An *in vitro* Screen to Isolate Developmentally Regulated Genes in Mice

by

Christina Ann Scherer

B.A., Biology
Amherst College, 1988

Submitted to the Department of Biology in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy
at the

Massachusetts Institute of Technology

February 1995

© 1995 Massachusetts Institute of Technology
All rights reserved

Signature of Author................................................................................

Department of Biology
February, 1995

Certified by.................................................................

Nancy Hopkins
Professor, Dept. of Biology
Thesis Supervisor

Henry Earl Ruley
Professor, Dept. of Microbiology and Immunology, Vanderbilt University
Thesis Supervisor

Accepted by.............................................................................

David Housman
Professor, Department of Biology
Committee Chairman
Abstract

In this thesis, a process to mutate and identify developmentally regulated genes in the mouse, *Mus Musculus*, was developed. To facilitate this, a new gene trap retrovirus (U3βgeoSupF) that selectively disrupts genes expressed in totipotent embryonic stem (ES) cells was constructed. This construct contains promoterless coding sequences for a LacZ-Neo fusion protein inserted into the long terminal repeats of a Moloney Murine retrovirus. Integration of the retrovirus into expressed chromosomal loci results in the expression of cellular-proviral fusion transcripts. Thus, integrations into expressed genes can be directly selected for by culturing infected cells in G418-containing medium. G418-resistant ES cell clones were selected and induced to form embryoid bodies, which contain a variety of differentiated cell types, in culture. Clones exhibiting regulated expression of the βgeo reporter gene (as assessed by X-Gal staining) were used to generate chimeric mice. Seven proviral insertions in differentially regulated genes were transmitted to the germline of chimeric mice, three of which are described in this thesis. All seven mouse lines exhibited *in vivo* X-Gal staining that was accurately predicted by the *in vitro* embryoid body assay. For instance, genes expressed in ES cells but repressed in embryoid bodies were expressed in blastocysts but repressed in post-implantation embryos. Conversely, genes expressed at low levels in ES cells but induced upon differentiation showed wide-spread expression in post-implantation embryos, but little or no expression in blastocysts. One of the proviruses inserted into a novel mouse gene with significant homology to a yeast open reading frame of unknown function. This gene exhibits diffuse expression throughout post-implantation embryos, but is expressed at higher levels in some regions, particularly neural cell lineages. Mice homozygous for the gene trap insertion into the *Neural Regionalized (Nrd)* gene die shortly after embryo implantation. The results presented in this thesis indicate that the *in vitro* embryoid body screen is an efficient and accurate method for the identification of novel genes involved in early mouse development.
Acknowledgments

Having spent my time at MIT in two different labs and three different floors of the Cancer Center, I have a great many people to thank for their help and encouragement throughout this time. They have made my experience at MIT an invaluable one.

I would first like to thank my two advisors, Earl Ruley and Nancy Hopkins, for their dedication and guidance throughout my graduate career. Earl's enthusiasm for science drew me to his lab many years ago, and continues to impress me now. He gave me a great deal of independence with this project (not completely by choice!), which I greatly appreciate. Nancy graciously allowed me to continue my studies at MIT in her laboratory. Her insight and dedication to science will always be an inspiration to me. I would also like to thank David Housman for his advice throughout the years, and Hazel Sive and Connie Cepko for their thoughtful comments on this manuscript.

To the members of the MIT branch of the Ruley lab: Sita Reddy, Wen Chang, Harald von Melchner and James DeGregori, who patiently helped me in my first experiences with molecular biology. I would especially like to thank Peggy Kolm, and Scott Lowe my fellow classmates, for many helpful discussions pertaining to science and life in general (the latter usually accompanied by beers...).

To all the members of the Hopkins lab, past and present, who made me feel welcome in the lab (even though I worked on an inferior genetic organism-hah!). In particular, I would like to thank Patty Culp, Caroline McMurtrie and Susan Winandy for their friendship and advice (scientific and otherwise); Nick Gaiano, Shuo Lin, Adam Amsterdam, Mike Farrell, Tina Kim and Tom Becker for helpful discussions; Miguel Allende for advice on in situ hybridizations; Kate Willett for help with sectioning and allowing me to use her microscope so many times; Caroline McMurtrie and Dean Thomson, who not only kept the lab running, but were also good friends (and great diving buddies!); and Marie Marshall, who made sure that everything in the office was running smoothly.

To all the members of the Vanderbilt Ruley lab, especially Geoff Hicks, Michael Roshon and Doug Williamson, who not only allowed me to use their benches and reagents on my periodic visits, but also made sure that I saw the sights of Nashville. In addition, Abudi Nachabeh provided excellent technical assistance and took good care of all my mice.

Special thanks are due to my collaborator, Jin Chen (formerly of the Hopkins lab!), whose cheerful outlook on science and life have inspired me the past year and a half. In addition, she was my gracious hostess (along with husband Mark and son Ian) on my frequent trips to Nashville.

I would also like to thank Marjorie Kummiskey, for her help with all aspects of keeping my mice at MIT, and Helen Rayburn for letting me use their facilities (and borrow their reference texts for extended periods of time).
To all of my friends, from Boston and elsewhere, without whose support I couldn't have made it. In particular, I want to thank the following: Chave Carr for those much needed lunches; Margaret Weigel, who introduced me to the local rock scene and the joys of East Cambridge (BIS, Madgel); Danyelle Desjardins, who made sure that I kept in touch with all our Amherst friends; Stacy Haughey and Shelley Hiatt, who were great roommates, friends and cat "aunts"; Ellie Hays, for all of her cheerful E-Mail messages; Sue Dahl and Brenda Williams, who were always around when I needed someone to talk to; my diving buddies Jeanne Century, Dean Thomson and David Schloerb, who ensured that I got some fresh air once in a while; the members of the Acid Blobs and Masters Swimming, with whom I spent many an hour away from the lab in search of athletic excellence; and Tim Dilligan, who introduced me to a completely different side of MIT (along with the other members of Türköiz Kööki) and who saw me through most of this thesis project. I would especially like to thank my current roommate, Leisa Johnson, who not only suffered through as my thesis took over the apartment, but also helped me maintain my sanity throughout the past year, especially in the final few weeks of this thesis (go Cornhuskers!).

To my cat Pandora, who kept me company and kept the computer warm while I stared at the computer screen for hours on end.

Lastly, I would like to thank my family for their love and support throughout my life. This thesis is dedicated to my parents, Barbara and Mike Scherer, whose constant love and encouragement kept me going, even when the going got tough. Thanks for everything- I love you both!
# Table of Contents

Abstract ................................................................. 2
Acknowledgments .................................................... 3
List of Figures and Tables ........................................ 7
Chapter 1: Introduction ............................................ 8
   Introduction ................................................................ 9
      Forward .................................................................. 9
      Early events in mouse embryogenesis: a brief description .. 9
      Methods for studying development: A brief overview ...... 11
      Embryonic Stem Cells .......................................... 12
      Targeted and random mutagenesis ......................... 13
      Retroviruses as insertional mutagens ..................... 14
      Increasing the efficiency of insertional mutagenesis ..... 15
      Gene Trap Vectors ............................................. 16
      Applications of gene trap technology ..................... 22
      Gene trapping versus cDNA subtraction techniques .... 24
      Methods for cloning sequences flanking gene trap insertions ... 25
      Development of an *in vitro* screen ....................... 26
      Genes regulated during development .................... 30
      Summary .......................................................... 34
Chapter 2: Development of New Retroviral Gene Traps .... 36
   Introduction ................................................................ 37
   Results ..................................................................... 40
      Construction and testing of U3LacZ-pgkNeo ................ 40
      Construction and testing of U3βgeoSupF .................... 45
   Discussion .................................................................. 50
   Materials and Methods .......................................... 53
      Plasmid Construction ............................................ 53
      Construction and titering of Y'2 Virus Producer Lines .... 53
      Infection of target cells ....................................... 54
      Southern and Northern analysis of individual clones .... 55
      RNase Protection Analysis .................................... 56
Chapter 3: The *in vitro* screen .................................. 59
   Introduction ................................................................ 60
   Results ..................................................................... 64
      Selection and analysis of regulated clones ................ 64
      β-galactosidase activity accurately reflects the level of transcription. .... 71
      Expression of βgal in transgenic mice is accurately predicted by *in vitro* gene expression. ........ 76
   Discussion .................................................................. 83
      U3βgeoSupF accurately reports endogenous gene expression .................. 83
      Negative selection against the viral primer binding site may cause proviral deletions in ES cells ... 84
      Summary .................................................................. 88
   Materials and Methods .......................................... 89
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Culture of ES Cells</td>
<td>89</td>
</tr>
<tr>
<td>RNase protection analysis</td>
<td>90</td>
</tr>
<tr>
<td>Construction of germline chimeras</td>
<td>91</td>
</tr>
<tr>
<td>Embryonic Expression of LacZ Fusion Genes</td>
<td>91</td>
</tr>
<tr>
<td>Chapter 4: Analysis of the Mike, Peabody and O3 Mouse Lines</td>
<td>93</td>
</tr>
<tr>
<td>Introduction</td>
<td>94</td>
</tr>
<tr>
<td>Results</td>
<td>96</td>
</tr>
<tr>
<td>Peabody</td>
<td>96</td>
</tr>
<tr>
<td>Mike</td>
<td>100</td>
</tr>
<tr>
<td>O3</td>
<td>105</td>
</tr>
<tr>
<td>Discussion</td>
<td>120</td>
</tr>
<tr>
<td>The mystery of the missing phenotypes</td>
<td>120</td>
</tr>
<tr>
<td>The O3 gene product: Form, function and further experiments</td>
<td>124</td>
</tr>
<tr>
<td>Material and Methods</td>
<td>127</td>
</tr>
<tr>
<td>PCR Cloning Methods</td>
<td>128</td>
</tr>
<tr>
<td>Cloning of PCR Products</td>
<td>132</td>
</tr>
<tr>
<td>Screening of cDNA library</td>
<td>134</td>
</tr>
<tr>
<td>Whole mount in situ hybridization:</td>
<td>136</td>
</tr>
<tr>
<td>Sectioning of LacZ stained embryos:</td>
<td>140</td>
</tr>
<tr>
<td>Chapter 5: Summary and General Discussion</td>
<td>142</td>
</tr>
<tr>
<td>Summary</td>
<td>143</td>
</tr>
<tr>
<td>Problems with this study</td>
<td>145</td>
</tr>
<tr>
<td>Future applications of the in vitro screen</td>
<td>146</td>
</tr>
<tr>
<td>Bibliography</td>
<td>150</td>
</tr>
</tbody>
</table>
## List of Figures and Tables

