EXPERIMENTAL CELL RESEARCH **244,** 125–136 (1998)
ARTICLE NO. EX984208

Efficient Gene Trap Screening for Novel Developmental Genes Using IRESβgeo Vector and *in Vitro* Preselection

Paolo Bonaldo,*,†,[1] Kamal Chowdhury,*,[1] Anastasia Stoykova,*,[1] Miguel Torres,*,‡,[1] and Peter Gruss*,[2]

**Department of Molecular Cell Biology, Max Planck Institute of Biophysical Chemistry, 37077 Göttingen, Germany; †Institute of Histology and Embryology, University of Padova, 35121 Padova, Italy; and ‡Departamento de Immunologia y Oncologia, Centro Nacional de Biotecnologia, Universidad Autonoma, Madrid 28049, Spain*

We have used different gene trap vectors and *in vitro* preselection of embryonic stem (ES) cells for a large scale screening of insertional mutations in developmentally regulated genes. A gene trap vector was constructed, which contains an internal ribosome entry site (IRES) upstream from a βgeo selectable-reporter fusion gene. Analysis of 801 independent integrations revealed that the IRESβgeo vector allows for a global enrichment of about 15 folds in the number of detectable gene trap events when compared with a conventional βgeo vector. Characterization of *in vitro* and *in vivo lacZ* expression suggested that this IRES-based vector is able to capture a wide range of genes expressed in a variety of tissues and developmental stages, and it can also allow trapping of genes expressed at very low levels in ES cells. A preselection protocol was devised, where gene-trapped ES cells were grown in the presence of specific growth/differentiation factors such as follistatin, nerve growth factor, and retinoic acid. Several gene trap integrations were found to be either activated or repressed by one of these factors. Characterization of *lacZ* expression during embryogenesis showed a strong enrichment of restricted patterns *in vivo* after ES cell preselection. These results suggest that a combination of IRESβgeo vector and *in vitro* preselection is more effective for the capture and mutation of a large number of developmental genes.

Key Words: **embryonic stem cells; gene trap; IRESβgeo vector; insertional mutagenesis; mouse development.**

INTRODUCTION

The understanding of mechanisms underlying mammalian embryonic development requires the identification and functional characterization of the genes controlling developmental events. A novel approach, called gene trap, was recently designed for the simultaneous identification and mutation of developmentally regulated genes in the mice [1–3]. This strategy is based on the transfection of murine embryonic stem (ES) cells with gene trap vectors containing a splice acceptor site upstream from a promoterless *lacZ* (β-galactosidase: β-gal) reporter gene. The *lacZ* gene can integrate randomly into the host genome, but it is activated only when a correct integration within a transcriptionally active endogenous gene has occurred. Due to transcriptional fusion between the reporter and a target gene, the *lacZ* expression closely resembles the spatial and temporal expression patterns of the target gene [3]. More importantly, gene trapping is mutagenic since the insertion events frequently interrupt the coding region of tagged genes [2, 3]. Therefore, mice can be bred to homozygosity to identify possible phenotypic alterations caused by the mutation of the interrupted gene. Cloning of the tagged gene is easily achieved by generation of cDNA from the *lacZ* fusion transcript using the quick 5′-RACE PCR method [4, 5]. Gene trapping has been successfully used by several groups and important functions for novel and known genes were described [6–11].

One major drawback of the gene trap approach is the necessity to generate a large number of mice from the corresponding ES cell clones to obtain few interesting genes. One way to circumvent this is offered by the possibility to preselect gene-trapped ES cell lines *in vitro* before generating the mice. This is especially interesting when searching for specific classes of genes [12, 13]. *In vitro* differentiation of ES cells can also be used for the selection of tagged genes that are expressed in one or more specific embryonic tissues [14]. We started a large-scale gene trap program aimed at the identification and mutation of a large number of developmentally regulated genes and were interested to devise some simple strategies to increase the efficiency of capture of developmental control genes. Toward this aim, different gene trap vectors were used and gene-trapped ES cells were screened *in vitro* according to several criteria. Currently used gene trap

[1] These authors contributed equally.
[2] To whom correspondence and reprint requests should be addressed. Fax: + 49 551 2011504.

0014-4827/98 $25.00

vectors are depending on fusion with an endogenous gene for the translation of the reporter gene. Although useful, this property has some drawbacks. Only those trapping events generating a transcriptional fusion between the reporter and the coding region of the tagged gene in the correct frame and orientation can be detected. Besides this, some translational fusions may lead to the inactivation of the reporter activity or may be targeted into subcellular sites where reporter activity is not easily accessible for detection. Therefore, we constructed a vector containing an internal ribosome entry site (IRES) from the encephalomyocarditis virus located between the splice acceptor and the reporter sequences [5]. In this vector, the IRES sequence allows for the cap-independent translation of the reporter gene from fusion transcripts [15]. We devised a simple and reproducible preselection protocol, where gene-trapped ES cell lines were screened for the β-gal staining patterns *in vitro* and tested for their responsiveness to specific growth/differentiation factors such as follistatin, nerve growth factor, and retinoic acid.

In this study we present the results obtained from the analysis of 801 gene-trapped lines generated with a conventional βgeo vector (pGT1.8geo) [12] and with the IRESβgeo vector. The global efficiency of IRESβgeo was about 15-fold higher than pGT1.8geo. Gene-trapped lines presented a broad range in the distribution and intensity of *lacZ* expression, and IRESβgeo allowed also the capture of genes expressed at very low levels in ES cells. Several genes were found to be specifically activated or repressed by retinoic acid, follistatin, or nerve growth factor. Generation of mice from trapped cell lines selected on the basis of their *in vitro lacZ* expression pattern and/or induction or repression by one of the above factors revealed a strong enrichment for lines with interesting restricted expression patterns during development. These observations provide evidence that a combination of the use of the IRESβgeo vector and an *in vitro* preselection protocol is more suitable for tagging genes at high efficiencies needed for large scale gene trapping.

MATERIALS AND METHODS

Gene trap vectors. The βgeo vector (pGT1.8geo, kindly provided by Dr. W. C. Skarnes) contains the splice acceptor sequence from the mouse *En-2* gene [1] joined in frame with the promoterless βgeo reporter-selectable marker gene [2], followed by the SV40 polyadenylation signal at the 3′-end (Fig. 1a). The βgeo gene of this vector does not contain the point mutation in the *neo* sequence known to reduce its enzymatic activity [12]. The bicistronic IRESβgeo vector (Fig. 1b) was obtained by introducing the internal ribosome entry site (IRES) from the encephalomyocarditis virus [16] between the splice acceptor and the βgeo sequences. This vector was generated as follows. An IRES-*lacZ* fusion plasmid was first constructed as described in detail by Kim *et al.* [16]. A fragment containing the entire IRES and part of the fused *lacZ* sequence was then excised from the IRES-*lacZ* plasmid by *Eco*RV digestion and cloned into the *Bgl*II (blunt ended)/*Eco*RV-cleaved βgeo vector. In this configuration, the translation initiation site for βgeo is provided by the IRES sequence, which allows independent translation of βgeo from gene trap fusion transcripts. Both vectors were linearized at the single *Sca*I site before electroporation.

ES cell culture. R1 ES cells [17] were routinely cultured on a feeder layer of mitomycin C-treated primary embryonic fibroblasts in ES medium. The ES culture medium consisted of Dulbecco's modified Eagle medium (DMEM) supplemented with 0.1 mM MEM nonessential amino acids, 1 mM sodium pyruvate, 2 mM L-glutamine, 0.1 mM 2-mercaptoethanol, 1000 U/ml leukemia inhibitory factor (LIF; ESGRO), and 20% heat-inactivated fetal calf serum (FCS; Gibco-BRL).

