therefore be at least partially a functional complement to the EST project and circumvent the need to isolate the genomic clones for future knockout experiments.

The third group (III) contains sequences which have no significant homology to any public domain databank sequence. Sequences 26–47 contain open reading frames (ORFs) and are

likely to represent coding exons. Sequences 48–50 do not contain any obvious ORF. They may represent 5′- or 3′-non-coding exons or simply putative sequence mistakes introduced by the molecular manipulations during the 5′-RACE PCR procedures. These sequences do not contain any region of identity to each other and therefore likely represent separate genes. However, there is a very remote possibility that two RACE products might belong to the same gene and represent insertions in separate introns of the same gene. We are currently analyzing the *lacZ* expression patterns of the mutant mouse lines. After completion of these studies, they will be published separately with the corresponding cDNA sequences.

The last group (IV) contains five ES cell clones from which multiple 5′-RACE products or mouse LINE (L1) repeat sequences were isolated. The reasons for obtaining multiple products are not clear. One possibility is splicing from several upstream donors to the acceptor of the vector. These donors may belong to the same or a different upstream gene. It should be possible to distinguish between these possibilities by isolating the cDNAs corresponding to the individual trapped exons; if the RACE product belongs to separate genes, one should isolate separate cDNAs. Another, less likely, possibility is insertion of the vector into a site upstream of the first splice donor of the targeted gene. In this kind of fusion transcript, the most 5′ splicing element will be an acceptor. Such transcripts may be unstable and *cis* or *trans* splicing to other splice donors may occur.

The total number of exons cloned and analyzed was 55. Among them, 16 (nos 1–14, 54 and 55) represent known genes, partial sequences of 11 are present in the databanks as ESTs and 28 represent novel genes (Table 2). Therefore, ~50% of the sequences cloned from gene trap lines are new. Furthermore, since little or no expression data is available for the ESTs, ~70% of the trapped genes can be considered as novel (Table 2).

Taken together, characterization of the nature of all trapped exons clearly indicates that gene trap insertion into the mouse genome is stochastic or unbiased. In this regard, it is worth mentioning that in one case we trapped the same gene (gt xvi-169, R-PTP-κ) as Skarnes *et al*. (10). This is interesting because different kinds of vectors were used by the two groups. We used the IRESβgeo vector, whereas Skarnes *et al*. used a specifically designed secretory trap vector (10). Thus the IRES vector can be used as a more general gene trapping vector. Trapping of the R-PTP-κ gene by us and by Skarnes group (10) suggests that besides predominant random integration of the gene trap vector, there might also be some hotspots for recombination in ES cells. Cloning and sequencing of a much larger number of trapped exons will be necessary to address this issue.

Known targeted genes include all major classes

Further analysis of 16 trapped genes with identity or similarity to known sequences shows that they fall into several different classes (Table 3). They include genes coding for nuclear, cytoplasmic, cytoskeletal, membrane and extracellular proteins. Proteins coded by these trapped genes play roles in different functions, like transcription, splicing, ribosome assembly, RNA binding, enzymatic reactions, vesicular transport and cell adhesion. The largest group (six genes) code for nuclear proteins and three of these are transcription factors. Therefore, several classes of genes were trapped, coding for proteins with many different functions and found at all major subcellular localizations.

If we extrapolate this finding to the trapped ESTs and unknown genes, one could assume that they will expand this list further to include many other kinds of genes. This provides for the first time direct evidence that gene trap vector insertion into ES cells by electroporation is truly a stochastic event and includes all major classes of genes.

Table 2. Gene trapping statistics

Known genes	17 (30%)
Unknown genes	28 (50%)
ESTs	11 (21%)

