



Rapid sequence analysis of gene trap integrations to generate a resource of insertional mutations in mice.

D J Townley, B J Avery, B Rosen, et al.

Genome Res. 1997 7: 293-298

Access the most recent version at doi:[10.1101/gr.7.3.293](https://doi.org/10.1101/gr.7.3.293)

References

This article cites 16 articles, 11 of which can be accessed free at:
<http://genome.cshlp.org/content/7/3/293.full.html#ref-list-1>

License

Email Alerting Service

Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Copyright © Cold Spring Harbor Laboratory Press

GENOME METHODS

Rapid Sequence Analysis of Gene Trap Integrations to Generate a Resource of Insertional Mutations in Mice

David J. Townley, Brian J. Avery,¹ Barry Rosen,² and William C. Skarnes^{3,4}

Biotechnology and Biological Sciences Research Council (BBSRC) Center for Genome Research, University of Edinburgh, King's Buildings, Edinburgh EH9 3JQ UK; ¹Exelixis Pharmaceuticals, Inc., Cambridge, Massachusetts 02139

Gene trapping in murine embryonic stem cells is a proven method for the simultaneous identification and mutation of genes in the mouse. Gene trap vectors are designed to detect insertions within genes through the production of a fusion mRNA transcript, making the identification of the endogenous gene possible by 5' rapid amplification of cDNA ends (RACE). Although the amplification of specific cDNAs can be achieved rapidly, cloning and screening of informative-sized cDNAs has proven to be time consuming. To eliminate the need for cloning, we have developed a method for solid-phase sequencing of 5' RACE products. More than 150 independent gene trap cell lines were analyzed, and sequence information was obtained for every line successfully amplified by RACE. With the vector used in this study, 40% of the cell lines were found to contain properly spliced gene trap events. The remaining lines were either spliced inefficiently or contained deletions of the vector. These results highlight the advantage of sequencing gene trap integrations before further characterization. This work now paves the way for large-scale gene trap screens in mice and should greatly accelerate the functional analysis of the mammalian genome.

The generation of mutations in murine embryonic stem (ES) cells using targeted and random approaches affords an opportunity to analyze the function of genes systematically in the mammalian genome. Gene trapping in ES cells provides a particularly useful method for the isolation of large numbers of random insertional mutations that are immediately accessible to molecular characterization (Gossler et al. 1989; Friedrich and Soriano 1991; Skarnes et al. 1992; von Melchner et al. 1992). Gene trap vectors are activated through the production of a reporter gene fusion transcript following insertions of the vector within endogenous transcription units. Target gene expression is tagged by the reporter, and a portion of the target gene may be cloned directly from the fusion transcript. The ability to conveniently isolate and characterize large numbers of gene trap insertions in ES cells and to

pass selected mutations to the germ line of mice potentially provides a powerful tool for the functional analysis of the mouse genome. Realization of this goal, however, will require the development of automated methods for the identification of target genes associated with gene trap insertions.

Identification of the endogenous genes associated with gene trap insertions has proven to be a major bottleneck, reflected by the fact that only a handful of genes have been cloned from gene trap cell lines (Skarnes et al. 1992, 1995; von Melchner et al. 1992; Chen et al. 1994; DeGregori et al. 1994; Gasca et al. 1995; Takeuchi et al. 1995; Chen et al. 1996; Forrester et al. 1996). A variety of methods based on 5' rapid amplification of cDNA ends (RACE; Frohman et al. 1988) have been used to amplify and clone endogenous cDNA sequences from the fusion transcript produced in gene trap cell lines (Skarnes et al. 1992; Takeuchi et al. 1995; Chen 1996). However, because of the need to clone desired products for each integration event, these approaches become impractical on a large scale. In this paper we describe a method for direct sequencing of 5' RACE reactions that enables high-throughput sequence analysis of gene trap integration events.

Present addresses: ²Universite de Nice, Unite Institut National de la Santé et de la Recherche Médicale (INSERM), Parc Valrose, Nice 06032, France; ³Department of Molecular and Cell Biology, University of California, Berkeley, California 94720
⁴Corresponding author.
E-MAIL skarnes@socrates.berkeley.edu; FAX (510) 643-6791.