Electroporation and selection. ES cells were trypsinized and resuspended in phosphate-buffered saline (PBS). In a typical experiment, 10^7 ES cells were electroporated with 30 μg linearized IRESβgeo or pGT1.8geo vector DNA in 1 ml PBS, by applying a single pulse at 250 V, 500 μF in a Bio-Rad Gene Pulser. After electroporation, cells were seeded onto a monolayer of G418-resistant primary embryonic fibroblasts at 3×10^6 cells/100-mm dish. Selection with G418 (Gibco-BRL) at 250 μg/ml was started 24 h after electroporation. After 7–10 days of selection, single G418-resistant colonies were picked under a microscope, trypsinized to disaggregate the cells, and plated into 96-well dishes containing feeder fibroblasts and 200 μg/ml G418. Cells were cultured in the 96-well dishes for 2–4 days, and the dishes were then processed for freezing of cells at −80°C and for X-gal staining.

X-gal staining. G418-resistant ES cell lines obtained after electroporation were screened for β-gal activity in the 96-well dishes. Positive lines were then thawed, expanded, and their *lacZ* expression pattern *in vitro* was analyzed again in 35-mm dishes. In some experiments, X-gal staining was also determined using partially differentiated cells obtained by growing the ES cell lines on gelatin-coated dishes without feeder fibroblasts and in the absence of LIF.

Detection of *lacZ*-expressing ES cell lines was done by histochemical staining for β-gal activity as previously described [18]. Briefly, dishes were rinsed with PBS and fixed with 0.2% glutaraldehyde for 3–5 min at room temperature. After being washed three times for 5 min with a solution containing 0.01% sodium deoxycholate and 0.02% Nonidet P-40, the cells were stained for 4 to 48 h at 37°C with a reaction mixture containing 10 mM potassium ferrocyanide, 10 mM potassium ferricyanide, and 0.5 mg/ml 4-chloro-5-bromo-3-indolyl-D-galactopyranoside (X-gal). After staining, dishes were rinsed with PBS and carefully analyzed under bright-field and phase contrast microscope.

Screening of gene trap lines responsive to soluble factors. To identify gene trap lines that were activated or repressed by specific soluble factors, we applied the screening procedure outlined in Fig. 4. After electroporation, single G418-resistant colonies were picked and grown into duplicate 96-well dishes containing feeder fibroblasts and 200 μg/ml G418. One dish (master plate) was used for freezing an aliquot of each cell line at −80°C. The other dish (screening plate) was further cultured for the screening of cells as follows. Cells from the screening plate were trypsinized, splitted into five different 96-well dishes without feeder fibroblasts, and cultured with normal ES medium. After 24 h, when small colonies were apparent, medium was removed and the dishes were cultured under different conditions, using ES medium without LIF. The following five conditions were used: (a) ES medium with 20% FCS (control dish); (b) ES medium with 20% FCS and 150 ng/ml follistatin (follistatin dish); (c) ES medium with 1% FCS (low serum control dish); (d) ES medium with 1% FCS and 100 ng/ml nerve growth factor (NGF dish); (e) ES medium with 1% FCS and 0.2 μM all-*trans*-retinoic acid (RA dish). Cells were additionally cultured for 48 h in the above growth conditions and then processed for X-gal staining. Staining was carefully compared for each cell line in the five different dishes in order to detect gene trap lines that responded to one (or more) factor(s). Selected cell lines were then thawed, expanded to 35-mm plates, and

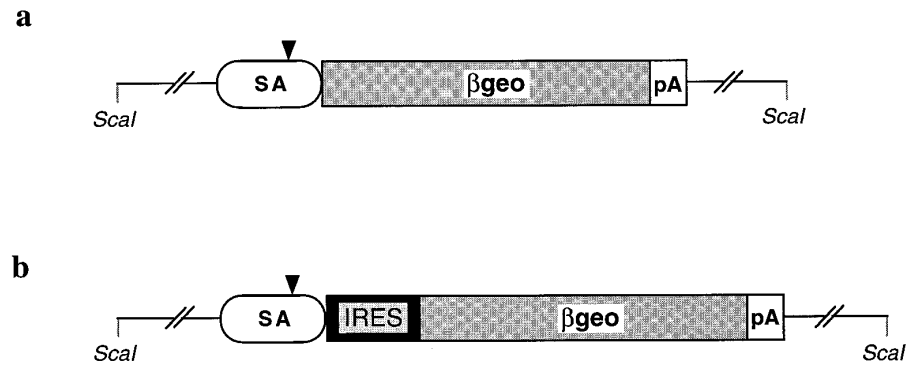


FIG. 1. Schematic diagram of the gene trap vectors pGT1.8βgeo (a) and IRESβgeo (b) used in this study. Relevant regions are depicted as boxes, plasmid backbone is represented by a plain line. SA, a 1.3-kb fragment containing a splice acceptor site (arrowhead) from the mouse *En-2* gene; βgeo, a 3.9-kb fragment consisting of the *E. coli lacZ* reporter gene fused in frame to the wild type *E. coli* neomycin phosphotransferase (*neo*) gene; pA, a 0.2-kb fragment containing the SV40 polyadenylation signal; IRES, a 0.6-kb fragment containing the internal ribosome entry site from the encephalomyocarditis virus. The *Sca*I site used for linearizing the vectors before electroporation is shown.

screened again with the above protocol to confirm their response to the factors.

Generation and analysis of mice. Selected ES cell lines were used for generating mouse chimeras by morula aggregation [19]. Analysis of the *in vivo* expression patterns of tagged genes was determined by whole mount X-gal staining [20], using mice that were heterozygous for the gene trap insertion.

RESULTS

Characterization of lacZ Expression in ES Cell Lines Obtained with the βgeo (pGT1.8geo) Vector

In our large-scale gene trap screening, we used different vectors and screened the tagged ES cell lines *in vitro* for the efficient detection of interesting genes. We first used a βgeo vector (pGT1.8geo, Fig. 1a) and tested its potential for generating efficiently *lacZ*-expressing ES cell lines. The βgeo gene encodes a protein containing both β-gal and G418-resistance activities and it serves as a selection-reporter marker [2]. After electroporation of ES cells and integration into a transcriptional unit, the βgeo gene can become activated by a host promoter. A transcriptional fusion mediated by the flanking splice acceptor sequence is necessary for the effective reporter and selection activity. This transcript is then translated into a fusion product between the βgeo protein and part of the endogenous protein. When the localization signals of the endogenous protein are retained after fusion, the βgeo is subcellularly targeted according to the endogenous protein. Therefore, a first information about the tagged gene can be obtained from the subcellular localization of β-gal activity. A detailed analysis of the subcellular localization of β-gal staining was performed in 117 *lacZ*-expressing ES cell lines obtained with pGT1.8geo. Six main subcellular patterns of β-gal staining were observed (Table 1). Some examples of the different subcellular staining patterns in gene-trapped ES cell lines are shown in Fig. 2.