Table 3. Various classes of genes are tagged by the gene trap

Class	Line, gene	Description
Nuclear proteins		
Transcription factors	gt xvi-23, NFκB	Mouse, p50 DNA binding subunit
	gt xviii-72, TUPLE1	Human, TUP1-like enhancer of split
	gt xvi-26	Mouse, new forkhead gene
Other nuclear proteins	gt xvi-74, ADPRP	Mouse, poly(ADP-ribose) polymerase
	gt xvi-36, NO38	Mouse, nucleolar protein
	gt 0–3, Prp8, p220	Human/yeast, splicing factor
Cytoplasmic enzymes		
	gt xvi-30, α-enolase	Mouse, converts phosphoglycerate to phosphoenol pyruvate
	gt xvi-46, PP1M M110	Rat, regulatory subunit of smooth muscle protein phosphatase 1M
	gt xvi-34	Human/yeast, ubiquitin C-terminal hydrolase
Cytoskeletal proteins		
	gt vii-28, Spnr	Mouse, microtubule-associated RNA binding protein
	gt xvi-78, MAP-1B	Mouse, microtubule-associated protein
Membrane-associated proteins		
Adhesion molecules	gt xvi-169, R-PTP-κ	Mouse, receptor protein tyrosine phosphatase κ, adhesion molecule
	gt vii-45, α E-catenin	Mouse, cadherin-associated molecule
Transport vesicle proteins	gt xviii-79, γ-COP	Human/bovine, subunit of coatomer-coated vesicles
Extracellular proteins		
	gt xvi-108, Laminin B2	Mouse, extracellular matrix protein
Unknown		
	gt xvi-178, S53	Human, hypothetical protein of Alzheimer locus

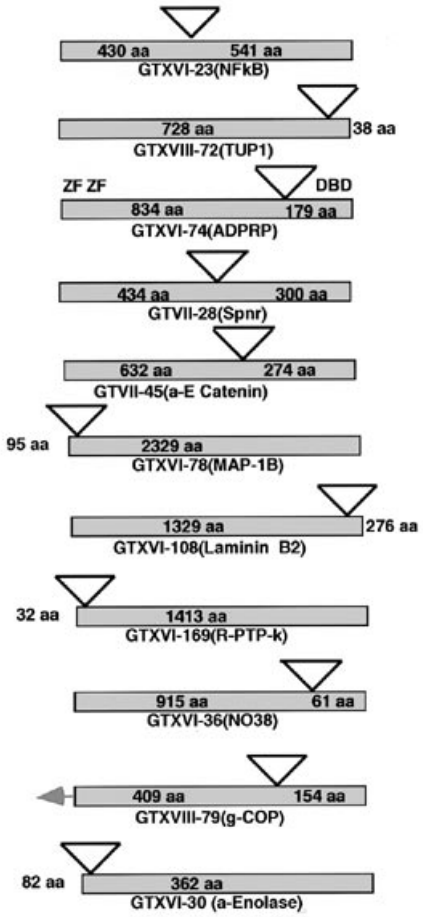

Figure 1. Insertion sites in known genes

Gene trap targeting events create complete or partial knockouts

The final goal of the gene trap project is the generation of mutant mice to carry out functional studies for a large number of genes. A prerequisite for this is the assumption that many insertions into individual genes will completely or partially inactivate the function of the endogenous protein. To investigate this, we analyzed the insertion sites in the known genes and the results are shown in Figure 1. Clearly, there is no preference for insertion site into individual mouse genes when using the SAβGeo or the IRESβGeo vectors. Insertions could be found either at the 5′-end, at the 3′-end or in the middle of a gene. Most insertions at the 5′-end of the trapped genes should completely knockout the corresponding genes. In other cases, one might obtain interesting phenotypes representing partial inactivation of the gene products or as a result of a dominant gain of function. A similar prediction can be done for insertions into unknown genes or into ESTs. However, the final phenotypic characterization will only be available after the corresponding mice have been bred to homozygosity. This is currently in progress. In all the cases, the trapped gene will be tagged with the *lacZ* sequence, thereby allowing easy analysis of the expression pattern and cell lineage analysis.

Since most of our trapped ES cell lines are in the germline and exist as heterozygous mice, they should provide a rich source of experimentally generated mouse mutants. Our current approaches are aimed at developing improved methods for 5′-RACE PCR which should allow us to produce a larger number of amplified products and to sequence them directly without cloning. We have also developed an exon trap vector which lacks the splice acceptor site and upon insertion should be directly fused to a coding exon to exert its selection and reporter activities. Attempts are also being made to obtain sequence information about the insertion site in these lines by direct genomic sequencing.