<table>
<thead>
<tr>
<th>Figure:</th>
<th>Page:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>20</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>29</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>41</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>43</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>46</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>48</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>68</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>70</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>73</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>75</td>
</tr>
<tr>
<td>Figure 3.5</td>
<td>77</td>
</tr>
<tr>
<td>Figure 3.6</td>
<td>80</td>
</tr>
<tr>
<td>Figure 3.7</td>
<td>81</td>
</tr>
<tr>
<td>Figure 3.8</td>
<td>82</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>98</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>99</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>101</td>
</tr>
<tr>
<td>Figure 4.4</td>
<td>103</td>
</tr>
<tr>
<td>Figure 4.5</td>
<td>107</td>
</tr>
<tr>
<td>Figure 4.6</td>
<td>110</td>
</tr>
<tr>
<td>Figure 4.7</td>
<td>113</td>
</tr>
<tr>
<td>Figure 4.8</td>
<td>114</td>
</tr>
<tr>
<td>Figure 4.9</td>
<td>116</td>
</tr>
<tr>
<td>Figure 4.10</td>
<td>118</td>
</tr>
<tr>
<td>Figure 4.11</td>
<td>129</td>
</tr>
<tr>
<td>Figure 4.12</td>
<td>131</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table:</th>
<th>Page:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1.1</td>
<td>22</td>
</tr>
<tr>
<td>Table 1.2</td>
<td>23</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>65</td>
</tr>
<tr>
<td>Table 3.2</td>
<td>66</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>104</td>
</tr>
</tbody>
</table>
Chapter 1:

Introduction
Introduction

Forward

Events during the first few days of post-implantation development in the mouse, *Mus musculus*, result in not only the rapid growth of the embryo but also the determination of primitive lineages and establishment of the primary body axis. Because of the inaccessibility of the implanted embryo, it has been very difficult to study these early events. Although the general mechanisms of early development are known, the molecular principles underlying these mechanisms are not well understood. Therefore, a number of approaches have been taken to better understand this period of development, with emphasis being placed upon isolating genes involved in these early patterning events. A number of genes have been identified either by homology to genes identified in other organisms or by mutational analysis; however, there are undoubtedly many which remain uncharacterized. Therefore, in this thesis I have explored the efficacy of an in vitro screen to identify novel genes that are differentially regulated during the early stages of embryogenesis.

Early events in mouse embryogenesis: a brief description

Mouse embryogenesis differs from that of other commonly studied organisms (such as the sea urchin, zebrafish and *Xenopus*) in that early development occurs very slowly (Hogan et al. 1986; Rugh 1990). For an excellent diagram of early mouse development, please refer to page 50 of *Manipulating the Mouse Embryo* (Hogan et al. 1986). Ovulation and fertilization of the egg usually occurs around midnight if the mice are kept on a 12 hour light/dark cycle. Noon of the following day is then defined as day 0.5 of development. Cleavage to the two cell stage does not occur until 24 hours after fertilization. As the embryo moves through the oviduct to the uterus, where implantation occurs, it continues to divide slowly without any increase in mass. The first cellular differentiation event occurs between the morula (16 cell) and blastocyst (64 cell) stages, when two distinct cell lineages, the trophectoderm (TE) and the inner cell mass (ICM), are formed.
The trophectoderm initially forms an epithelial layer which surrounds the blastocoel and the 20-30 cells of the ICM. The "mural" trophectoderm cells surrounding the blastocoel stop dividing and become large polyploid cells (primary trophoblastic giant cells), which first invade the uterine epithelium during embryo implantation. The "polar" TE cells, which are located around the ICM, remain diploid and proliferate further to form a number of structures in the post-implantation embryo. Some cells migrate around the embryo and replace the primary trophoblastic cells, while others invade the uterine epithelium to form the bulk of the placenta. A third subset migrates into the blastocoel cavity to form the extraembryonic ectoderm of the egg cylinder at about 6 days post coitum (p.c.). The extraembryonic ectoderm eventually recedes towards the placenta to form the chorion.

The second differentiation event occurs at about 4.0 days p.c., when cells of the ICM form the primitive endoderm and primitive ectoderm. The primitive endoderm colonizes the extraembryonic endoderm of the yolk sac, whereas the primitive ectoderm gives rise to the extraembryonic mesoderm as well as the embryo proper. The blastocysts then implant into the uterus between days 4.5 and 5.5 of development, after which the rate of cell division in the embryo (particularly in the primitive ectoderm) rapidly increases. By day 6.0 p.c., the primitive ectoderm has formed an organized epithelial layer surrounding a central proamniotic cavity.

The process of gastrulation results in the formation of the three germ layers (endoderm, mesoderm and ectoderm) and the establishment of the basic body plan of the embryo. Gastrulation begins at the egg cylinder stage (about 6.5 days p.c.) with the formation of the primitive streak. Cells ingress from the epithelial layer of the primitive ectoderm and move through the streak to form mesoderm and definitive endoderm (Hogan et al. 1986). As gastrulation proceeds, a transient structure known as the node appears at the anterior end of the streak. Cells moving through the node towards the anterior end of the egg cylinder form a strip of mesodermal cells. These will organize into the notochord and somites as the node moves toward the posterior of the embryo. Thus, there is an anterior to posterior gradient of development; while mesoderm is still being formed at the posterior end of the embryo, the somites begin condensing in the anterior end of the embryo. Around 7-7.5 days p.c., the ectoderm anterior to the primitive streak and above the notochord is induced to form the neural plate, which later gives rise to the central nervous system. In addition, the anterior
ectoderm thickens to form the head process. Starting around 8 days p.c., the neural plate gradually folds up in an anterior to posterior movement to form the neural tube. By day 8.5 p.c., a recognizable embryo has been formed, with a bulging head process, neural fold/tube (in the process of closing) and primitive heart.

Although this is a brief summary of the events occurring during early embryogenesis, it can be seen that many important mechanisms are taking place. The complex cell movements which occur during gastrulation and additional morphogenic events have been identified to a certain extent through embryological studies. However, the molecular changes associated with these movements have not been elucidated. Inductive events (defined as the ability of one group of cells to change the fate of another group of cells) are known to result in the determination of virtually all of the primitive and differentiated cell lineages in the embryo (Kessler and Melton 1994). Although many factors, including members of the FGF and TGF-β families, have been found to act as inductive agents, the endogenous agents have not yet been determined (ibid). Thus, many questions regarding early post-implantation development need to be answered on a molecular level. These include the following:

-- How do differentiated cells arise and organize to form tissues and organs?
-- How is pattern generated- i.e. what starts the cell movements which result in primitive streak formation?
-- What are the endogenous inducing agents for mesoderm and neural lineages?
-- What factors allow cells to respond to particular inducing agents (defined as competence)?
-- How is cellular competence modified?

In order to identify some of the molecular processes taking place in early post-implantation development, I was interested in identifying genes differentially regulated during this time period. The in vitro screen described in this thesis could be one method for isolating such genes.

Methods for studying development: A brief overview

One method for investigating the complicated events of early embryogenesis involves identifying factors which are responsible for the various inductive activities in the early embryo. This has been quite successful in organisms which develop externally, such as Xenopus, which has become a
model vertebrate system for pattern formation (Melton 1991), axis formation (McMahon 1993; Sive 1993), and mesoderm and neural tissue induction (Kessler and Melton 1994). Unfortunately, the sorts of experiments easily conducted in *Xenopus* are not very feasible in the mouse because of its intrauterine development. Therefore, genetics has been the approach of choice in the mouse.

A classical genetic approach to development involves characterizing existing or new embryonic mutations and then isolating the genes responsible for them. Again, this is not a very efficient process in the mouse. Because of the large size of the (mouse) genome, positional cloning of genes is extremely time consuming, despite advances made in the genome project. Although a few genes responsible for known embryonic mutations have been cloned recently [for instance the Steel locus and *Brachyury*, (Copeland *et al.* 1990; Hermann *et al.* 1990; Huang *et al.* 1990; Zsebo *et al.* 1990)], most of the known (spontaneous and induced) mutations involve easily identified postnatal characteristics (Green 1989). Finding new embryonic mutations is also not a trivial undertaking. Large-scale genetic screens utilizing chemical mutagenesis such as those done in *Drosophila* and now also in the zebrafish, *Brachydanio rerio* (Rossant and Hopkins 1992; Nüsslein-Volhard 1994), are infeasible in a mammalian system because of prohibitive costs and inefficiency. The intrauterine development and small litter size characteristic of mammalian development make the identification of mutations quite difficult; recessive mutations involving developmental phenotypes (especially lethal ones) can often only be identified by sacrificing pregnant females. Therefore, a number of alternative methods for identifying genes involved in mouse embryogenesis have been developed.