The βgeo gene is depending on the expression of the endogenous gene for the production of *neo* and *lacZ* activities, therefore trapping with a βgeo vector should theoretically select for genes that are expressed in ES cells and the proportion of β-gal-positive colonies among the G418-resistant ones should be close to 100% [2]. Using the pGT1.8geo vector, we observed a much lower proportion of β-gal-positive colonies, in the range of 24 to 37% of total G418-resistant colonies (Table 2). There was a broad range in the distribution and intensity of β-gal staining among the different gene-trapped cell lines, suggesting that the vector is able to capture genes with different levels of expression in ES cells. The detailed analysis of *in vivo* distribution of *lacZ* expression patterns in several gene-trapped lines suggests that this βgeo vector is able to capture both

TABLE 1

Subcellular β-gal Localization Patterns in Cell Lines Obtained with pGT1.8geo

Diffuse[a]	49 (42%)
Cytoplasmic[b]	12 (10%)
Dot[c]	10 (8%)
Cytoplasmic + dot[d]	29 (25%)
Cell surface[e]	6 (5%)
Nuclear[f]	11 (9%)

Note. Frequencies were calculated on the total number of β-gal-positive cell lines.

[a] β-gal staining was found throughout the cell.
[b] β-gal staining was evenly distributed in the cytoplasm.
[c] β-gal staining was restricted to single cytoplasmic dots.
[d] β-gal staining was present both in cytoplasm and dots.
[e] β-gal staining was found at the periphery of the cell.
[f] β-gal staining was present only in the nucleus.

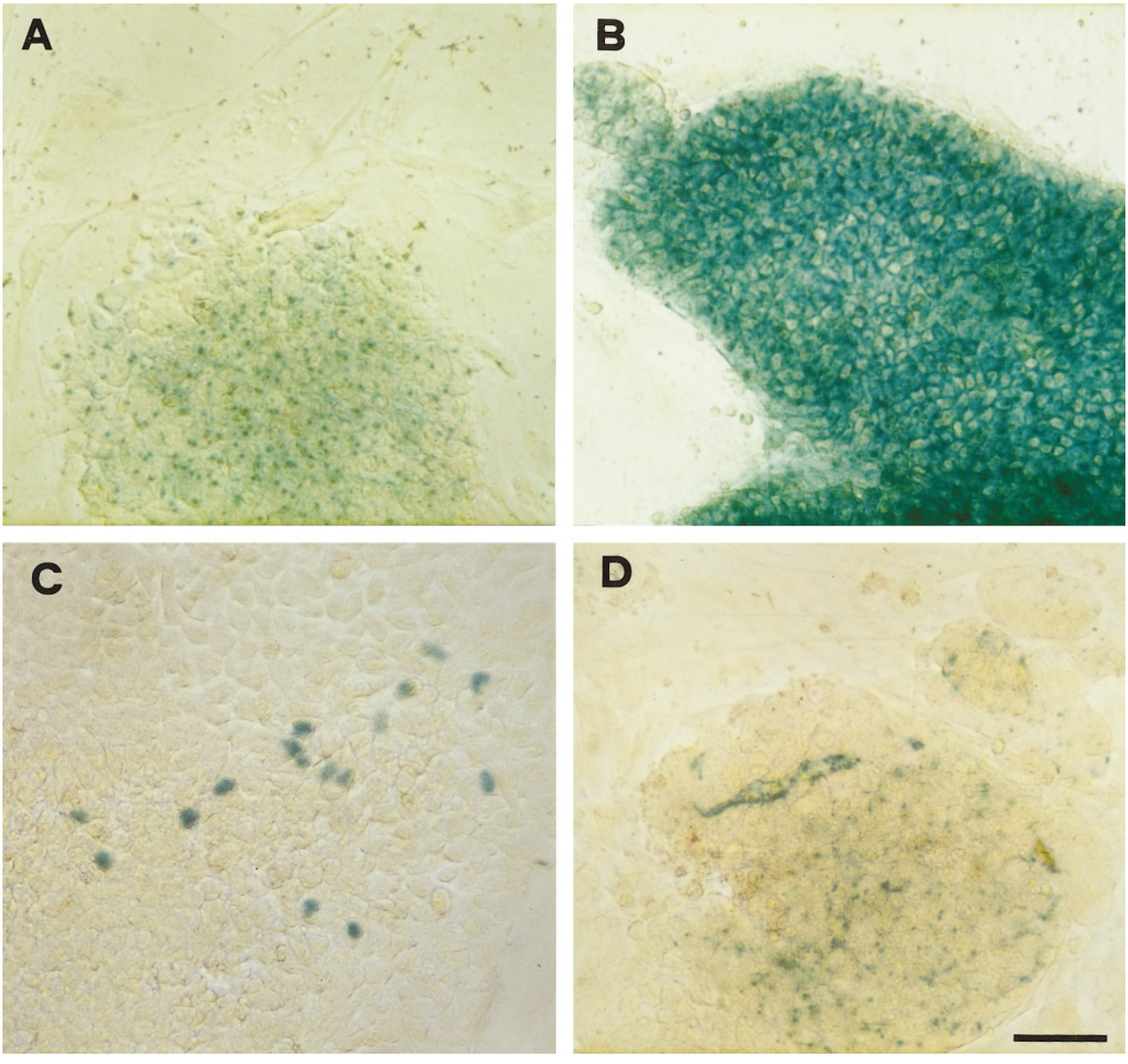


FIG. 2. Subcellular localization of β-gal staining in gene trap ES-cell lines obtained with βgeo vector. (A) Line XIII-4, staining was restricted to distinct cytoplasmic dots. (B) Line XIII-23, a diffuse β-gal staining was detected in the cytoplasm, but not in the nucleus. (C) Line X-3, β-gal was detected in the cell nuclei. (D) Line VII-45, staining was localized close to the cell surface; the gene tagged in this line coded for β-E-catenin [10], a component of the membrane-associated cadherin adhesion complex. Scale bar, 100 μm.

ubiquitously expressed genes as well as genes whose expression is restricted to a wide range of tissues and embryonic stages [20].

Construction of an IRESβgeo Vector and Analysis of Its Trapping Efficiencies

Although the βgeo vector works efficiently as shown above, it has several limitations due to the requirement of translational fusions with endogenous tagged sequences and only some insertions will be fully productive. Insertions of the βgeo sequences in the wrong frame or orientation or fusion to the 5′- or 3′-untranslated regions of the host transcript will not be detected. Moreover, a partial or complete loss of reporter and/or selection activity might occur when βgeo protein is fused to certain endogenous protein sequences and some classes of genes might be absent or under-represented. For instance, it has been shown that β-gal activity is lost when βgeo is fused to a signal peptide, therefore βgeo fusions with secreted molecules cannot be detected [12].