RESULTS AND DISCUSSION

The main problem encountered with 5' RACE stems from the preferential amplification and subsequent cloning of short, uninformative cDNAs (J. Moss and W. Skarnes, unpubl.). To obtain high-quality sequence directly from 5' RACE reactions, it was essential to size-select products that extend well beyond the primer sites used for PCR amplification. In the absence of size selection, direct solid-phase sequencing (Hultman et al. 1989) of 5' RACE reactions was poor, presumably because of interference from superimposed anchor oligonucleotide sequence at the 5' ends of short products (data not shown). Methods for purifying cDNAs of desired length, such as gel electrophoresis or gel filtration, were considered to be impractical on a large scale. Instead, we chose microdialysis, a convenient method for processing many samples and suitable for adapting to a multiwell format.

The direct sequencing protocol was pioneered using previously characterized ES cell lines carrying insertions of the secretory trap vector pGT1.8TM in the *LAR* and *Sek* genes (Skarnes et al. 1995). 5' RACE (Fig. 1) was carried out from total RNA using microdialysis between each step to remove excess primer, nucleotides, and salts, and to size-select cDNAs >300 bp in size. In the second round of PCR, the 5' anchor primer was biotinylated to capture sense-strand cDNAs on streptavidin-coated beads. Immobilized cDNAs were cycle-sequenced using a ³²P-end-labeled primer near the splice acceptor site of the vector. Direct sequencing of 5' RACE products derived from insertions in *LAR* and *Sek* yielded 100–200 bases of unambiguous sequence that matched the sequence of RACE cDNAs cloned from these lines by conventional means (data not shown).

To test this method on a large scale, 5' RACE was performed on total RNA from 153 independent ES cell lines carrying insertions of pGT1.8TM. The efficiency of each RACE reaction was monitored by dot blot hybridization of second-round PCR products using *En-2* exon sequences from the vector as a probe (Fig 2). RACE products that hybridize with *En-2* exon sequences were obtained from 115 lines, whereas 38 lines failed to amplify. These nonamplifiable lines likely represent deletions of the vector because RACE products could be obtained using nested primers further 3' in the vector (data not shown). Several nonamplifiable lines were characterized further, and deletion breakpoints in the vector were mapped to various points downstream of the primer sites used in the original 5' RACE reactions (J. Brennan, B.J. Avery, and W.C. Skarnes, unpubl.).

The high proportion of cell lines that fail to amplify may be particular to the vector used in this study and to the mode of transfection, as it is known that electroporated DNA is prone to degradation prior to or at the time of integration in the genome (Niwa et al. 1993). Minimizing the distance between the splice acceptor and selectable marker or the use of retroviral-based vectors would be predicted to reduce the proportion of nonamplifiable lines.

Informative sequence was obtained for 115 lines amplified successfully by RACE (Table 1). A single sequence ladder corresponding to a properly spliced fusion mRNA transcript was observed for 57 lines, indicating the efficient use of the splice acceptor (Fig 3). For these lines an average of 115 bases of sequence could be read upstream of the splice site. Twenty-nine sequences corresponded to known genes, and among the set of novel sequences, 11 of 25 matched dbEST entries in the National Center for Biotechnology Information database.

Fourteen insertions were spliced properly but yielded two or more sequences 5' of the splice site. This could arise from alternative splicing or mixed colonies. An additional three lines were spliced correctly to a cryptic donor site within CD4 sequences of the vector. The most likely explanation for these events is the integration of, and splicing within, tandem copies of the vector.

Direct sequencing was found to work equally well with cell lines expressing a low level of fusion transcript as judged by Northern blot analysis of total RNA (Figs. 3 and 4, lanes 4 and 5). Typically, these cell lines express detectable levels of β -Gal activity in a subset of spontaneously differentiated cells but not in ES cells themselves. Such integrations currently represent the limit of recoverable gene trap events in genes expressed at low levels in ES cells. Whether this method will be successful for rare transcripts remains to be tested and may require the use of poly(A)⁺-selected mRNA or an additional round of PCR with a third nested primer.