*Embryonic Stem Cells*

Many of the methods described in the following sections take advantage of a feature unique to mouse molecular genetics. Embryonic stem (ES) cells are totipotent cells derived from the inner cell mass (ICM) of mouse blastocysts (Evans and Kaufman 1981; Martin 1981). ES cells are much like the cells of the early ICM, as they can contribute to the trophotoderm and primitive endoderm when injected into blastocysts (Beddington and Robertson 1989). However, they primarily colonize the primitive ectoderm, which results in the formation of chimeric embryos. ES cells contribute to the germline of these chimeras at high
frequencies (Bradley et al. 1984). Like embryonal carcinoma (EC) cell lines (which are derived from teratocarcinomas), ES cells can be maintained in an undifferentiated state in culture by growing on feeder layers and/or leukemia inhibitory factor (LIF), a factor which inhibits differentiation (Nichols et al. 1990; Pease et al. 1990). Moreover, ES cells can be experimentally manipulated in vitro just like other cultured cell lines without compromising totipotency (Wagner et al. 1985; Robertson et al. 1986; Reddy et al. 1992). Alternatively, ES cells can be induced to differentiate into a variety of cell types in vitro. When ES cells are plated at a high density in suspension culture, "embryoid bodies" form which have an inner layer of ES cells and an outer endoderm layer (Martin 1981). If allowed to attach to a gelatinized tissue culture plate, these embryoid bodies then differentiate into a wide variety of cell types, including melanocytes, cartilage, skeletal and smooth muscle, basal lamina, and even beating heart structures (Doetschman et al. 1985). Lastly, ES cells isolated from embryos homozygous for various mutations can be used to study these mutations in vitro (Martin et al. 1987). This provides a system in which one can perform biochemical studies of both lethal and viable null mutations. In summary, ES cells provide the mouse geneticist with an efficient medium for the generation and analysis of mutations in vitro and in vivo.

Targeted and random mutagenesis

One very successful application of reverse genetics in the mouse has involved the targeted mutagenesis of genes suspected to be involved in developmental processes. The cloning of mouse homologues to many of the genes originally defined in early saturation mutagenesis of Drosophila [reviewed in (Nüsslein-Volhard 1994)] revealed many highly conserved gene families including the homeobox (Hox) genes and paired box (Pax) genes [reviewed in (Kessel and Gruss 1990; McGinnis and Krumlauf 1992)]. Targeted mutagenesis by homologous recombination in ES cells of these and other genes has revealed many conserved functions [reviewed in (Capecchi 1989a; Capecchi 1989b; Rossant 1991)]. Gene targeting techniques have been further refined with the so-called "hit and run" and "double replacement" vectors (Bautista and Shulman 1991; Hasty et al. 1991b; Wu et al. 1994). These vectors have enabled investigators to engineer subtle mutations in specific genes, allowing for more sophisticated analyses of gene functions during development.
Although targeted mutagenesis has provided much information about the contributions of various gene products, its main disadvantage is that only known genes can be disrupted. Therefore, a number of labs have used random insertional mutagenesis with DNA or retroviral vectors to identify novel genes involved in development. These vectors integrate randomly throughout the genome and can produce mutations at a certain frequency. A major advantage of insertional mutagenesis over chemical or radiation-induced mutagenesis is that the insertion vector creates a convenient "tag" for cloning mutant genes. These insertions can be transmitted to the germline of mice by retroviral infection (Jaenisch 1976) or microinjection (Harbers et al. 1981) of early embryos. Alternatively, insertion vectors can be introduced into embryonic stem (ES) cells, where integrations are selected for and positive clones can be used to generate chimeric mice.

*Retroviruses as insertional mutagens*

Retroviruses have several advantages over DNA vectors as insertional mutagens (Gridley et al. 1987; Jaenisch 1988). Retroviruses are naturally occurring transposable elements. The complex life cycle of retroviruses includes the reverse transcription of the viral RNA genome into double stranded DNA and the integration of the DNA copy (known as the provirus) into a single site in the genome of infected cells. Integration is accompanied by a short duplication of host DNA sequences (4-6 bp); no other rearrangements occur (Fields et al. 1986). DNA vectors, on the other hand, tend to integrate in long head to tail tandem arrays at a single chromosomal location (Palmiter and Brinster 1986). Integration as a result of either microinjection or transfection often causes deletions, duplications, rearrangements and translocations of both the insert and host sequences. Thus, observed phenotypes may be due to these alterations rather than being a consequence of integration into a specific locus. Another advantage of retroviruses is the ability to infect cells with a given number of proviruses, depending on the multiplicity of infection (moi). Thus, if only single integrations are desired, a low moi is used, whereas if multiple integration sites are preferable, a high moi can be used. In contrast, DNA copy number can not be predicted. Finally, although retroviruses might preferentially integrate into expressed regions (Rohdewohld et al. 1987; Shih et al. 1988; Sandmeyer et al. 1987; ).
Retroviruses are also amenable to genetic manipulation. Most viral sequences can be replaced by foreign genes without affecting viral integration. The only required sites are the \( \Psi \) packaging signal sequence, the \( (R) \) repeats which flank the virus and promote strand transfer during reverse transcription, the tRNA and polypurine primer binding sites which initiate plus and minus strand DNA synthesis, and the ends of U3 and U5 which are recognized by the viral integrase for integration of the provirus (Fields et al. 1986). Thus, not only can most of \( gag, pol \) and \( env \) be replaced by foreign genes, but genes can also be inserted into the LTR, where they will be duplicated along with normal LTR sequences. Recombinant retroviruses can be produced \textit{in vitro} by transfecting a plasmid retroviral element into a packaging cell line. These cell lines contain one or more defective helper viruses which express all of the proteins necessary for viral packaging but cannot be packaged themselves due to the lack of the \( \Psi \) signal sequence. Packaging cell lines are capable of producing high titer recombinant retroviruses for long periods of time, as they are not killed upon release of retroviruses. Since the host range of a retrovirus is specified by the envelope (env) protein, different helper cell lines can produce retroviruses capable of infecting most cultured cell types. Murine ecotropic packaging lines produce viruses capable of infecting mouse and rat cells, whereas murine amphototropic or Gibbon ape leukemia helper cell lines produce retroviruses which can infect many different cell types, including human, monkey, mouse, rat, dog, cat and chicken. Retroviruses pseudotyped with the vesicular stomatitis virus G-protein have even been used to infect fish embryos (Lin et al. 1994), thus further expanding the scope of retroviral mutagenesis.

\textbf{Increasing the efficiency of insertional mutagenesis}

A major disadvantage with random insertional mutagenesis is its inefficiency [reviewed in (Gridley et al. 1987; Jaenisch 1988)]. Because most insertions occur in non-expressed regions of the genome, only 10% of DNA insertions and 5% of retroviral insertions cause obvious phenotypes in homozygous mice. To increase the efficiency of random insertional mutagenesis, various different constructs which select for integration into or near expressed genes have been developed. One family of vectors, known as
enhancer traps, contain a reporter gene (usually LacZ) driven by a weak promoter. In order to detect expression of the reporter gene, cis-acting enhancer sequences must elevate transcription levels. P-element-based enhancer traps have been used very effectively in Drosophila (O’Kane and Gehring 1987; Bellen et al. 1989), where hundreds of single insertion lines with many different patterns of expression (at least 65% of which were developmentally regulated) were generated. This was possible using a small number of so-called "jump start" males which carry a stable P-transposase source as well as a recombinant P-element enhancer trap. Activation of the transposase in the germline of the male can result in excision of the original enhancer trap and insertion into a new locus. Positional cloning of insertional loci is relatively trivial in Drosophila because of the small genome size and the ability to do cytological mapping on giant chromosomes of the salivary gland. In addition, genes are easier to identify (than vertebrate genes) by sequence analysis and the fact that they have fewer introns. Although enhancer traps have also been used in ES cells, the process is not as efficient as in Drosophila. Approximately 10% of enhancer trap insertions result in the expression of LacZ; of these, 10-30% exhibit regulated expression when used to generate chimeric mice (Gossler et al. 1989; Johnson and Mahon 1993). The lower activation frequency makes sense given the significantly larger genome size of the mouse. Other characteristics of mammalian cells also influence the efficiency of this method. For instance, because of the complex regulatory pathways in vertebrates and the ability of enhancers to influences sequences many kilobases away, it is more difficult to clone the regulatory elements affecting expression of the reporter gene. Unlike Drosophila, expression of an enhancer trap in mammalian cells has rarely been linked with expression of an endogenous gene (Bettenhausen et al. 1994). In addition, it is unlikely that enhancer traps would be highly mutagenic, unless they integrate into a gene controlled by the regulatory sequences.

**Gene Trap Vectors**

In order to enrich for mutagenic insertion events in mammalian cells, a number of groups have developed gene trap vectors to select for integrations into actively expressed genes (Brenner et al. 1989; Gossler et al. 1989; von Melchner and Ruley 1989; von Melchner et al. 1990; Friedrich and Soriano 1991; Reddy et al. 1991; Skarnes et al. 1992; von Melchner et al. 1992; Chang et al. 1993).
These vectors are based on the P-lac fusion constructs originally used in bacteria to isolate new transcription units (Casadaban and Cohen 1979). Promoterless reporter genes are used to usurp cellular promoters, thereby creating an expression tag for that specific gene as well. Integration into expressed genes often results in the functional disruption of that gene. Two basic designs have been used for mammalian gene trap constructs: those utilizing 5' splice sites (splice acceptor, or SA, traps) and those utilizing consensus translation initiation sites (AUG traps). Splice acceptor gene traps can be either DNA or retroviral constructs which contain a consensus 5' splice site (5' ss) upstream of a promoterless reporter gene (usually \textit{LacZ}). An internal drug resistance marker is usually also included to allow selection for all integration events. The vectors can contain the 5' ss sequence alone (Gossler \textit{et al.} 1989; Skarnes \textit{et al.} 1992), or in addition to an ATG which can function as either a translation initiation signal or an internal methionine (Brenner \textit{et al.} 1989; Friedrich and Soriano 1991). In either case, activation of the reporter gene requires integration of the construct in the correct orientation for production of a spliced fusion transcript and protein (see Figure 1.1). Retroviral SA traps contain the splice site-reporter gene cassettes in the opposite orientation to viral transcription to prevent splicing out of the \Psi \Theta (packaging) sequences and polyadenylation in the LTR (Friedrich and Soriano 1991). In general, splicing vectors have a higher activation frequency than AUG traps, presumably because their target size (i.e. introns) is much larger; AUG traps can only be activated by integration into 5' exons or introns (see below). Consistent with the finding that retroviruses tend to integrate in expressed regions of the genome (see above), the activation frequency of retroviral SA traps is higher than that of plasmid-based SA traps (Friedrich and Soriano 1991). As expected, SA traps greatly enrich for mutations; approximately 40% of the insertions transmitted to the germline of transgenic mice have resulted in obvious homozygous phenotypes, including embryonic lethality (Friedrich and Soriano 1991; Skarnes \textit{et al.} 1992).