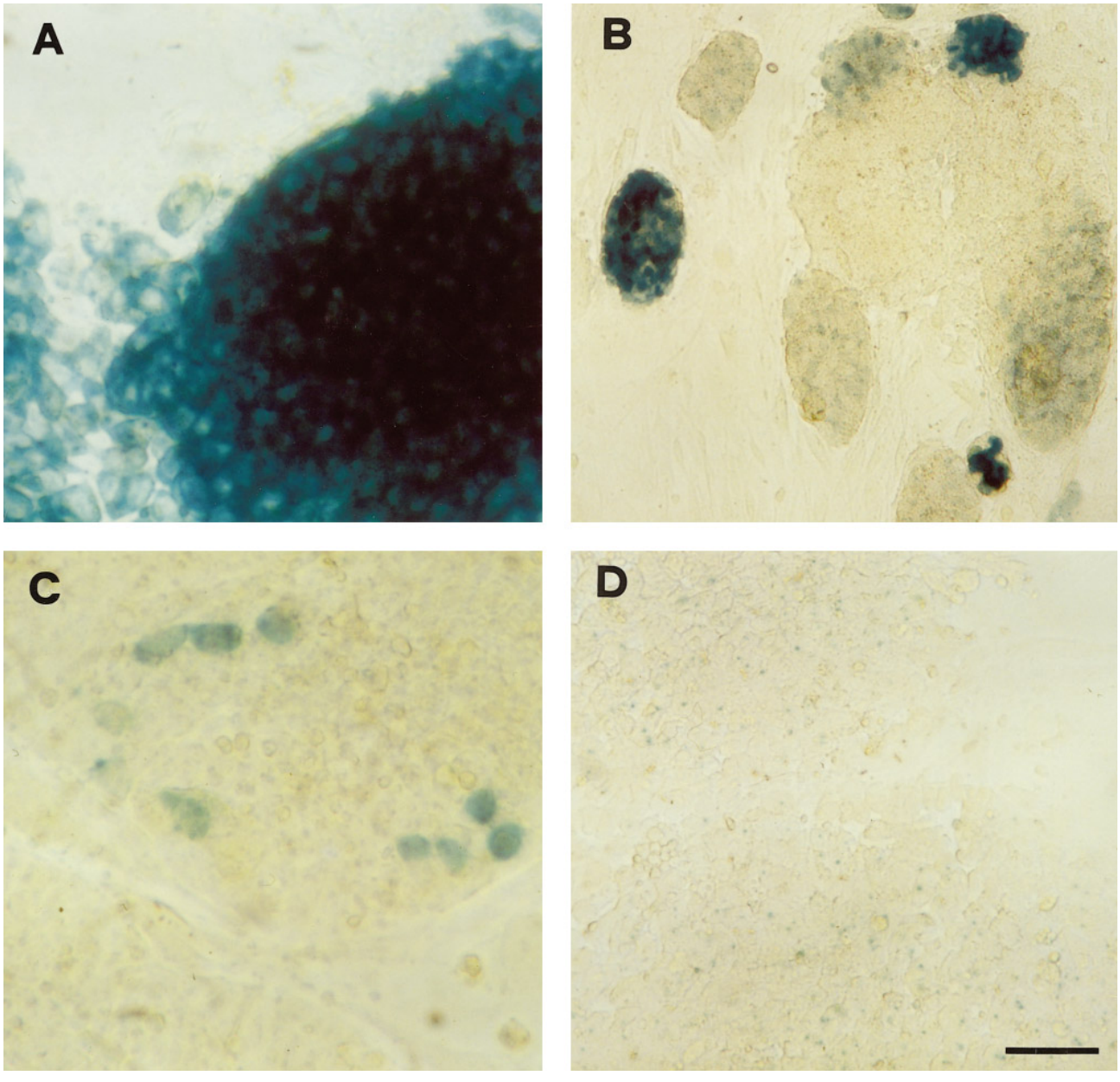


FIG. 3. Examples of gene trap ES-cell lines with different percentages of stained and unstained cells. (A) Line XV-56, strong staining was present in all the cells. (B) Line X-82, *lacZ* expression was differently regulated with stronger staining in small undifferentiated colonies and weaker or absent in larger colonies containing differentiated cells. (C) Line XVI-83, only about 1% of the cells expressed *lacZ*. (D) Line XVI-91, very low levels of *lacZ* expression were revealed as faint dots, barely detectable even after extensive staining. Gene trap cell lines were obtained with βgeo (A, B) and IRESβgeo (C, D) vectors. Cells were stained with X-gal for 8–12 h (A–C) or 48 h (D) and viewed under a phase contrast microscope. Scale bar, 100 μm (A, C, D) and 30 μm (B).

In order to overcome these limitations, we constructed an IRESβgeo vector containing an internal ribosome entry site (IRES) between the splice acceptor and the βgeo sequences (Fig. 1b). After trapping with this vector, the IRES sequence should allow for the cap-independent translation of βgeo from fusion transcripts with host sequences, without the above constraints. As shown in Table 2, the average number of G418-resistant colonies we obtained with IRESβgeo vector was about sixfold higher than that obtained with pGT1.8geo vector. The proportion of β-gal-positive colonies was also increased by 2.5-folds, since about 80% of the G418-resistant colonies obtained with IRESβgeo expressed β-gal activity (Table 2). As expected, subcellular localization of β-gal staining was similar for all the β-gal-positive cell lines obtained with

TABLE 2

Gene Trap Efficiencies of pGT1.8geo and IRESβgeo Vectors

Experiment	Vector	No. of cells[a]	G418^{R}[b]	β-gal +[c]
1	pGT1.8geo	$8 \cdot 10^6$	15	5 (33%)
2	pGT1.8geo	10^8	131	36 (30%)
3	pGT1.8geo	$9 \cdot 10^7$	84	20 (24%)
4	pGT1.8geo	$8 \cdot 10^7$	89	26 (30%)
5	pGT1.8geo	$6 \cdot 10^7$	90	33 (37%)
6	pGT1.8geo	10^7	21	6 (29%)
7	IRESβgeo	10^7	39[d]	28 (72%)
8	IRESβgeo	$8 \cdot 10^7$	172[d]	157 (91%)
9	IRESβgeo	10^7	77	55 (71%)
10	IRESβgeo	10^7	83	64 (77%)

[a] Number of ES cells used for each electroporation experiment.
[b] Number of G418-resistant colonies obtained on each experiment. The average number of G418^{R} colonies/10^7 electroporated cells was 12 (pGT1.8geo) and 80 (IRESβgeo).
[c] Number of β-gal-positive colonies detected on each experiment. The proportion of β-gal-positive colonies among the total G418-resistant colonies is shown in parenthesis. The mean frequencies of β-gal-positive colonies were 29% (pGT1.8geo) and 82% (IRESβgeo).
[d] Only some G418-resistant colonies were picked in these two experiments (remaining colonies were not included in the counting).

TABLE 3

Distribution of *lacZ* Expression in Gene-Trapped Cell Lines

In vitro lacZ expression[a]	βgeo[b]	IRESβgeo[b]
Ubiquitous	18 (14%)	43 (15%)
Widespread	25 (20%)	22 (8%)
Ubiquitous/widespread, regulated	n.d.[c]	96 (34%)
Restricted	34 (27%)	51 (18%)
Very restricted	50 (39%)	69 (24%)
Faint (dot) expression[d]	n.d.[c]	68 (24%)

[a] β-gal staining patterns of trapped cell lines were classified into five different groups as follows: ubiquitous (all cells in the dishes were positive, Fig. 3A), widespread (60–95% of the cells were positive), regulated (low levels of β-gal staining in most cells and much stronger staining in few cells, Fig. 3B), restricted (20–50% of the cells were positive), very restricted (1–15% of the cells were positive, Fig. 3C).
[b] Frequencies of lines for each group were calculated on the total of β-gal-positive lines (127 lines for pGT1.8geo and 281 lines for IRESβgeo).
[d] *lacZ* expression was only detected after 48 h staining as a faint dot present in a variable proportion of cells (Fig. 3D).
[c] n.d., not determined.

IRESβgeo. In agreement with the results obtained with pGT1.8geo vector, we observed different temporal and spatial patterns of *lacZ* activation *in vivo* during embryogenesis using gene-trapped cell lines generated with the IRESβgeo [20].

Theoretically, every G418-resistant cell clone should be positive for β-gal activity when the IRESβgeo vector is used. The lower frequencies of β-gal-positive cells we observed might be due to the capture of genes with weak promoter activities so that *lacZ* expression cannot be detected. This is also suggested by the very faint and sometimes barely detectable dot-like expression found in several gene-trapped cell lines only after extensive β-gal staining (Fig. 3D and Table 3). When we introduced *in vivo* some of the β-gal-negative cell lines and checked their *lacZ* expression during embryogenesis, we found that most lines were still negative but *lacZ* expression could be detected in few cases (see also Table 4). These observations suggest that the IRESβgeo vector works efficiently and it can also allow capture of genes expressed at very low levels in ES cells.