En-2 intron sequence was detected in RACE products from 44 lines, indicating that the fusion mRNAs in these lines are spliced inefficiently. Direct sequence of these lines showed either intron sequence alone or intron sequence superimposed on another distinct sequence (data not shown). The presence of unspliced fusion transcripts was confirmed in these lines by Northern blot hybridization (Fig 4, lanes 8,9), whereas in contrast, intron-containing transcripts were not detected in lines that yielded a single, properly spliced sequence (Fig. 4, lanes 1–7). Whereas inefficient splicing of the

GENE TRAPPING FOR ID AND MUTATIONS IN MICE

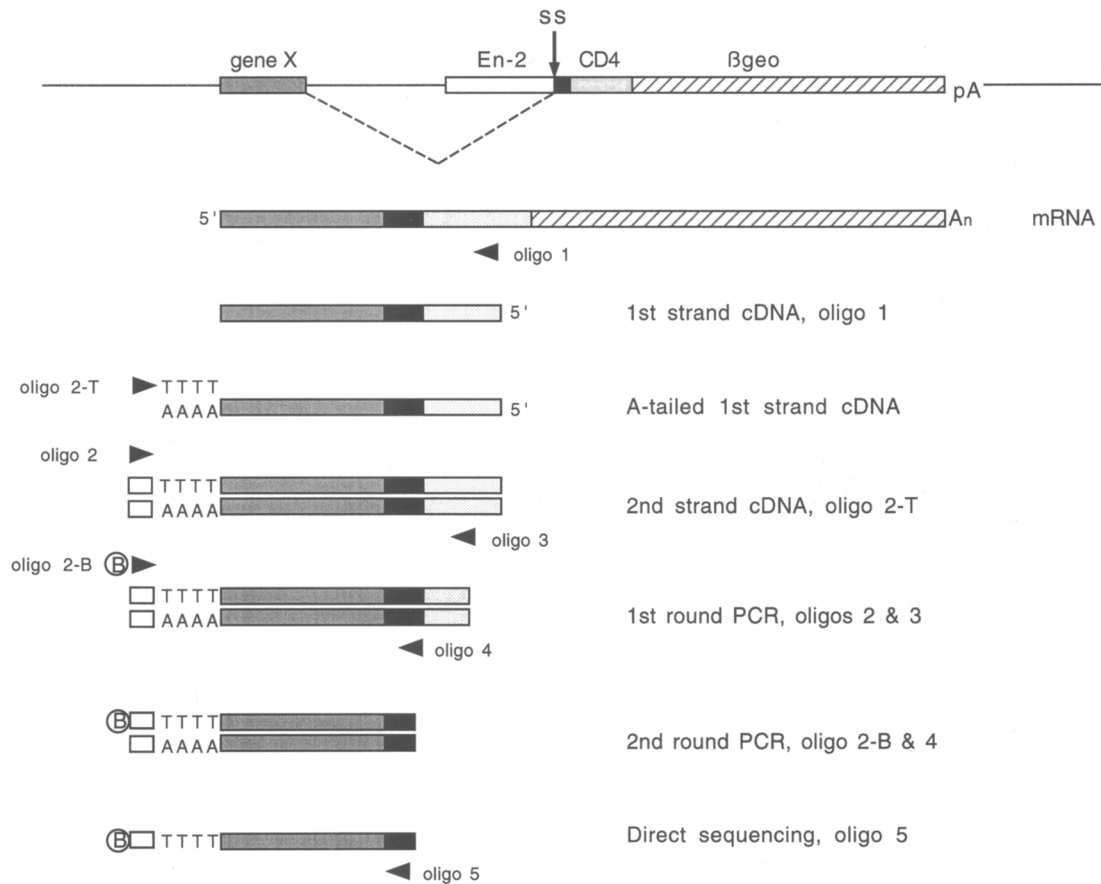


Figure 1 Protocol for direct sequencing of 5' RACE products. Insertion of the pGT1.8TM secretory trap vector in the intron of an endogenous gene is predicted to generate a β -*geo* fusion mRNA transcript through the use of the *En-2* splice acceptor (SA) contained in the vector (open box, *En-2* intron; solid box, *En-2* exon). cDNA prepared from fusion mRNA using oligonucleotide 1 (CD4 sequences, +364 bases from SA) is A-tailed, and second-strand cDNA is synthesized using a T-tailed anchor primer (oligonucleotide 2-T). Two rounds of PCR are then carried out using nested primers in CD4 and *En-2* exon (oligonucleotides 3 and 4; +281 and 130 bases from the SA, respectively) and the anchor primer (oligonucleotide 2). In the second round of PCR, a 5'-biotinylated anchor primer (oligonucleotide 2-B) is used. Biotinylated sense-strand cDNA is captured on streptavidin-coated beads and provides the template for cycle sequencing reactions using a ^{32}P -end-labeled primer in *En-2* exon sequence (oligonucleotide 5; +74 bases from SA).

vector to upstream exons may be a property of some endogenous loci, we also predict that integration of the vector in exons of genes or in non-polymerase II transcription units would generate intron-containing fusion transcripts. Further work will be required to define the genomic structure of insertions leading to inefficient splicing.