The AUG traps constructed in our lab contain selectable promoterless reporter genes in the 3' LTR of an enhancerless Moloney murine leukemia virus. In addition, some vectors carry an internal drug resistance marker which allows for selection of all viral integration events. Integration of the provirus into the genome of infected cells results in duplication of the 3' LTR such that the reporter gene is placed only 30 nucleotides from cellular flanking sequences (see Figure 1.1). Activation of reporter gene expression requires integration in or near 5'
exons of transcriptionally active genes which results in a cellular-proviral fusion transcript. Usually, appended RNA from the fusion transcripts is less than 500 nt long. To aid in the translation of such transcripts, the reporter genes contain a Kozak consensus translation initiator site (Kozak 1986) and a Shine-Dalgarno ribosome binding site (Shine and Dalgarno 1974; Steitz and Jakes 1975). Presumably, efficient translation selects for integration events that position the initiation codon near the 5' end of resulting fusion transcript (Kozak 1991). For instance, slightly longer leader sequences seem to result in more efficient translation (i.e. AUG sites too close to the 5' cap will not be recognized by the ribosome machinery). In addition, a strong, upstream, out-of-frame start codon will decrease translation from the proviral start codon. Depending on the reporter gene used, the fraction of activating integrations varies from 1/200-1/2000 of the total integrations. This is partly explained by the presence of in-frame upstream stop codons in some of the gene trap constructs (i.e. U3-His), whereas others, like U3-LacZ or U3βgeoSupF, contain only out-of-frame stop codons. The presence of an in-frame stop codon imposes a strong selective advantage on integration events which result in the provirus providing the first initiating AUG codon (von Melchner et al. 1990). The average activation frequency of AUG traps is approximately 20-fold less than that associated with splice acceptor traps. This can be explained by the smaller target size allowing activation of the AUG traps compared to splice acceptor traps, which can be activated in a number of introns. However, the frequency of mutations as the result of integration into active genes is very similar to that of splice acceptor traps; approximately 40% result in obvious recessive phenotypes. Although the activation frequency of AUG traps is lower than that of SA traps, a potential advantage is that they integrate close to 5' ends of genes, so it is more straightforward to clone 5' regulatory sequences.

Analysis of cellular flanking sequences indicates that the retroviruses do indeed usurp active promoters, rather than activating cryptic ones (von Melchner et al. 1990; Reddy et al. 1991; von Melchner et al. 1992). Therefore, gene traps can function much like expressed sequence tags in creating a library of genome markers/tags (Adams et al. 1991; Adams et al. 1993a; Adams et al. 1993b). Genes trapped have the characteristics of genes transcribed by RNA Polymerase II. In particular, they are expressed prior to integration, undergo splicing, and tend to hybridize to single copy DNA. RNA Pol I and Pol III transcribed genes
Figure 1.1  Integration and activation of retroviral gene traps. Dark stippled boxes represent genomic exon sequences and light stippled boxes represent the proviral LTR. Reporter gene sequences are shown in white. (A) Diagram of a splice acceptor gene trap provirus. The 3' splice site (SA), reporter gene and polyadenylation signal (pA) are in the opposite orientation to viral transcription. Integration of the provirus into an intron of a hypothetical gene is shown. Integration in the correct orientation allows the production of a spliced fusion transcript and protein. (B) Diagram of an AUG gene trap provirus. Duplication of the 3' LTR during the viral life cycle places the reporter gene only 30 nucleotides from cellular flanking sequences. The reporter gene is expressed from fusion transcripts extending from the flanking cellular DNA into the provirus. Integration into the first intron (top) or second exon (bottom) of a hypothetical gene are shown.
Figure 1.1
Retroviral Insertional Mutagenesis Vectors

A. Splice Acceptor Trap

![Diagram of Splice Acceptor Trap]

B. AUG "Promoter" Trap

![Diagram of AUG "Promoter" Trap]
would not be trapped because the transcripts would lack 5' caps and therefore would not be processed or translated efficiently (Banerjee 1980; Sisodia et al. 1987). Analysis of 5' flanking regions of over 500 proviral integrations resulted in approximately 30 being matched to known genes (von Melchner et al. 1992; Chen and Ruley 1994; Hicks et al. 1994). Approximately half of the proviral integrations occurred into intron sequences and half into exons. In addition, half appeared to be upstream of the endogenous gene's AUG initiator codon, whereas the other half appeared to be downstream. Since many of these matches are to cDNA sequences, integrations into introns, alternative exons or near 5' splice sites would probably not reveal homologies to known genes. Therefore, this analysis probably underestimates the likelihood of disrupting known genes. However, it does provide the best indication to date of the mechanisms surrounding gene trapping events.

It appears that most expressed genes can be disrupted using AUG traps. Viral integration is random throughout the genome (Withers-Ward et al. 1994), although there seem to be preferred sites which tend to be near expressed genes (as characterized by DNase hypersensitivity) (Rohdewohld et al. 1987). In most cell lines there seem to be between 10^4 and 10^5 target genes for AUG traps, which corresponds well to the number of active genes (approximately 20,000) as estimated by RNA renaturation experiments (Lewin 1975). The number of gene targets can be estimated by two different methods. One way is to use the activation frequency of the gene traps themselves. The total size of the genome (3 x 10^9 nt) is multiplied by the activation frequency of the gene traps (which range between 1/200 and 1/2500 integration events). This number indicates the total target allowing expression of the gene traps (1.2 x 10^6 - 1.5 x 10^7 nt). Dividing this number by the average size of the appended message for each gene trap gives the total number of genes which are capable of activating U3 gene expression (approximately 1.2-7.5 x 10^4). This is summarized in Table 1.1. In the second method, two AUG trap vectors were used to estimate how many insertions had to take place to knock-out one single copy gene (Chang et al. 1993). Cells lines containing single U3-TK proviruses were subsequently infected with U3-Hygro. Targeted integrations into the U3-TK locus were selected for by growth in hygromycin and gangcyclovir-containing medium. The ratio of hygromycin and gangcyclovir double resistant clones to the total number of Hygro^R clones ranged between 5/1 x 10^5 and 2/2 x 10^5. This represents a total target size of 2 x 10^4 to 1 x 10^5.
Table 1.1

<table>
<thead>
<tr>
<th>U3 Gene</th>
<th>Fraction of proviruses that express U3 gene</th>
<th>Average length of appended cellular RNA (nt)</th>
<th>Maximum number of sites in the genome that can express U3 gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>βgeo</td>
<td>ND (~1:200)</td>
<td>400</td>
<td>4 x 10⁴</td>
</tr>
<tr>
<td>hisD</td>
<td>1:2500</td>
<td>100</td>
<td>1 x 10⁴</td>
</tr>
<tr>
<td>hygro</td>
<td>1:400</td>
<td>270</td>
<td>3 x 10⁴</td>
</tr>
<tr>
<td>lacZ</td>
<td>1:200</td>
<td>400</td>
<td>4 x 10⁴</td>
</tr>
<tr>
<td>tk</td>
<td>1:200</td>
<td>250</td>
<td>6 x 10⁴</td>
</tr>
</tbody>
</table>

Applications of gene trap technology

Gene trap retroviruses can be used in a number of different genetic screens in any susceptible cell line. However, because most cell lines are diploid, it is difficult to isolate new genes responsible for recessive phenotypes (it hasn't been done yet). This problem could be circumvented by using cell lines such as CHO (Chinese Hamster Ovary), which are known to be hemizygous at multiple loci. Using the U3-Hygro gene trap, Chang et al. selected for N-glycosylation mutations (Chang et al. 1993). This system was chosen because a wide variety of mutants can be selected in media containing wheat germ agglutinin. Also, mutations spontaneously arise at a frequency of 1/1 x 10⁵, suggesting that CHO cells may be functionally hemizygous for a number of genes. In fact, they were able to target the GlcNAc transferase I gene at a frequency of 1/5 x 10⁵ gene trap events, or 2 x 10⁸ proviral integrations. Such numbers can be easily achieved with gene trap vectors produced at high titers. Another potential method for achieving recessive phenotypes is to select at higher than usual drug concentrations to try to achieve two hits in the same gene. Although this method has worked for targeted knock-outs in ES cells (Mortensen et al. 1992), it would probably not be very efficient for non-targeted insertional mutations.

Many of the gene traps constructed in our laboratory facilitate selection both for and against reporter gene expression (summarized in Table 1.2), which
allows one to screen for differentially regulated genes. One potentially powerful application of gene trap technology could identify targets of transcription factors or oncogenes in cultured cells. To select genes that are down regulated by the factor of interest, one would start with a library of clones expressing a reporter gene such as tk or gpt (selected in HAT or X-HAT). One would then supply the regulator in trans and select against expression of the reporter gene (in gangcyclovir or 6-thioguanine). Alternatively, to isolate targets that are activated by the regulator, one would start with a library of non-expressing clones and select for acquisition of reporter gene expression. In either case, several rounds of positive/negative selection would be performed, in order to enrich for regulated genes. Depending on the screen used, this may not always be possible.