Distribution of lacZ Expression Patterns in Gene-Trapped Cell Lines

Analysis of β-gal staining in cell lines obtained with IRESβgeo and pGT1.8geo vectors revealed other features (Table 3). Ubiquitous distribution of β-gal staining throughout all the cells (Fig. 3A) was found only in a limited number of lines, while for the most part of the lines β-gal staining was restricted to a subset of cells (Figs. 3B and 3C). The proportion of β-gal-positive cells was variable: from 100% to less than 1% of the cells of the dish could be stained, depending on the specific tagged cell line. Several lines showed a dynamically regulated *lacZ* expression, with low levels of β-gal staining in most cells and a much stronger staining in few cells (Fig. 3B). When the β-gal staining was performed using par-

TABLE 4

Comparison of *lacZ* Expression Patterns in Trapped Cell Lines and in Mice

		Expression *in vivo* (9.5–14.5 dpc)[b]		
Expression *in vitro*[a]		Ubiquitous	Restricted	Negative
Ubiquitous/widespread	20	14	1	4
Ubiquitous/widespread, regulated	14	6	7	1
Restricted	13	2	8	3
Very restricted	7	0	3	4
Faint (dot) expression	7	1	5	1
Negative	12	0	1	11

[a] For the classification of β-gal staining patterns *in vitro*, see Table 3.
[b] β-gal staining was checked using mouse embryos heterozygous for the gene trap insertion. At least three different developmental stages were stained for each line: 9.5/10.5, 11.5/12.5, 13.5/14.5 dpc (days post coitum). Ubiquitous: most or all tissues were positive. Restricted: staining was restricted to specific tissues or organs in all the stages. Negative: no b-gal activity was detected in any of the stages.

tially differentiated colonies, *lacZ* expression was found to be confined either to differentiated cells or to undifferentiated ES cells in certain lines. In few distinct cases, *lacZ* expression was found exclusively in some specific cell types like fibroblasts or neuron-like cells, based on the gross cellular morphology. As shown in Table 3, a restricted or regulated pattern of *lacZ* expression *in vitro* was detected altogether in about 70% of the β-gal-positive lines.

We generated mice from gene-trapped cell lines of each group according to their *in vitro* β-gal staining (ubiquitous, regulated, restricted, very restricted) and analyzed their *lacZ* expression *in vivo* during midgestation. This analysis revealed a correlation between the *lacZ* expression in the cells and in mice (Table 4). Fourteen (70%) of the 20 lines with an ubiquitous or widespread *lacZ* expression *in vitro* revealed an ubiquitous expression *in vivo.* On the contrary, restricted expression during embryogenesis was found for 7 (50%) of the 14 lines with regulated expression *in vitro* and for 8 (62%) of the 13 lines with a restricted expression *in vitro.* Of the 7 lines with a very restricted expression *in vitro,* 3 had a restricted pattern and 4 were negative during embryogenesis; interestingly, ubiquitous expression *in vivo* was never detected in lines showing a very restricted expression in the cells. We also generated mice from 7 cell lines obtained with IRESβgeo vector and having a faint dot-like *lacZ* expression in ES cells (Fig. 3D): five (71%) of these lines showed clearly restricted expression during embryogenesis (Table 4), suggesting that the corresponding tagged genes are activated later during development.

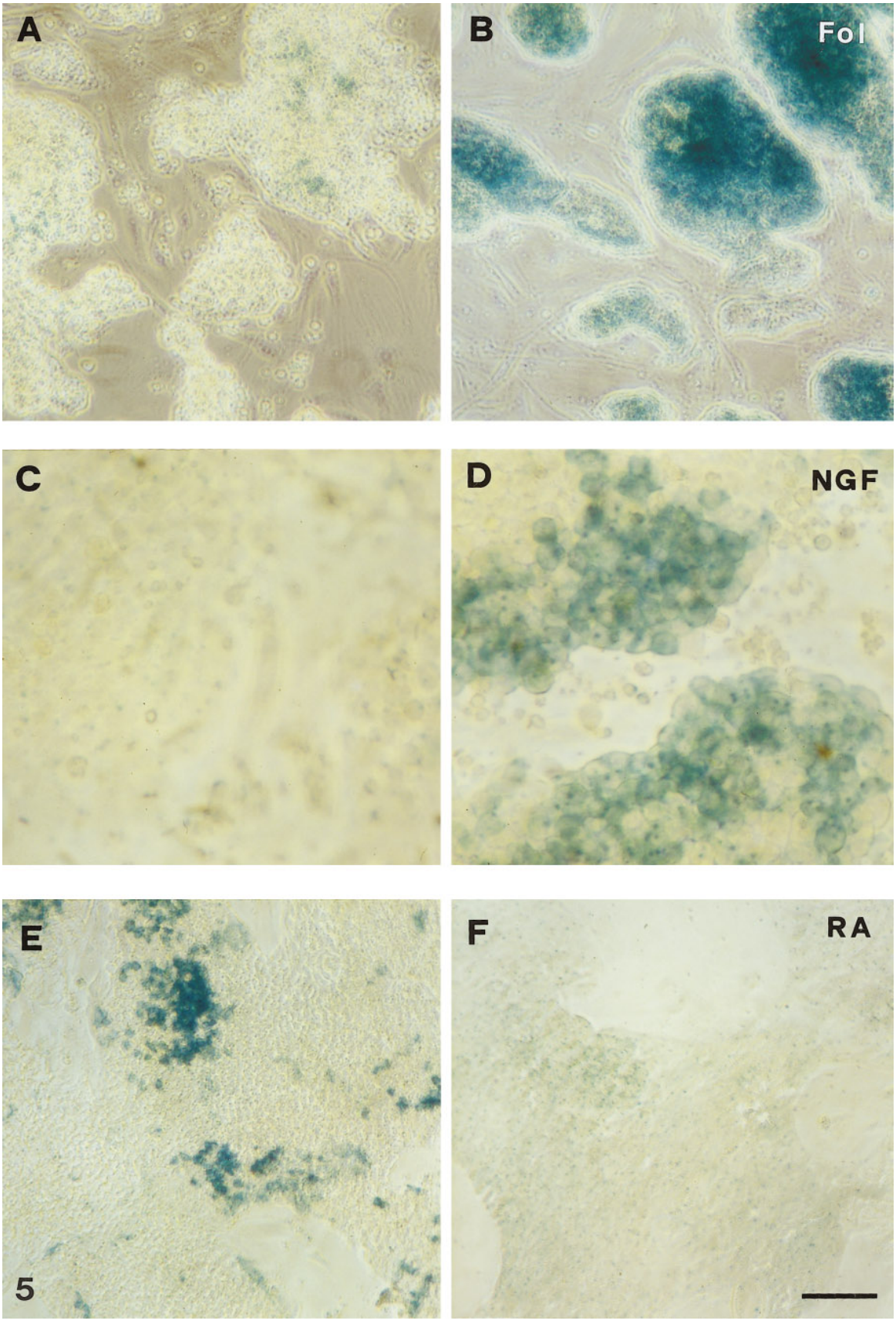
In Vitro Preselection of ES Cells Can Enrich for Lines with Restricted in Vivo Expression Patterns