We have demonstrated the feasibility of direct sequencing of 5' RACE reactions as a rapid alternative to cloning and sequencing individual RACE cDNAs. On average, >100 bp of sequence was obtained from all lines amplified by RACE. The method worked equally well with lines that express a low level of fusion transcript, demonstrating the reliability and sensitivity of this pro-

cedure. The use of microdialysis and solid-phase sequencing readily lends itself to automation and is suitable for automated fluorescent sequencing. The high-throughput capability of this technique now enables the selection of desired mutations prior to germ-line transmission. This is particularly important in view of the large proportion of aberrant events that are associated with vector deletion or inefficient splicing.

In the near future all of the genes in the mammalian genome will be sequenced, and functional analysis of genes will take precedence over gene discovery. Mutant mice can be generated by gene trapping at a rate far exceeding that by conventional gene targeting as the need to build constructs for each target gene is

TOWNLEY ET AL.

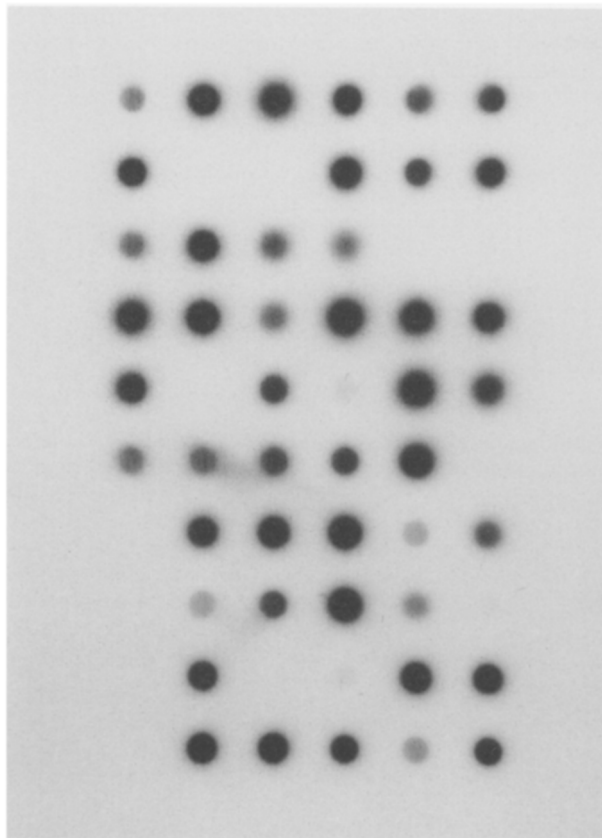


Figure 2 Monitoring the efficiency of 5' RACE. Dot blot analysis of second-round RACE products hybridized with the *En-2* exon probe (4-hr exposure).

eliminated. Theoretically, an unlimited number of gene trap cell lines can be isolated, partially sequenced, and stored indefinitely to create a permanent library of insertional mutations in ES cells. Large-scale gene trapping should thus provide an important future source of new mouse mutations. Such a resource of characterized mutations should complement and extend the current genome effort by providing both expression and phenotypic data for new

genes as they emerge as well as for previously cloned genes of unknown function.

NOTE: In an ongoing gene trap screen, >50 characterized mutations have been transmitted to the germ line of mice. Sequence and expression information for these lines is currently available at <http://socrates.berkeley.edu/~skarnes/resource.html>.

METHODS

5' RACE

CGR8.8 ES cells, a subline of the parental CGR8 cell line, were electroporated with the secretory gene trap vector pGT1.8TM, and G418-resistant colonies were picked and expanded as described previously (Skarnes et al. 1995). Total RNA from ES cell lines grown to confluence in 24-well plates was prepared using 0.25 ml of RNazol B (Tel-Test, Inc.) according to the manufacturer's instructions. An additional phenol/chloroform extraction prior to RNA precipitation was found to improve the quality of the final sequence. No improvement was observed when starting with poly(A)⁺ RNA. First-strand cDNA was reverse transcribed in a volume of 20 μ l, where 5 μ g of total RNA was annealed to 10 ng of oligonucleotide 1 (5'-CCAGAACCAGCAAAC-TGAAGGG; +364 bp from the splice site) and incubated for 1 hr at 37°C with Superscript II (GIBCO) in buffer supplied by the manufacturer. The reaction was stopped by adding one-tenth volume of 1 M NaOH and incubating at 65°C for 20 min, and then neutralized with one-tenth volume of 1 M HCl.