Table 1.2
Selection Strategies for and Against Expression of Various U3 Reporter Genes

<table>
<thead>
<tr>
<th>U3 Reporter Gene</th>
<th>Positive Selection</th>
<th>Negative Selection</th>
</tr>
</thead>
<tbody>
<tr>
<td>βgeo (LacZ-Neo fusion)</td>
<td>G418, FACS, X-Gal</td>
<td>FACS, X-Gal</td>
</tr>
<tr>
<td>CD4 (T-lymphocyte surface antigen)</td>
<td>Anti-CD4 antibodies, FACS</td>
<td>Anti-CD4 plus complement, FACS</td>
</tr>
<tr>
<td>gpt (E. coli xanthine guanine phosphoribosyl transferase)</td>
<td>X-HAT medium</td>
<td>6-Thioguanine</td>
</tr>
<tr>
<td>hisD (Salmonella histidinol dehydrogenase)</td>
<td>L-Histidinol</td>
<td>None</td>
</tr>
<tr>
<td>hygro (hygromycin phosphotransferase)</td>
<td>Hygromycin B</td>
<td>None</td>
</tr>
<tr>
<td>lacZ (E. coli β-galactosidase)</td>
<td>FACS, X-Gal</td>
<td>FACS, X-Gal</td>
</tr>
<tr>
<td>neo (Tn5 neomycin phosphotransferase)</td>
<td>G418</td>
<td>None</td>
</tr>
<tr>
<td>TK (herpes simplex virus-2 thymidine kinase)</td>
<td>HAT medium</td>
<td>8-BrdU, gangcyclovir</td>
</tr>
</tbody>
</table>
LacZ has been the reporter gene of choice for situations in which direct visualization of gene expression is desired. LacZ encodes the Escherichia coli β-galactosidase, for which several cytological stains exist. Fixed cells can be stained with the chromogenic substrate X-Gal, which is cleaved by β-gal to produce a blue precipitate. Alternatively, live cells can be stained with the fluorescent substrate FDG, and either be visualized by fluorescent microscopy or be analyzed using a cell sorter. These methods allow easy in vitro and in vivo analysis of regulated gene expression. LacZ traps have been used to isolate genes which are regulated during the cell cycle by using FACS analysis (Brenner et al. 1989; Reddy et al. 1991). One of the most elegant applications of LacZ gene traps has been in the study of development. Like cultured cells, embryos can be stained with X-Gal to assess gene expression (Gossler et al. 1989; Reddy et al. 1992; Skarnes et al. 1992; Johnson and Mahon 1993). This allows a direct visualization of where "trapped" genes are expressed, which is not possible with gene traps utilizing drug resistance genes as reporters.

Gene trapping versus cDNA subtraction techniques

An approach which has been used to directly isolate genes differentially expressed during development has been cDNA subtraction analysis. For instance, subtractive cDNA libraries specific for unfertilized eggs, 2-cell, 8-cell and blastocyst-stage embryos have been developed, wherein one can identify stage-specific RNA expression in pre-implantation embryos (Rothstein et al. 1992). Genes regulated around the time of implantation have been identified by several cDNA subtraction schemes. Era-1 and REX-1, for example, were isolated from cDNA libraries specific for genes induced and repressed upon retinoic acid treatment of EC cells (LaRosa and Gudas 1988; Hosler et al. 1989). The murine H19 gene was isolated from a library specific for genes expressed in embryoid bodies (Poirier et al. 1991). The isolation of genes expressed in specific subsets of tissues during development is more problematic, although genes expressed in the embryonic nervous system of Xenopus have been identified by subtractive cloning (Richter et al. 1988). This is more difficult in mouse embryos, because of the small size of the embryo and because specific tissues are not easily separated from each other.
Although subtractive cDNA analysis is an efficient method for cloning novel genes, gene trap analysis of development offers several advantages. Firstly, gene trap selection is extremely sensitive; even weakly expressed genes can be detected. Secondly, gene trapping is not particularly biased for highly expressed genes. Although highly expressed genes might provide a greater target for gene trap selection (since large amounts of fusion transcripts might compensate for translational suppression due to appended cellular sequences), the magnitude of the bias is no more than 3 fold, as estimated from the variation in the size of appended RNA. Indeed, regulated genes may have longer untranslated 5' leaders than constitutively expressed genes and may be preferentially targeted. Thirdly, cells recovered after gene trap selection have a reporter gene expressed from the gene's natural promoter, which is usually a faithful reproduction of actual expression, (Lé Mouellic et al. 1990; Mansour et al. 1990). This can be used to determine expression patterns in vivo and select those mouse lines showing gene expression in specific cell lineages. Finally, insertion of the gene trap retrovirus into actively expressed loci generates mutations in those genes. These mutations can be transmitted to the germline at a high frequency.

*Methods for cloning sequences flanking gene trap insertions*

Genes disrupted by gene trap selection must be characterized from sequences adjacent to the integrated provirus. The first step is to clone cellular flanking sequences, which are then used as probes to screen cDNA libraries. We have used several techniques for cloning 5' flanking sequences. Inverse PCR (iPCR) is used to clone genomic flanking sequences upstream of the provirus (von Melchner et al. 1990). Genomic DNA from the cell line of interest is digested with a restriction enzyme which cuts relatively frequently in the genome, such as *Hinfl*. The linear pieces of DNA are then circularized by ligating at a low concentration which encourages intramolecular ligations instead of intermolecular ligations. These circles are then digested with another enzyme (often *PvuII*) which cuts next to the 3' LTR so that only cellular sequences flanking the 5' LTR will be amplified. Primers to the known proviral sequences are used to amplify these sequences by the polymerase chain reaction (PCR). An alternative method is to clone cellular transcripts appended to the proviral transcript by 5' RACE (Rapid Amplification of cDNA Ends) (Frohman et al. 1988). Using a
primer specific for the reporter gene in U3, RNA from infected cell lines is reverse transcribed to synthesize cDNAs. A poly-C tail is then added to the cDNA, and the tailed cDNA is amplified by PCR using a nested U3 primer and a universal primer which can bind to the poly-C tail. Both of these methods are diagrammed in the "Methods" section of Chapter 4. If neither of these methods produces suitable probes, exon trapping can be used to identify potential exons in the flanking regions (Buckler et al. 1991).

Several of the new vectors have additional features designed to make the cloning of cellular flanking sequences easier. U3βgeoSupF, for instance, includes a bacterial supF (amber suppressor) gene in the internal sequences. Using a restriction enzyme which does not cut within the retrovirus, the entire provirus and flanking sequences can be excised from genomic DNA and ligated to phage arms. The recombinant phage can then be used to infect amber mutant bacteria, which will only grow if infected with a phage particle containing the supF gene. Another new vector constructed by Geoff Hicks in the lab incorporates a shuttle vector design. This gene trap uses a promoterless neo gene in U3 as the selectable reporter gene. The body of the provirus contains a pBR322 plasmid origin, an ampicillin resistance gene, and the lac operator (lac O) sequence. Excision of the provirus and intramolecular ligation yields a plasmid carrying the 5' LTR and 5' flanking sequences (termed plasmid rescue). lac O permits partial purification of flanking sequences by selecting for sequences which bind to the lac repressor protein. This is an extremely efficient process; in the past year, it has been used to isolate and sequence over 500 insertion sites in ES cells (Hicks et al. 1994)

Development of an in vitro screen

As we were interested in studying genes that are regulated during early development (and thus potentially involved in early developmental events), we wanted to find a way to enrich for these genes before constructing germline chimeras. Two potential methods existed for isolating genes regulated during early embryogenesis. The first method, which was already in use in several laboratories, involved generating chimeric animals from ES cell clones containing activated gene traps (Gossler et al. 1989; Friedrich and Soriano 1991; Johnson and Mahon 1993). Chimeric embryos are then sacrificed and expression of the transgene is analyzed. The advantage of such a screen is that expression of the
transgene is observed in the proper environment, assuming that most of the embryo is generated from the donor ES cells. However, it is only possible to look at a few time points, generally later in development, unless numerous animals are generated. This is due to the difficulty in isolating embryos younger than 8 days. In addition, chimera construction is relatively labor-intensive and expensive. Although new methods for generating chimeras have reduced the labor involved (Nagy et al. 1993; Wood et al. 1993a; Wood et al. 1993b), interesting clones still have to be reinjected into blastocysts to generate germline mice.

Because of these problems, I decided to use a screen which takes advantage of the ability of ES cells to form embryoid bodies in culture. Like embryonal carcinoma (EC) cells, ES cells can differentiate into a variety of cell types in vitro. When ES cells are plated at a high density in bacterial petri dishes, they form embryoid bodies, which have an inner layer of ES cells and an outer endoderm layer (Martin 1981). If allowed to attach to gelatinized tissue culture plates, these embryoid bodies can differentiate into a wide variety of cell types and tissues, including melanocytes, skeletal and smooth muscle, blood islands and beating heart structures (Doetschman et al. 1985). Previous studies indicated that, in a number of cases, expression of regulated genes in vitro accurately reflected in vivo expression. Mouse embryonic globin genes, for instance, are expressed in the correct temporal order in embryoid bodies, and further differentiation results in an appropriate switch to fetal/adult genes (Brown et al. 1987). Other examples of genes appropriately regulated in vitro include GATA-1, REX-1, murine H19 and gap junction genes (Hosler et al. 1989; Nishi et al. 1991; Poirier et al. 1991; Simon 1993).

In order to determine if embryoid bodies contain cell lineages from early development, Yamada et al. looked at several molecular markers for specific cell types during EB differentiation. They found expression of brachyury (a mesodermal marker) and Nkx-1.1 (a homeobox gene expressed in neuroectoderm) in subsets of non-endodermal cells. Furthermore, expression of brachyury and Pax-3 (which is involved in somitogenesis and neural differentiation) could be induced by appropriate factors (Activin A and basic FGF, and NGF, a neurotrophic factor, respectively), indicating that cells in EBs are able to respond to external signals in a physiologically relevant manner (Yamada et al. 1994). Additional studies in EC cells also indicate that brachyury is transiently expressed in cells which differentiate into mesoderm-derived cells, including
skeletal muscle (Vidricaire et al. 1994). Thus, embryoid bodies will differentiate into early cell lineages before being committed to more specific cell types.