In order to enrich for genes that are expressed in a restricted manner during development, especially in the nervous system, we devised a simple *in vitro* preselection protocol of gene-trapped cell lines using specific growth and differentiation factors. This protocol allows the detection of trapped lines where *lacZ* expression is activated or repressed by a certain factor. The procedure is outlined in Fig. 4. The following three soluble factors were used separately: follistatin, nerve growth factor (NGF), and retinoic acid (RA). To avoid the presence of additional factors that might interfere with the interpretation of results, a low serum concentration was needed when ES cells were cultured in the presence of NGF or RA. We tested the growth and survival of ES cells cultured in medium containing different concentrations (from 20 to 0.5%) of serum and found that a serum concentration as low as 1% was still able to support the survival of ES cells at a good rate for the duration of the experiment (data not shown). Since follistatin has been demonstrated to need the presence of activin for its activity [21], ES cells were cultured in the presence of normal serum concentration (20%) when follistatin was added. Therefore, each tagged ES cell line was cultured in five different conditions: (a) 20% serum (control), (b) 20% serum and follistatin, (c) 1% serum (control), (d) 1% serum and NGF, (e) 1% serum and RA. In order to detect only those genes exhibiting an early response to the factors, cells were kept under the above conditions for a relatively short time (24 h), and then stained for β-gal activity. Staining of each cell line in the different conditions was carefully compared, and those lines revealing activation or repression of *lacZ* expression by one (or more) factors were identified (Table 5 and Figs. 5A–5F). Altogether, 262 G418-resistant lines were screened with follistatin and NGF and 157 lines were screened with RA. Activation or repression by RA was found in 22 (14%) lines, while the proportion of lines where *lacZ* expression was activated or repressed by follistatin or NGF was much lower (1.5–3%).

Gene-trapped cell lines tested with this protocol were selected for the generation of heterozygous mice. Only the lines showing induction or repression by one factor and those with a regulated or a faint *lacZ* expression *in vitro,* were selected and introduced *in vivo.* Analysis of the *lacZ* expression during embryogenesis (9.5–14.5 dpc) revealed that there was a strong enrichment for restricted patterns after preselection of the gene-trapped cell lines, with a corresponding decrease of the ubiquitous and negative patterns (Table 6). Either 50 or 25% of the lines had a restricted embryonic *lacZ* expression after *in vitro* preselection or when no selection was applied, respectively. The proportion of lines showing restricted or prominent *lacZ* expression in the developing nervous system was also clearly increased: almost half of the lines obtained after *in vitro* preselection had a prominent expression in the nervous system, but this expression pattern was found only in one-fourth of the lines when no selection was applied for the cells.

DISCUSSION

We have used a IRESβgeo vector and *in vitro* preselection of ES cells for large-scale gene trap search of developmentally regulated genes. Eight hundred ES cell lines were generated with either IRESβgeo or a βgeo (pGT1.8geo) vector. We screened these lines for *lacZ* expression *in vitro* and compared the relative efficiencies of the two types of vectors. The data presented here provide evidence that IRESβgeo is able to tag and mutate genes at higher efficiencies than the


A
B
Fol
C
D
NGF
E
F
RA
5

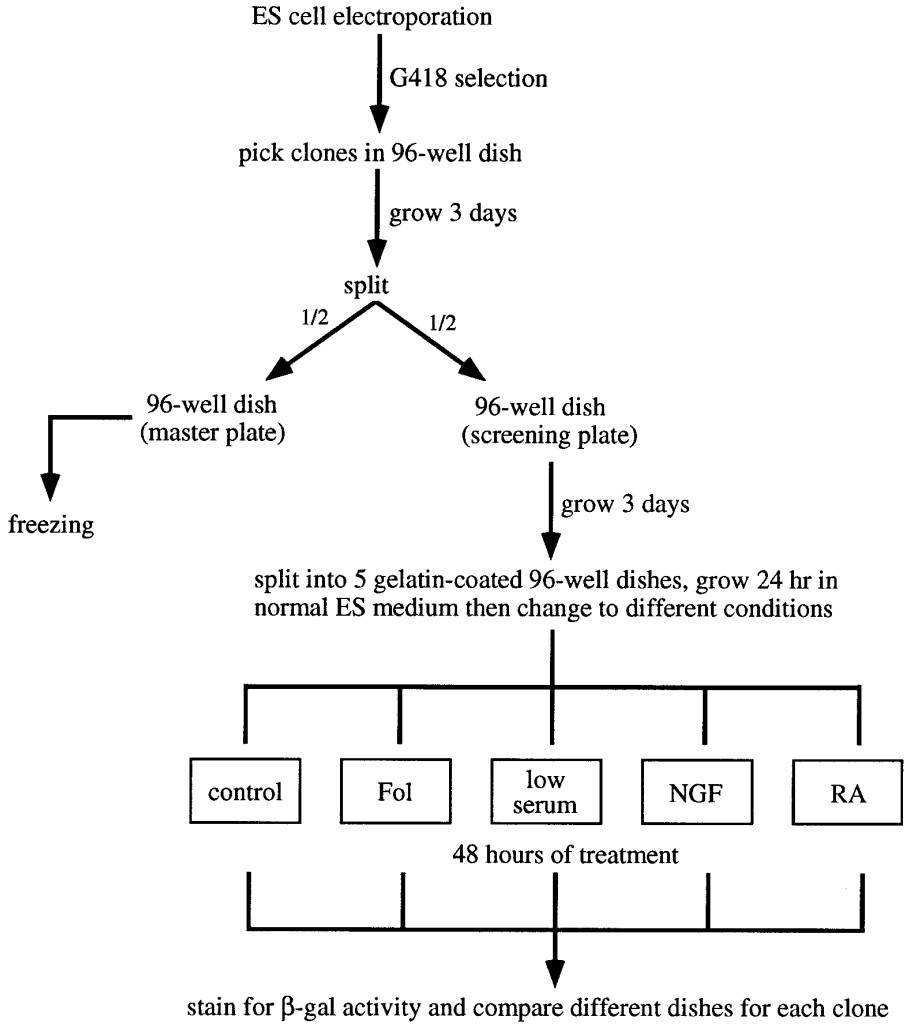


FIG. 4. Schematic diagram of the screening protocol of gene-trapped ES cell lines with soluble factors (see the Materials and Methods for explanations).

commonly used βgeo vectors. The use of pGT1.8geo for gene trap was already described [12]. The results we have obtained with this vector are in good agreement with those reported earlier. Skarnes *et al.* demonstrated that pGT1.8geo produces a broader range of staining intensities and greater proportion of β-gal-negative colonies than another βgeo vector (pSAβgeo). They found that the *neo* sequence of pSAβgeo, but not pGT1.8geo, has a mutation that reduces β-gal activity and therefore selects for highly expressed genes [12]. Using pGT1.8geo we found a large variety in the intensity, distribution, and subcellular localization of β-gal staining, therefore confirming that this vector was working properly in our screen.

The internal ribosome entry site (IRES) from the encephalomyocarditis virus has been widely used for the construction of bicistronic vectors and cap-independent translation of reporter or selectable marker genes. The activity of this IRES element in transgenic mice has been previously investigated and it has been dem-

FIG. 5. *lacZ* expression in gene trap cell lines specifically activated or repressed by follistatin, nerve growth factor (NGF), or retinoic acid (RA). Cells were screened with the protocol described in the legend of Fig. 4 and then stained with X-gal for 8 h. (A, B) Line XVI-75, activated by follistatin; very few cells were stained in the control (A) but strong expression was found in all the cells after follistatin (B). (C, D) Line XVI-97, activated by NGF; in the control dishes a very faint dot-like expression was found in few cells (C), while much stronger staining was detected after NGF (D). (E, F) Line XVI-73, repressed by RA; strong staining was present in several cells of control dishes (E), but *lacZ* expression was clearly reduced after RA (F). The corresponding genes for these three lines are novel [5]. Scale bar, 100 μm (A, B, E, F) and 30 μm (C, D).