Following alkaline hydrolysis, RNA, salts, and excess primer were removed by microdialysis on 0.025- μ m nylon filters (Millipore, cat. no. VSWP 025 00) for 4 hr against TE buffer. In a volume of 25 μ l, first-strand cDNA was A-tailed for 5 min at 37°C in the presence of 0.2 mM ATP and 30 units of terminal deoxynucleotidyl transferase (GIBCO) in buffer provided by the manufacturer. The reactions were terminated by incubation for 2 min at 70°C. Half of the A-tailed cDNA was used directly for second-strand cDNA synthesis in a volume of 20 μ l of 1 \times restriction buffer M (Boehringer-Mannheim), containing 10 ng of a complementary T-tailed anchor primer [oligonucleotide 2-T; 5'-GGTTGTGAGCTCTCTAGATGGT₍₁₇₎], 2 units of Klenow enzyme (Boehringer Mannheim) and 0.5 mM dNTPs. The reaction was incubated for 30 min at room temperature, shifted to 37°C for 30 min, and then stopped by heating at 70°C for 5 min.

To size-select cDNAs, the reactions were spotted on 0.1 μ m nylon filters (Millipore, cat. no. VCWP 025 00) and microdialyzed

Table 1. Summary of 5' RACE Results

Total lines	Efficiently spliced		Inefficiently spliced		Nonamplifiable
	single sequence	multiple sequence ^a	intron alone	endogenous plus intron ^b	
153	57	14	23	21	38

^aSuperimposition of two or more sequences 5' of splice site.

^bSuperimposition of vector intron and endogenous gene sequences 5' of splice site.

GENE TRAPPING FOR ID AND MUTATIONS IN MICE

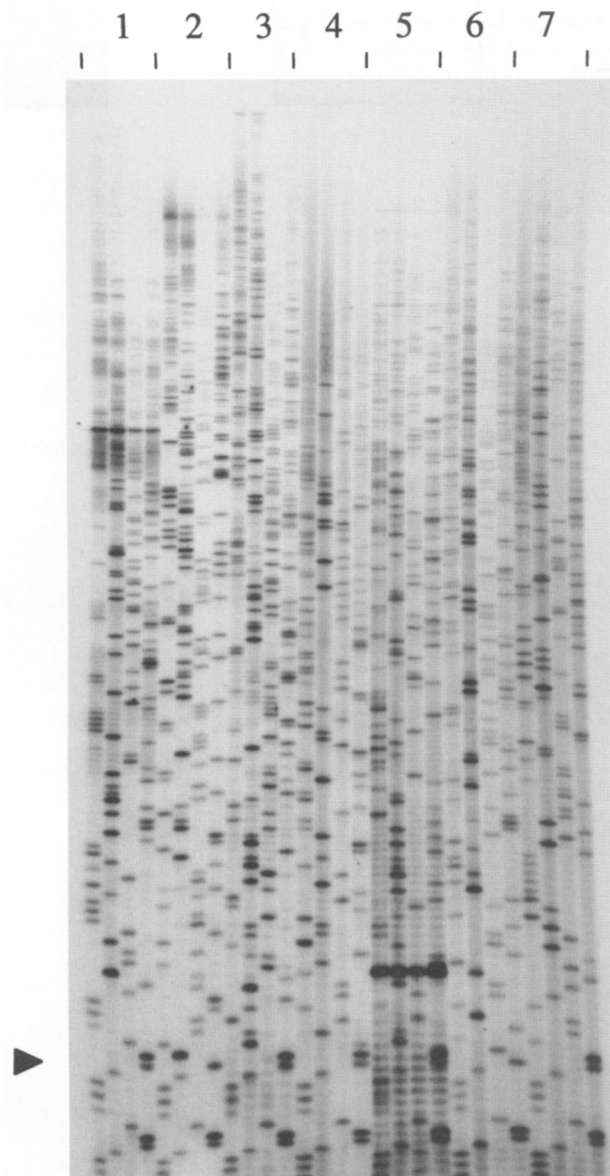


Figure 3 Direct sequence of 5' RACE products associated with insertions in known genes. Lanes 1–7 correspond to α -mannosidase II (GenBank accession no. X61172), adenosine kinase (U50196), APLP-2 (U15571), BMP-8 (U39545), laminin B1 (M15525), megalin (L34049), and K-glypican (X83577), respectively. The arrowhead shows the position of the splice site in vector. Sequencing reactions were loaded G, A, T, C.