This evidence indicated that a screen utilizing embryoid bodies would likely be an effective way to enrich for genes regulated during early development. Preliminary experiments by Sita Reddy in the lab confirmed this hypothesis. In three cases, *in vitro* expression of an activated U3-LacZ provirus was a good indicator of *in vivo* gene expression (Reddy et al. 1992). One problem with the U3-LacZ gene trap, however, is that the thymidine kinase promoter driving the internal Neo gene is inactive in ES cells. In order to isolate individual trapping events in ES cells, multiple rounds of Fluorescence Activated Cell Sorting (FACS) were required (Reddy et al. 1992), a time-consuming and expensive endeavor. Thus, I constructed two new LacZ gene trap constructs which contained functional Neo genes to allow direct selection of clones with integrated proviruses. The first trap, U3LacZpgkNeo, consisted of a simple replacement of the tk promoter with the phosphoglycerate kinase-1 (pgk-1) promoter, which is active in ES cells (Adra et al. 1987). The second trap construct utilizes a LacZ-Neo fusion gene, βgeo (Friedrich and Soriano 1991), as the selectable reporter gene in U3. Thus, proviral integrations that occur in active genes can be directly selected for in G418-containing media. In addition, X-Gal staining provides a general indication of gene expression whereby weakly expressed genes stain white or light blue and genes expressed at higher levels stain blue. This allows one to screen not only for genes that are repressed upon differentiation (blue to white), but also for those that are expressed at higher levels (white to blue).

The general scheme for isolating regulated genes using an embryoid body based assay is diagrammed in Figure 1.2. ES cells are infected with a trap containing the LacZ-Neo fusion gene, βgeo. Integrations which result in the expression of βgeo are selected for in G418-containing medium, and all NeoR clones are tested for X-Gal expression. All clones are then allowed to differentiate by plating in suspension culture on bacterial petri dishes for 5 days in media lacking differentiation inhibiting factors. The resulting cystic embryoid bodies are replated on gelatinized plates and allowed to further differentiate an additional 4 days. The time point of nine days was chosen for two reasons. The first reason was to ensure that changes in RNA levels were also reflected by changes in the amount of βgeo protein in the cells. Secondly, because we were interested in events during early embryogenesis, we did not want to allow differentiation to proceed for too long. Many terminally differentiated tissues
Figure 1.2

EXPERIMENTAL STRATEGY

Infect ES (D3) cells with U3βgeoSupF

Select in G418

Pick colonies and stain with X-Gal

Expand cultures for freezing

Allow cultures to differentiate *in vitro*

1. Grow cells in suspension culture in DME/10% FCS
2. After 5 days, replate embryoid bodies on gelatinized plates and culture for four days
3. Stain cells with X-Gal and pick those that have changed βgal expression (blue to white or white to blue)
4. Make transgenic mice with regulated lines
   - Look at *in vivo* staining patterns
   - Clone out flanking sequences and identify gene (if known)
(such as melanocytes, neurons, and cartilage) appear after 2-3 weeks in culture, whereas the complex embryoid bodies seen after 9 days in culture exhibit various early cell lineages, including ectoderm, endoderm, mesoderm and their derivatives. These complex embryoid bodies are then stained with X-Gal. Clones exhibiting regulated expression of the LacZ transgene are then chosen for further analysis and injection into blastocysts to generate chimeric animals.

**Genes regulated during development**

Theoretically, we should be able to isolate most genes that are expressed in early embryos. An important consideration for devising any genetic screen is what kinds of genes will be isolated. Our screen can pick up two types of regulated genes; those that are restricted or shut off upon differentiation (blue to white) and those whose expression increases (white to blue). There are a number of examples of genes which are down-regulated upon differentiation *in vitro* and *in vivo*; these include Oct3 (Rosner *et al.* 1990; Schöler *et al.* 1990), REX-1 (Hosler *et al.* 1989; Rogers *et al.* 1991) PEA3 (Xin *et al.* 1992), Fgf-4 (Ma *et al.* 1992; Niswander and Martin 1992) and the $\beta$ subunit of activin (Albano *et al.* 1993). It is possible that the expression of these genes in ES cells and blastocysts is simply due to a partial state of derepression, as has been observed in Xenopus embryos after the mid blastula transition [Rupp and Weintraub, 1991, cited in (Kafri *et al.* 1992)]. It has been previously shown that the amount of methylation in the preimplantation mouse embryo is very low, and is subsequently increased via *de novo* methylation sometime after implantation (Monk *et al.* 1987; Kafri *et al.* 1992). This could result in a loosening of transcriptional control.

It is also possible, however, that many of the genes expressed in the ICM and ES cells do have specific functions. For instance, two of the genes cited above (*Fgf-4* and the $\beta$-activin subunit) are implicated in the induction of mesoderm, one of the first determinative events in the developing embryo [reviewed in (New 1991; Kessler and Melton 1994)]. *Fgf-4* (kFGF), which contains an octamer motif, is one of the first members of the FGF family to be expressed during development. *In vivo*, *Fgf-4* is first detected in late blastocysts (approx. day 4.5). After implantation, its expression is restricted to the primitive streak, where mesoderm and definitive endoderm form. At day 10, *Fgf-4* is restricted to the tail bud, which is the primary source of mesoderm at this stage.
The expression of the β subunit of activin is also repressed upon differentiation both *in vitro* and *in vivo* (Albano *et al.* 1993). This subunit is expressed in all cells of the preimplantation embryo until the blastocyst stage, where it is only expressed in the inner cell mass. By 4.5 days, however, expression disappears in the ICM and reappears in the trophectoderm. Preliminary data suggested that it was not expressed in day 6.5 embryos, but was expressed in some of the surrounding decidual cells. Very little is known about endogenous inductive events in vertebrate embryogenesis. It is quite probable that a "combinatorial action of inducers, having both redundant and antagonistic functions, underlies the regional specification of cell fate" [(Kessler and Melton 1994), p. 603]. Therefore, a screen which isolates developmentally regulated genes could potentially identify novel genes which are involved in early inductive events.

One subset of these genes might include "competence" factors. Competence is defined as the ability of a cell to respond to specific inductive signals in an appropriate manner. Thus, genes that are involved in determining the competence of a particular cell would allow that cell to respond to morphogens such as FGF and activin. As cells become more specified, some competence genes would be down-regulated, whereas others might be activated, to allow for more specific differentiation pathways. Experiments in *Xenopus*, for instance, have shown that the competence of ectoderm to respond to basic FGF changes with time (Kengaku and Okamoto 1993). Early ectoderm is induced to form mainly neurons of the central nervous system (CNS), but with increasing age, the ectoderm becomes less competent to form neurons and forms melanophores instead. The change in response in ventral ectoderm precedes that in dorsal ectoderm, which could explain the regional specification of ectoderm into different lineages (neural tube vs. neural crest). Competence appears to be due to intrinsic rather than extrinsic factors, as the same response is seen *in vitro*. Since ES cells are totipotent, multiple competence genes could be expressed, each participating in different developmental fates. Commitment of the stem cell to differentiation could restrict expression of the genes to cells of the appropriate type. Early (or primary) competence genes would be able to participate in early development without needing to be induced. Competence genes could include receptors for the various peptide factors implicated in inductive events. Cells expressing receptors with different affinities for inductive signals could respond in diverse ways resulting in the formation of tissues with
altered developmental programs (i.e. dorsal and ventral mesoderm, which differentiate into completely different tissues) (Melton 1991; Jessell and Melton 1992). Other components of the signal transduction pathway might act as competence modifiers, which modulate specific cellular responses to developmental signals (Moon and Christian 1992).

Another set of genes could be required for the maintenance of pluripotency in ES cells. Such genes might be actively involved in preventing cellular differentiation until the appropriate time, perhaps by acting as transcriptional repressors. Alternatively, these genes might allow ES cells to respond to a variety of different signals, thus acting as general competence genes. An example of a pluripotency gene could be the POU-domain transcription factor oct-3 which is linked to pluriotypy in vitro and in vivo (Rosner et al. 1990; Schöler et al. 1990). Oct-3 expression is consistently decreased as cells become committed to differentiated lineages. In vitro, it is expressed in ES and EC cells, but not in embryoid bodies. In vivo, oct-3 is one of the first homeodomain proteins to be expressed in the developing embryo; it is expressed in early embryos through the blastocyst stage, and in primitive cell lineages until day 8.5. After this stage, expression is only seen in germ cells. As very little is known about any of the above processes, it is essential that new genes involved in these functions be isolated. Our screen could be one method for isolating such genes.

Other genes which are repressed upon differentiation might be involved in more specific functions. For instance, REX-1 (Zfp-42) is a zinc finger gene which was cloned because its expression was reduced upon retinoic acid (RA) induced-differentiation in EC cells. Its promoter region contains an octamer motif (the binding site for POU-domain proteins) which appears to be required for negative regulation by RA (Hosler et al. 1993). In vivo, REX-1 is expressed in the ICM of preimplantation (day 3.5-4.5) embryos, but is limited to trophoblast-derived tissues shortly thereafter. In adult mice, REX-1 is only expressed in spermatocytes (Rogers et al. 1991). Based on expression patterns, it has been hypothesized that Rex-1 is involved in trophoblast development and spermatogenesis.

Genes whose transcription is very weak in pluripotent cells but increased upon differentiation would be scored as white to blue in our assay. These might include so-called housekeeping genes, which would be transcribed at higher levels upon differentiation and morphogenesis as cells require more "supplies"
for the rapid growth of the embryo. Other genes might encode gene products required for differentiation and growth or even for specific morphogenetic events, either of which would be very interesting. Two kinds of genes whose levels are increased upon differentiation are those involved in cell cycle functions and gap junction formation. The mouse homologue for the cdc25 mitotic inducer is expressed in EC cells, is RA-inducible and is widely expressed in differentiating tissue in the embryo (Kakizuka et al. 1992). Gap junction genes are also expressed in ES cells and their abundance increases with development (Nishi et al. 1991). As cells go through the complex processes of differentiation, morphogenesis, and organogenesis, gap junction communication is probably crucial. Indeed, blocking gap junction formation during development is associated with defects in embryo patterning [reviewed in (Guthrie and Gilula 1989)].

At the time this study was begun, it was impossible to predict whether or not genes involved in pattern formation could be identified in vitro. Although many different cell types are formed in embryoid bodies, there is no obvious dorsal-ventral or anterior-posterior patterning present. With the exception of what is known about homeodomain proteins, little is known about axial patterning in mammalian development. Therefore, it is conceivable that expression of some early axial determinants could be induced in the ICM prior to pattern formation and maintained only in cells that adopt the proper pattern. In vitro, this might be represented by expression in ES cells but repression in embryoid bodies, where the genes are not in the correct (patterned) environment.