TABLE 5

Identification of Cell Lines Responding to Follistatin, NGF, and Retinoic Acid

	Total G418^R clones	Retinoic acid[a]		Follistatin		NGF	
		Activated	Repressed	Activated	Repressed	Activated	Repressed
Exp. I	157	6	16	4	3	3	5
Exp. II	45	—	—	3	0	4	0
Exp. III	60	—	—	2	1	1	0
		4%	10%	3%	1.5%	3%	2%

Note. ES cells were electroporated with IRESβgeo vector, and gene-trapped colonies were tested for their responsiveness to retinoic acid, follistatin, or NGF with the protocol outlined in the legend of Fig. 4. The number of lines showing specific activation or repression of *lacZ* expression (assessed by X-gal staining) is indicated. The results of three different experiments are shown.

[a] Retinoic acid was tested only in the first experiment.

onstrated that it works stably in mouse embryos and it is functional in different tissues [16, 22]. The use of the IRES sequence in mammalian transgenesis has been reviewed by Mountford and Smith [15]. Our data represent the first extensive analysis of the application of an IRES-based vector for gene trap search. Comparison of the results obtained with IRESβgeo and pGT1.8geo showed that a more efficient and accurate detection of gene trap insertions was achieved when IRESβgeo was used.

First, the number of G418-resistant colonies obtained with IRESβgeo was about sixfold higher than pGT1.8geo (Table 2). Production of reporter/selection activity is independent of the reading frame of the tagged gene when IRESβgeo is used. Therefore, translation of βgeo depends only on the orientation, and half of the insertions within chromosomal transcriptional units should be productive. As a consequence, the total number of G418-resistant colonies is higher and fewer experiments are needed to obtain more colonies.

Second, IRESβgeo allowed for a more reliable detection of *lacZ* expression since β-gal is not fused with a host protein. This is especially important because certain βgeo fusions may lead to a partial or complete inactivation of β-gal enzymatic activity or to targeting into a subcellular site where β-gal cannot be detected [12]. Indeed, higher proportions of β-gal-positive colonies were detected with IRESβgeo: about 30% of the lines obtained with pGT1.8geo had detectable levels of *lacZ* expression, but the ratio rose to about 80% when IRESβgeo was used (Table 2). As expected, β-gal was detected in different subcellular localizations in lines generated with pGT1.8geo (Fig. 2), while only cytoplasmic β-gal localizations were found in cell lines generated with IRESβgeo (Figs. 3 and 5). Altogether, the overall efficiency of detectable gene trap events produced by IRESβgeo was about 15-fold higher than pGT1.8geo.

Analysis of a large number of insertion events obtained with IRESβgeo suggested that this vector has no particular constraint or limitation for its use in gene trap screens. IRESβgeo allowed capture of genes with a broad range of expression in ES cells, including genes expressed at very low levels in ES cells (Fig. 3). This is interesting because it avoids the need of a SAβgal-PGKneo vector [1], where the *neo* gene is under the control of a constitutive promoter. The SAβgal-PGKneo vector is able to trap genes not expressed in ES cells but it is very inefficient because of the very low proportion of β-gal-positive cells among the total G418-resistant cells [1, 2]. Our results suggest that even those genes whose expression in ES cells is very low can be efficiently captured using the IRESβgeo vector. By changing the concentration of G418 during the selection step, it should be possible also to select for genes with low, medium, or high expression in ES cells (unpublished results). Cloning of genes tagged by

TABLE 6

Embryonic Expression Patterns before and after *in Vitro* Preselection

lacZ expression *in vivo* (9.5–14.5 dpc)	No *in vitro* preselection	After *in vitro* preselection
Widespread or ubiquitous patterns[a]	17 (47%)	11 (34%)
Restricted patterns[b]	9 (25%)	16 (50%)
Negative (no expression)[c]	10 (28%)	5 (16%)
CNS[d]	10 (28%)	16 (50%)

Note. lacZ expression was checked by X-gal staining of heterozygous embryos or, for some lines, of chimeric embryos; in most cases, three different stages were analyzed for each line: 9.5–10.5, 11.5–12.5, 13.5–14.5 days post coitum (dpc). Altogether, *in vivo* expression data were obtained for 32 preselected gene-trap lines and for 36 lines where no preselection was applied.

[a] Staining was found in most or all tissues.

[b] Staining was restricted to specific tissues or organs in all the stages.

[c] No β-gal activity was detected in any of the stages.

[d] Staining was found mostly or exclusively in the central nervous system (this pattern represents a subclass of restricted and widespread patterns).

IRESβgeo is not more complicated than any other vector currently in use. Isolation of the tagged cDNA from 42 gene trap lines produced with IRESβgeo showed that 30% corresponded to known genes and 70% were new genes [5]. The known genes belonged to several classes coding for proteins with different functions and found at all main subcellular localizations, and vector insertions were present at different positions in the individual genes. These observations suggest that IRESβgeo integrates randomly in the genome, without any apparent bias. Generation of transgenic mice from a large number of gene-trapped lines showed that IRESβgeo allows production of chimeric mice and germ-line transmission at the same efficiencies of the original βgeo vector. Analysis of the spatiotemporal distribution of *lacZ* expression in heterozygous mice indicated that IRESβgeo works efficiently in a wide range of tissues both during development and after birth [20].

One of the main aspects of a large gene trap search is the number of mutant mice that can be generated and analyzed for the identification of interesting novel genes. Although effective, gene trap works randomly in its basic design and isolation of a large number of developmentally regulated genes requires screening of a vast amount of integration events. A large gene trap screen [23] showed that most part of the integrations either had ubiquitous *lacZ* expression or no *lacZ* expression when tested at 8.5 days of embryonic development. The most laborious and time-consuming step of gene trapping is by far the generation and analysis of mice for each line, therefore it is more convenient to preselect tagged cell lines *in vitro* in order to enrich for genes with potential interesting expression patterns *in vivo,* while excluding the ubiquitously expressed genes. Comparison of the *lacZ* expression patterns *in vitro* and *in vivo* for several gene trap lines showed that most lines with ubiquitous staining in the cells had also ubiquitous *lacZ* expression during embryogenesis, while tissue- or region-specific expression during development was found in most part of the lines with restricted or very faint staining *in vitro* (Table 4). ES cells can be differentiated *in vitro* into embryoid bodies that can give rise to several different cell types as neurons, cardiomyocites, skeletal myocites, and hematopoietic cells [reviewed in 24]. *In vitro* differentiation of ES cells has been used for the selection of gene-trapped lines with expression in one or more defined embryonic tissues [14]. Although very useful, differentiation protocols of ES cells through the formation of embryoid bodies may be too laborious to be used on a large-scale gene trap program. Another possibility is offered by the screening of tagged cell lines *in vitro* with growth/differentiation factors, for the identification of genes that are activated (or repressed) by a certain factor. In one study [13], 20 gene trap integrations were identified that responded to retinoic acid (RA) *in vitro;* interestingly, 19 of these 20 integrations showed restricted expression at 8.5–11.5 days of embryogenesis.