against TE for 4 hr. Microdialysis filters of different pore sizes (0.025–0.2 μ m) were tested empirically using double-stranded DNA markers (1-kb ladder, GIBCO), and DNA fragments <300 bp were removed effectively on 0.1- μ m filters. Size-selected cDNAs were taken through two rounds of nested PCR amplification and microdialyzed on 0.1- μ m filters after each round. For the first round, the entire sample was amplified for 30 cycles (1.5 min at

94°C, 1.5 min at 60°C, 3 min at 72°C) using 5 units of AmpliTaq (Perkin Elmer), 100 ng of anchor oligonucleotide (oligonucleotide 2, 5'-GGTTGTGAGCTCTTCTAGATGG), 100 ng of oligonucleotide 3 (5'-AGTAGACTTCTGCACAGACACC; +281 bp from splice acceptor) in 50 μ l of 1 \times PCR buffer [100 mM KCl; 20 mM Tris (pH 8.3 at 37°C); 2 mM MgCl₂] containing 0.2 mM dNTPs.

Following microdialysis, one-eighth of the first-round products was added to the second-round PCR reaction containing 100 ng of biotinylated anchor (oligonucleotide 2-B; 5'-biotin-GGTTGTGAGCTCTTCTAGATGG) and 100 ng of oligonucleotide 4 (5'-TGCTCTGTCAGGTACCTGTTGG; +130 bp from the splice acceptor site) using conditions described for first-round PCR.

Dot Blot Analysis of 5' RACE Products

Three microliters of second-round RACE products was denatured for 30 min at 37°C in 0.1 ml of 0.4 M NaOH and transferred to Hybond N+ filters (Amersham) using a 96-well dot blot apparatus (Bio-Rad). Following transfer, the filters were probed with an α -³²P-labeled, random-primed DNA fragment containing 120 bp of *En-2* exon sequence immediately 3' of the splice site of pGT1.8TM. Hybridizations were performed at 55°C in buffer containing 0.5 M sodium phosphate (pH 7.2), 7% SDS, and 0.5% Marvel milk powder (Church and Gilbert 1984). Filters were washed three times at 50°C in 150 mM sodium phosphate containing 0.1% SDS and exposed to film for 2–4 hr.

Direct Solid-Phase Sequencing

5'-Biotinylated RACE products were bound to 200 μ g of streptavidin-coated magnetic beads (Dynabeads M-280) according to the manufacturer's instructions. Depending on the inten-

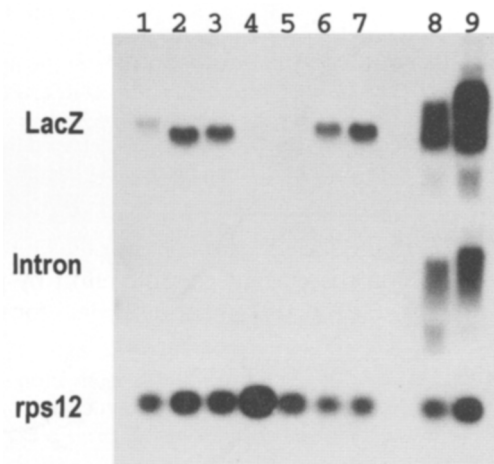


Figure 4 Northern blot analysis of ES cell lines carrying insertions of pGT1.8TM. In cases where direct sequencing yielded a single, properly spliced sequence (Fig. 3, lanes 1–7), a unique fusion transcript was detected using a *lacZ* probe (lanes 1–7). In lines that showed vector intron sequence by direct sequencing, RNA transcripts hybridizing to both *lacZ* and vector intron probes were detected (lanes 8,9). The ribosomal protein S12 (*rps12*) gene was used to control for loading.