Earlier reports indicated that transcription of several mouse homeobox genes expressed in EC cells increased upon in vitro differentiation (Colberg-Poley et al. 1985; Chavrier et al. 1988). In vivo, however, Hox genes are not expressed until after embryo implantation. In ES cells, very low levels of Hox genes controlled by a retinoic acid response element, or RARE, are observed. This is most likely due to some sort of inducing effect from serum in the medium, as de-lipidized serum greatly reduces the basal level of transcription of a number of RARE-containing genes (L. Gudas, personal communication). Thus, although expression of these RARE-containing genes is artifactual, proviral integration into such genes might result in clones exhibiting the white to blue phenotype observed in our screen.

There are several kinds of genes which could probably not be isolated in our embryoid body screen using U3βgeoSupF. Many genes, such as brachyury
(Vidricaire et al. 1994), nodal (Zhou et al. 1993) and snail (Nieto et al. 1992; Smith et al. 1992) are expressed after embryo implantation and thus could not be isolated via a screen which requires ES cell expression. However, such genes could be isolated by using the U3LacZpgkNeo gene trap. Libraries of NeoR, LacZ- clones can be generated which contain gene trap integrations into non-expressed regions of the genome. Differentiation of such clones and selection for activation of the LacZ reporter gene would allow one to identify genes which are only expressed in differentiated tissues.

Summary

To circumvent the difficulties inherent in studying mouse development, a number of laboratories have developed gene trap vectors which enrich for insertional mutations in ES cells. Because activation of reporter genes requires integration into or near expressed genes, gene traps produce mutations at a high frequency, approaching 100%. Unlike random insertion techniques, approximately 40% of insertions selected in this manner result in embryonic lethal mutations. Both novel and known genes have been identified in gene trap studies. Skarnes et al. identified three novel genes, including a zinc finger-containing protein (Skarnes et al. 1992). Chen et al. recently identified one of their gene trap lines to be an insertion into Transcription Enhancing Factor 1 (TEF1), which causes heart defects and embryonic lethality (Chen et al. 1994). Insertions with our AUG traps have identified insertions into a number of known genes, including the developmentally regulated gene REX-1, and the epithelial cell kinase, ECK, which is expressed in the mouse node (von Melchner et al. 1992; Chen and Ruley 1994). In addition, the gene originally designated fugl (DeGregori et al. 1994) is probably a GTPase activating protein for the G-Protein Ran (personal communication to H.E.R.). This confirms that interesting genes which are involved in early embryogenesis can be isolated by gene trap mutagenesis.

In this thesis, I have investigated the potential of an in vitro screen to enrich for genes regulated during early development. A retroviral gene trap construct utilizing the βgeo fusion protein as a selectable marker was used to infect totipotent ES cells. NeoR clones expressing varying amounts of the βgeo fusion transcripts (as assayed by X-Gal staining) were differentiated into embryoid bodies in culture. Clones exhibiting regulated reporter gene expression
in vitro were used to generate chimeric mice. In every case, expression of the reporter gene in vitro accurately predicted in vivo expression (including REX-1 and ECK described above). For instance, genes expressed in ES cells but repressed in EBs were expressed in blastocysts but repressed in post-implantation embryos. Conversely, genes expressed at low levels in ES cells but induced upon differentiation showed wide-spread expression in post-implantation embryos, but little to no expression in blastocysts. In addition, one out of the three insertions discussed in this thesis resulted in an embryonic lethal phenotype in homozygous mutants. In summary, this in vitro screen provides a viable alternative to both chimera-based screens and cDNA subtraction techniques for the isolation of genes involved in early embryonic development.
Chapter 2:

Development of New Retroviral Gene Traps
Introduction

The goal of this thesis project was to develop an *in vitro* screen to isolate and mutagenize developmentally regulated genes in mouse embryonic stem (ES) cells. Several conceptual and technical developments made the undertaking possible. First, retroviral gene trap constructs containing promoterless coding sequences for reporter genes in the LTR of Moloney Murine Retroviruses had been constructed in our lab (von Melchner and Ruley 1989). Upon integration of the provirus, the reporter gene is placed only 30 nucleotides from flanking cellular DNA. Activation of the reporter gene typically requires that the retrovirus integrate into or near 5' exons of expressed genes (von Melchner and Ruley 1989; Chen et al. 1994). In addition to providing a measure of gene activity, insertion of the retrovirus into 5' sequences usually disrupts gene function completely (Gossler et al. 1989; von Melchner et al. 1990; Friedrich and Soriano 1991; Skarnes et al. 1992; von Melchner et al. 1992). Second, lines of embryonic stem cells had been developed that not only maintain the ability to transmit genes into the mouse germline but are also capable of differentiation in culture. The first step of the *in vitro* screen consisted of identifying genes that are expressed in ES cells by using retroviral gene traps as insertional mutagens. Clones containing integrations in active genes were allowed to differentiate *in vitro* by removing differentiation-inhibiting factors. Those clones exhibiting regulated expression of the viral reporter gene were used to generate transgenic mice, where gene expression and function could be investigated *in vivo*.

In order to make this screen as efficient as possible, there were several requirements for the gene trap retrovirus to fulfill. The first was that selection of individual clones had to be simple and efficient. In addition, a reporter gene which could clearly indicate changes in gene expression was necessary. High titer virus stocks were an advantage, as was the ability to select both for and against viral gene expression (to isolate expressed and non-expressed clones). A logical candidate for a reporter gene with which one could monitor changes in gene expression was LacZ, which had previously been incorporated into the U3-LacZ gene trap vector (Reddy et al. 1991). This vector contains an internal Neomycin transferase (Neo) gene under the control of a thymidine kinase (tk) promoter which is inactive in ES cells. Thus, isolation of ES cell clones expressing LacZ fusion genes required multiple rounds of Fluorescence
Activated Cell Sorting (FACS), a time-consuming and expensive endeavor (Reddy et al. 1992). A high priority, then, was to design new gene trap constructs which contained functional Neo genes as well as a LacZ reporter gene. This chapter describes the development and testing (in 3T3 cells) of two new vectors, U3LacZpgkNeo and U3βgeoSupF.

Construction of pU3LacZpgkNeo consisted of a simple replacement of the tk promoter from U3LacZ with the phosphoglycerate kinase (pgk-1) promoter (Adra et al. 1987), which is active in ES cells. Therefore, all clones containing integrated proviruses will acquire NeoR. To select for gene trap events, colonies on plates can be stained with the live stain for β-galactosidase, fluorescein-di-galactosidase (FDG). An advantage of this approach is the ability to select directly both for and against LacZ expression without altering cell viability. Individual βgal positive and negative colonies can be picked after selection in G418, or populations of FDG positive and negative cells can be isolated using the cell sorter a few days post infection. These populations can then be screened for regulated expression of the reporter gene.

U3βgeoSupF utilizes a βgal-neo fusion construct, βgeo (Friedrich and Soriano 1991), which replaces most of the LacZ coding sequences in U3LacZ. The advantage of this vector is that very few manipulations are required to isolate gene trap integrations into expressed genes. Theoretically, all G418R colonies contain integrations into active genes. Indeed, all of the 3T3 clones tested (and subsequently, all ES clones as well) expressed fusion transcripts originating in cellular sequences upstream of the provirus. One drawback of this vector is that one cannot generate a null population of clones which have proviruses integrated into inactive chromosomal sites.

The data in this chapter indicate that both U3LacZpgkNeo and U3βgeoSupF are efficient insertion vectors. However, due to some uncertainties concerning the U3LacZpgkNeo vector (described in this chapter), the latter vector was used for the in vitro screen described in Chapter 3. U3βgeoSupF is an efficient gene trap which allows one to select for insertions downstream of both strong and weak promoters. This is an ideal vector for the in vitro screen which we have utilized.

All of the work described in this chapter was performed by myself. Sita Reddy and Harald von Melchner suggested how the new constructs should be made. Sita helped me with some technical aspects of the cloning and RNase
Protection assays. In addition, she kindly provided me with control blue and white RNA samples to use in the RNase Protection assays.
Results

Two new gene trap retroviruses were constructed in order to facilitate an in vitro screen for developmentally regulated genes. Before being used in the screen, both vectors were tested in 3T3 cells to ensure that they were indeed acting as gene trap vectors. This section describes the construction and testing of the two new gene traps in 3T3 cells.

Construction and testing of U3LacZ-pgkNeo

pU3LacZ-pgkNeo was constructed from pGgTKNeolacZen(-) (BG2) by replacing the thymidine kinase promoter with the phosphoglycerate kinase (pgk-1) promoter, which is active in ES cells. Cell lines producing the U3LacZpgkNeo retrovirus were generated by transfecting ψ2 ecotropic packaging cells (Mann et al. 1983) with the plasmid. Ten ψ2 clones were titered on NIH 3T3 cells to find a high-titer producer line. Four of the ten lines produced virus titers less than $1 \times 10^2$ colony forming units (cfu)/ml per $10^7$ producer cells and were discarded. The other six lines exhibited titers ranging from $8 \times 10^3$ to $2.7 \times 10^5$ cfu/ml per $10^7$ producer cells. The highest titer virus producer line was pgkΨ8, which was used in subsequent experiments.

In order to test whether U3LacZ-pgkNeo was acting as a gene trap, 3T3 cells were infected with the virus and both general infection statistics and individual clones were analyzed. The frequency of proviral activation was estimated to be 1 in 200 integration events, which is consistent with the observed activation frequency of the U3LacZ retrovirus (Reddy et al. 1991).

Figure 2.1 Southern blot of DNA isolated from clones infected with U3LacZpgkNeo. 10 μg of genomic DNA was restriction digested with ClaI, fractionated on agarose gels, and transferred to nitrocellulose membranes. The blots were then hybridized to a LacZ probe. (A). Diagram of a properly integrated provirus. (B). Autoradiogram of the Southern blot. Lanes 1-3 contain DNA from NeoR βgal- clones, and lanes 4-7 contain DNA from NeoR βgal+ clones.
Figure 2.1

A.  