Due to the simple detection of expression patterns by β-gal staining, it should be possible to tag a large number of genes in ES cells, analyze their expression pattern in the embryo and finally clone and sequence them. This should generate an extremely valuable source of mouse mutants, providing tools for or leading to molecular dissection of vertebrate development. Therefore, using these criteria, large scale gene trapping in ES cells and subsequent production of the corresponding mutant mice will be fruitful and feasible.

ACKNOWLEDGEMENTS

We gratefully acknowledge the excellent technical assistance of F.Uhde-Micknaus for DNA sequencing and of M.Breuer, S.Mahsur, R.Scholz and M.Walther for cell culture, morula aggregation and mouse work. We thank H.P.Geithe for oligonucleotide synthesis and the laboratory of E.Wimmer for providing the IRES–CAT plasmid used as the source of the IRES sequence. Thanks are also due to R.Fritsch and D.Treichel for help with the computer, A.Mansouri for initial participation in the project and A.Kobbe for sharing the unpublished Prp8 cDNA data. This project was generously supported by Amgen Inc., Thousand Oak, CA and by the Max Planck Society. PB and MT were supported by EC fellowships.

REFERENCES

1 Jaenisch,R. (1988) *Science*, **240**, 1468–1474.
2 Gossler,A., Joiner,A., Rossant,J. and Skarnes,W. (1989) *Science*, **244**, 463–465.
3 Wilson,C., Kurth Pearson,R., Bellen,H., O'Kane,C., Grossniklaus,U. and Gehring,W. (1989) *Genes Dev.*, **3**, 1301–1313.
4 Bellen,H., O'Kane,C.,Wilson,C., Grossniklaus,U., Kurth Pearson,R.and Gehring,W. (1989) *Genes Dev.*, **3**, 1288–1300.
5 Skarnes,W.C., Auerbach,A. and Joyner,A.L. (1992) *Genes Dev.*, **6**, 903–918.
6 Frohman,M.A., Dush,M.K. and Martin,G.R. (1988) *Proc. Natl. Acad. Sci. USA*, **85**, 8988–9002.
7 Rossant,J. and Hopkins,N. (1992) *Genes Dev.*, **6**, 1–13.
8 Forrester,L., Nagy,A., Sam,M., Watt,A., Stevenson,L., Bernstein,A., Joyner,A. and Wurst,W. (1996) *Proc. Natl. Acad. Sci. USA*, **93**, 1677–1688.
9 Chen,Z., Friedrich,G. and Soriano,P. (1994) *Genes. Dev.*, **8**, 2293–2301.
10 Skarnes,W.C., Moss,J., Hurtley,S. and Beddington,R. (1996) *Proc. Natl. Acad. Sci. USA*, **92**, 6592–6596.
11 Takeuchi,T., Yamazaki,Y., Katoh-Fukui,Y., Tsuchya,R., Kondo,S., Motoyama,J. and Higashinakagawa,T. (1995) *Genes Dev.*, **9**, 1211–1222.
12 Serafani,T., Colamarino,S., Leonardo,E., Wang,H., Beddington,R., Skarnes,W. and Tessier-Lavigne,M. (1996) *Cell*, **87**, 1001–1014.
13 Wurst,W., Rossant,J., Prideux,V., Kownacka,M., Joyner,A., Hill,D., Guillermot,F., Gasca,S., Cado,D., Auerbach,A. and Ang,S. (1995) *Genetics*, **139**, 889–899.
14 Friedrich,G. and Soriano,P. (1991) *Genes Dev.*, **5**, 1513–1523.
15 Kim,D., Kang,H., Jang,S. and Shin,H. (1992) *Mol. Cell. Biol.*, **12**, 3636–3643.
16 Mountford,P. and SmithA. (1995) *Trends Genet.*, **11**, 179–184.
17 Wurst,W. and Joyner,A.L. (1993) In Joyner,A.J. (ed.), *Gene Targeting: A Practical Approach.* Oxford University Press, Oxford, UK, pp 33–61.
18 Nagy,A. and Rossant,J. (1993) In Joyner,A.J. (ed.), *Gene Targeting: A Practical Approach.* Oxford University Press, Oxford, UK, pp. 147–178.