We used a simple *in vitro* preselection protocol, where gene-trapped cell lines generated with IRESβgeo were cultured in the presence of some specific growth/differentiation factors, namely follistatin, NGF, and RA. The choice of the factors was based on our interest in the identification of genes involved in the control of embryogenesis, particularly of the neural development. RA is known to have a wide role in tissue patterning during vertebrate development, and it is able to promote ES cell differentiation into neurons under certain conditions [25]. NGF acts as a trophic and differentiation factor for neurons [26], and it can also promote neuronal differentiation of ES cells *in vitro* [27]. Follistatin has been shown to display direct neutralizing activity, possibly acting as a neural inducer *in vivo* [21]. Using this protocol, we were able to identify several gene trap integrations where *lacZ* expression was either activated or repressed by one of the above factors. Cloning of the tagged cDNAs from six activated/repressed lines showed that the corresponding genes were all novel and their expression was mostly confined to nervous and mesodermal tissues [20]. These observations suggest that this is an efficient method for the identification of novel developmental control genes acting downstream of specific signaling pathways. Preselection of gene-trapped cell lines having a regulated or faint *lacZ* expression *in vitro* and responding to the above factors, resulted in a marked increase for lines with restricted *in vivo* expression patterns during development (Table 6). More interestingly, half of these lines showed a prominent expression in the developing nervous system.

In summary, we showed that IRESβgeo is more suitable than conventional βgeo vectors for the identification of a large nubmer of gene trap events or when preselection procedures are applied to ES cells. The application of IRES elements to more sensitive reporter systems such as the vital marker green fluorescent protein [28] and the use of defined ES cell preselection protocols are expected to increase the frequencies at which novel developmental genes are mutated by gene trapping, thus rendering this approach more powerful and feasible.

We thank Drs. A. Nagy for the R1 ES cell line, W. C. Skarnes for the pGT1.8geo vector, and T. Takeuchi for the IRES sequence. We also thank Dr. A. Mansouri for initial participation in the project. The excellent technical help of S. Mahsur, M. Breuer, R. Scholz, and M. Walther is gratefully acknowledged. This project was generously supported by the Amgen Inc., Thousand Oak, CA, and by the Max Planck Society. P.B. was supported in part by a European Community fellowship and by a grant from Telethon-Italy (Grant 367). M.T. was partly supported by fellowships from the European Molecular Biology Organization and the European Community.

REFERENCES

1. 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.
2. 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.
3. Skarnes, W. C., Auerbach, B. A., and Joyner, A. L. (1992). A gene trap approach in mouse embryonic stem cells: The *lacZ* reporter gene is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice. *Genes Dev.* **6,** 903–918.
4. 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.
5. Chowdhury, K., Bonaldo, P., Torres, M., Stoykova, A., and Gruss, P. (1997). Evidence for the stochastic integration of gene trap vectors into the mouse germ line. *Nucleic Acids Res.* **25,** 1531–1536.
6. Chen, Z., Friedrich, G. A., and Soriano, P. (1994). Transcriptional enhancer factor I disruption by a retroviral gene trap leads to heart defects and embryonic lethality in mice. *Genes Dev.* **8,** 2293–2301.
7. DeGregori, J., Russ, A., von Melchner, H., Rayburn, H., Priyaranjan, P., Jenkins, N. A., Copeland, N., and Ruley, H. E. (1994). A murine homolog of the yeast RNA1 gene is required for postimplantation development. *Genes Dev.* **8,** 265–276.
8. Takeuchi, T., Yamazaki, Y., Katoh-Fukui, Y., Tsuchya, 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.
9. 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 commissural axon guidance in the developing vertebrate nervous system. *Cell* **87,** 1001–1014.
10. 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.
11. Yeo, T. T., Yang, T., Massa, S. M., Zhang, J. S., Honkaniemi, J., Butcher, L. L., and Longo, F. M. (1997). Deficient LAR expression decreases basal forebrain cholinergic neuronal size and hippocampal cholinergic innervation. *J. Neurosci. Res.* **47,** 348–360.
12. Skarnes, W. C., Moss, J., Hurtley, S., and Beddington, R. (1995). Capturing genes encoding membrane and secreted proteins important for mouse development. *Proc. Natl. Acad. Sci. USA* **92,** 6592–6596.
13. 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,** 1677–1688.
14. Baker, R. K., Haendel, M. A., Swanson, B. J., Shambaugh, J. C., Micales, B. K., and Lyons, G. E. (1997). *In vitro* preselection of gene-trapped embryonic stem cell clones for characterizing novel developmentally, regulated genes in the mouse. *Dev. Biol.* **185,** 201–214.
15. Mountford, P. S., and Smith, A. G. (1995). Internal ribosome entry sites and dicistronic RNAs in mammalian transgenesis. *Trends Genet.* **11,** 179–184.
16. Kim, D. G., Kang, H. M., Jang, S. K., and Shin, H.-S. (1992). Construction of a bifunctional mRNA in the mouse by using the internal ribosome entry site of the encephalomyocarditis virus. *Mol. Cell Biol.* **12,** 3636–3643.
17. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W., and Roder, J. C. (1993). Derivation of completely cell culture-derived mice from early passage embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **90,** 8424–8428.
18. Gossler, A., and Zachgo, J. (1993). Gene and enhancer trap screens in ES cell chimeras. *In* "Gene Targeting: A Practical Approach" (A. L. Joyner, Ed.), pp. 181–225, IRL Press, Oxford.
19. Nagy, A., and Rossant, J. (1993). Production of completely ES cell-derived fetuses. *In* "Gene Targeting: A Practical Approach" (A. L. Joyner, Ed.), pp. 147–178, IRL Press, Oxford.
20. Stoykova, A., Chowdhury, K., Bonaldo, P., Torres, M., and Gruss, P. (1998). Gene trap expression and mutational analysis for genes involved in the development of the mammalian nervous system. *Dev. Dynamics* **212,** 198–213.
21. Hemmati-Brivanlou, A., Kelly, O. G., and Melton, D. A. (1994). Follistatin, an antagonist of activin, is expressed in the Spemann organizer and displays direct neuralizing activity. *Cell* **77,** 283–295.
22. Mountford, P., Zevnik, B., Düwel, A., Nichols, J., Li, M., Dani, C., Robertson, M., Chambers, I., and Smith, A. (1994). Dicistronic targeting constructs: Reporters and modifiers of mammalian gene expression. *Proc. Natl. Acad. Sci. USA* **91,** 4303–4307.
23. Würst, W., Rossant, J., Prideux, V., Kownacka, M., Joyner, A., Hill, D. P., Guillermot, 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.
24. Weiss, M. J., and Orkin, S. H. (1996). *In vitro* differentiation of murine embryonic stem cells: New approaches to old problems. *J. Clin. Invest.* **97,** 591–595.
25. Bain, G., Kitchens, D., Yao, M., Huettner, J. E., and Gottlieb, D. I. (1995). Embryonic stem cells express neuronal properties *in vitro. Dev. Biol.* **168,** 342–357.
26. Jelsma, T. N., and Aguayo, A. J. (1994). Trophic factors. *Curr. Opin. Neurobiol.* **4,** 717–725.
27. Wobus, A. M., Grosse, R., and Schöneich, J. (1988). Specific effects of nerve growth factor on the differentiation pattern of mouse embryonic stem cells *in vitro. Biomed. Biochim. Acta* **47,** 965–973.
28. Zernicka-Goetz, M., Pines, J., McLean-Hunter, S., Dixon, J. P., Siemering, K. R., Haseloff, J., and Evans, M. J. (1997). Following cell fate in the living mouse embryo. *Development* **124,** 1133–1137.

Received April 14, 1998
Revised version received June 24, 1998