TOWNLEY ET AL.

sity of the dot blot signal, between 5 and 40 μ l of dialyzed second round PCR products was used. Antisense strand cDNA (nonbiotinylated) was eluted in alkali, and sense-strand biotinylated cDNA bound to the beads was washed several times and resuspended in water. A total of 10.5 pmoles of sequencing primer (oligonucleotide 5; 5'-AGCAGTGAAGGCTGTGC, +74 bases from the splice acceptor) was end-labeled in a volume of 25 μ l containing 10 units of polynucleotide kinase (Boehringer Mannheim) and 50 μ Ci of [γ - 32 P]ATP (>5000 Ci/mole) in buffer supplied by the manufacturer. Cycle sequencing reactions (Amplicycle Sequencing Kit, Perkin Elmer) were carried out in the presence of 0.1 pmole of end-labeled primer as per the manufacturer's instructions (25 cycles of 95°C for 1 min, 60°C for 1 min, 72°C for 1 min).

RNA Analysis

Duplicate Northern blots were prepared by dry blot transfer to Hybond N membrane (Amersham) of ~25 μ g of total RNA purified with RNeasy B and electrophoresed on a 0.9% agarose gel containing 1 \times MOPS and 6.7% formaldehyde (Sambrook et al. 1989). Filters were UV cross-linked and probed with random primed, α - 32 P-labeled DNA fragments containing 2 kb of *lacZ* or 1.2 kb of *En-2* intron sequences from the pGT1.8TM vector. Hybridizations were carried out at 55°C in buffer containing 0.35 M sodium phosphate (pH 7.2), 7% SDS, and 30% formamide buffer containing 1% bovine serum albumin (Sigma) (Church and Gilbert 1984). The blots were washed three times at 50°C with 150 mM sodium phosphate buffer containing 0.1% SDS and exposed to film for 2 days. The filters were stripped and reprobed with a 0.5-kb probe containing the ribosomal protein S12 gene as a control for loading.

ACKNOWLEDGMENTS

We thank Julie Moss for her help in optimizing the RACE protocol and Tom Gardner for assisting with tissue culture and photographic work. This work was funded by the BBSRC Centre for Genome Research and by a grant from EXELIXIS Pharmaceuticals, Inc.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Chen, J., A. Nachabe, C. Scherer, P. Ganju, A. Reith, R. Bronson, and H.E. Ruley. 1996. Germ-line inactivation of the murine Eck receptor tyrosine kinase by gene trap retroviral insertion. *Oncogene* **12**: 979–988.
- Chen, Z. 1996. Simple modifications to increase specificity of the 5' RACE procedure. *Trends Genet.* **12**: 87–88.
- Chen, Z., G.A. Friedrich, and P. Soriano. 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.
- Church, G.M. and W. Gilbert. 1984. Genomic sequencing. *Proc. Natl. Acad. Sci.* **81**: 1991–1995.

DeGregori, J., A. Russ, H. von Melchner, H. Rayburn, P. Priyaranjan, N.A. Jenkins, N.G. Copeland, and H.E. Ruley. 1994. A murine homolog of the yeast *RNAI* gene is required for postimplantation development. *Genes & Dev.* **8**: 265–276.

Forrester, L.M., A. Nagy, M. Sam, A. Watt, L. Stevenson, A. Bernstein, A.L. Joyner, and W. Wurst. 1996. An induction screen in embryonic stem cells: Identification of genes that respond to retinoic acid in vitro. *Proc. Natl. Acad. Sci.* **93**: 1677–1682.

Friedrich, G. and P. Soriano. 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., M.K. Dush, and G.R. Martin. 1988. Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene specific oligonucleotide primer. *Proc. Natl. Acad. Sci.* **85**: 8998–9002.

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

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

Hultman, T., S. Stahl, and M. Uhlen. 1989. Direct solid phase sequencing of genomic and plasmid DNA using magnetic beads as solid support. *Nucleic Acids Res.* **17**: 4937–4946.

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

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. *Molecular cloning: A laboratory manual*, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Skarnes, W.C., B.A. Auerbach, and A.L. Joyner. 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**: 903–906.

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

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

von Melchner, H., J.V. DeGregori, H. Rayburn, S. Reddy, C. Friedel, and H.E. Ruley. 1992. Selective disruption of genes expressed in totipotent embryonic stem cells. *Genes & Dev.* **6**: 919–927.

Received October 10, 1996; accepted in revised form January 16, 1997.