B.  

6.9kb
This frequency was calculated by dividing the number of X-Gal positive colonies by the total number of viral integrations (Neo^R colonies). To ensure that acquisition of Neo^R was due to integrated proviruses, Southern analysis was performed on genomic DNA isolated from seven individual Neo^R clones (Figure 2.1). Panel A indicates the expected proviral structure and the size of the internal Clal restriction fragment. Genomic DNA samples were digested with Clal, fractionated on agarose gels, transferred to nitrocellulose membranes and hybridized to a LacZ probe. Lanes 1 through 3 contain DNA isolated from LacZ^-, Neo^R colonies, and lanes 4 through 7 contain DNA isolated from LacZ^+, Neo^R colonies. All seven lanes show a 6.9 kb band indicative of a correctly integrated provirus.

Northern analysis and RNase Protection assays on RNA from the same seven clones showed that the Neo transcript driven by the pgk promoter was expressed at high levels in all seven clones (Figure 2.2). Panel A diagrams the possible transcripts originating outside and within integrated proviruses. Transcripts originating in the pgk promoter and terminating in 3' U5 sequences will be 5.4 kilobases (kb) in length, whereas transcripts originating upstream of the provirus will vary in length but should be at least 3.6 kb. Panel B shows a Northern blot on the seven clones depicted in Figure 2.1 (B). As before, lanes 1-3 contain RNA from LacZ^-, Neo^R clones and lanes 4-7 contain RNA from LacZ^+, Neo^R clones. There is a strong band at 5.4 kb in all seven lanes which corresponds to the pgkNeo message. In addition, there are bands approximately 4kb in size in lanes 4 through 7, indicating possible fusion transcripts from flanking DNA. However, there also appear to be very faint but similarly sized bands in lanes 2 and 3.

**Figure 2.2** Analysis of RNA isolated from clones infected with U3LacZpgkNeo. (A) Diagram of possible proviral transcripts. Stippled bars indicate fragments protected in RNase Protection assays. (B) 10μg of total RNA from each clone were fractionated on 1% formaldehyde-agarose gels and transferred to nitrocellulose membranes. The blots were then hybridized to a LacZ probe. Lanes 1-3 and lanes 4-7 contain RNA from the same clones as in Figure 2.1. (C) 30 μg of total RNA was hybridized to a 689 nt LacZ anti-sense riboprobe. Bands at 643 nt represent protected fragments from RNA extending through the 3' LTR, whereas bands at 501 nt represent protected fragments from fusion transcripts initiating in 5' cellular sequences. Lanes 1-7 are the same as in (B). Lanes 8-11 contain the following controls: 8: blue ES clone, 9: blue 3T3 clone, 10: white 3T3 clone, and 11: yeast tRNA.
RNase Protection analysis on the same samples also detected both internal and cellular fusion transcripts. Figure 2.2 (C) shows results from a typical RNase Protection assay. The expected sizes of protected fragments are indicated in panel A by the stippled bars. The LacZ probe will protect a 643 nt fragment from internal transcripts through the 3' LTR, whereas a 501 nt fragment will be protected from transcripts initiating upstream of the provirus. Lanes 1-3 and 4-7 again represent the same clones. All 7 show a very strong band at 643 nt which corresponds to the protected fragment from the 3' LTR. In addition, all lanes except for lane 1 have bands at 501 nt. If lanes 2 and 3 represent truly LacZ- clones, then there should be no band at that position. However, this result is consistent with the presence of faint transcripts in the Northern blot. Lanes 8 through 11 contain control RNA samples. Lane 11 contains yeast tRNA, to control for nonspecific hybridization of the riboprobe. Lane 8 contains RNA isolated from a blue ES cell clone infected with U3LacZ (Reddy et al. 1992). Note that there is very little of the 643 nt protected fragment corresponding to the 3' transcripts. This is because neither tkNeo nor the viral promoter in U3 are expressed in ES cells. Thus, any protected fragment present represents transcripts which extend through the entire retrovirus. Lanes 9 and 10 contain RNA from 3T3 cells also infected with U3LacZ; lane 9 is a blue clone and lane 10 is a white clone. There is a clear difference between the blue and white clones in the amount of the 501 nt protected fragment, which is not readily apparent in my clones. As there was a chance that my clones were not clonal, I subcloned a number of putative white clones by fluorescence activated cell sorting. Upon repeating the RNase protection, I again observed bands at the 501 nt position (data not shown), although these were very faint. One possible explanation is that the bands represent breakdown products of the 643 nt fragment, as the 32P-labeled probes are very labile; lane 12 shows a dilution of the unhybridized probe; although it was stored at -20°C for only 24 hours (the time period of the experiment), a significant amount of breakdown is evident. In lane 1, a faint band can be seen at the 501 nt position which is probably a breakdown product. In those lanes with more of the 643 nt protected fragment, there is likely to be more of this breakdown product. I did not feel confident that these bands were indeed indicative of transcripts originating outside of the provirus and not just a side effect of having an extremely strong promoter driving the internal neo gene. Reversing the orientation of the internal pgkNeo was considered, but this may have inhibited viral transcription and titers due to
excessive transcription in the opposite direction. Because of these uncertainties, I did not pursue the use of this gene trap vector.

Construction and testing of U3\(\beta_{geo}\)SupF

The pU3\(\beta_{geo}\)SupF plasmid was constructed using the \(\beta\)-galactosidase-Neomycin transferase fusion gene (\(\beta_{geo}\), gift of Philippe Soriano) and BG2. In brief, BG2 Lac\(Z\) sequences from the ClaI to NheI sites were replaced with \(\beta_{geo}\) sequences, leaving the Shine-Delgarno (Shine and Dalgarno 1974; Steitz and Jakes 1975) consensus ribosome binding site and the Kozak (Kozak 1986) AUG site intact. In addition, the bacterial amber suppressor gene, sup\(F\) (Reik et al. 1985), was introduced in place of tkNeo for the purpose of cloning cellular flanking sequences. Ecotropic retrovirus producer lines were generated by transfecting \(\Psi\)2 cells with pU3\(\beta_{geo}\)SupF. Ten Neo\(^R\) \(\Psi\)2 lines were assessed for virus titer. Because there is no independent selectable marker in U3\(\beta_{geo}\)SupF, only clones containing integrations which activate the proviral reporter gene will be Neo\(^R\). Therefore, the titers reported here are approximately 200 fold less than the actual viral titer (if we assume that the trapping frequency is similar to that of U3-Lac\(Z\)). Although the viral LTR contains a 4 kb insertion, high titer producer lines were obtained. Three produced virus with titers less than \(10^2\) cfu per \(10^7\) producer cells. Virus from the other seven ranged in titer from \(7 \times 10^3\) to \(4.2 \times 10^4\) cfu/ml per \(10^7\) producer cells. The two highest titer producer lines, \(\beta_{geo}\)\(\Psi\)2 and \(\beta_{geo}\)\(\Psi\)8, had titers of \(3.4 \times 10^4\) and \(4.2 \times 10^4\), respectively. These two producer lines were used in subsequent experiments.

To ensure that U3\(\beta_{geo}\)SupF was functioning as a gene trap, both producer lines were first tested in 3T3 cells. Although all Neo\(^R\) clones should be expressing \(\beta_{gal}\) as well, 30\% of NIH 3T3 colonies infected with \(\beta_{geo}\)\(\Psi\)2 and 35\% of colonies infected with \(\beta_{geo}\)\(\Psi\)8 were white in X-Gal staining assays. Southern blot analysis indicated that all clones did have integrated proviruses (Figure 2.3).

**Figure 2.3** Southern blot of DNA isolated from 3T3 clones infected with U3\(\beta_{geo}\)SupF. Genomic DNA was prepared as in Figure 2.1. Blots were hybridized to a Neo probe. (A) Diagram of a properly integrated provirus. (B) Autorad of the southern blot. Lanes 1-4 contain DNA from clones infected with \(\beta_{geo}\)\(\Psi\)2 and lanes 5-6 contain DNA from clones infected with \(\beta_{geo}\)\(\Psi\)8. Lane 7 contains DNA from uninfected 3T3 cells.
Figure 2.3

A. Cla1-6.1kb

B. 6.1kb

1 2 3 4 5 6 7
Panel A depicts the expected proviral structure with the size of the internal *ClaI* fragment shown. Panel B shows a Southern blot of genomic DNA digested with *ClaI* and hybridized to a Neo probe. Lanes 1 through 4 contain DNA isolated from clones infected with βgeoΨ2 and lanes 5 and 6 contain DNA isolated from clones infected with βgeoΨ8. Lane 7 contains DNA from uninfected 3T3 cells. It can be seen that lanes 1 through 4 exhibit bands at 6.1 kb indicative of properly integrated proviruses, whereas samples 5 and 6 contain differently sized bands. This was seen with other clones infected with βgeoΨ8 as well (data not shown). It was assumed that the retroviral construct in this producer line had been rearranged or mutated during transfection and was producing defective retroviruses. As a result, virus from the βgeoΨ2 producer line was used exclusively in further experiments.

Analysis of RNA isolated from the same 3T3 clones is shown in Figure 2.4. Panel A indicates the expected transcripts and protected fragments from proviruses integrated into active genes. Bands larger than 4.3 kb are expected for transcripts originating outside the provirus, whereas bands approximately 6 kb in length are expected for transcripts originating in the 5' U3 region. Panel B shows a Northern blot on RNA from the clones infected with βgeoΨ2 shown in Figure 2.3. Lanes 1 through 4 show transcripts around 4 kb, whereas lane 5, the uninfected 3T3 control, shows no transcripts. RNase Protection Analysis on the same clones confirmed that NeoR in these clones was due to fusion transcripts originating in flanking cellular DNA. Lanes 1 through 4 contain RNA isolated from clones infected with U3βgeoSupF and Lanes 5 through 7 contain the same controls as Figure 2.2 (C). Lane 5 contains RNA from a blue ES clone, lane 6 contains RNA from a white 3T3 clone, and lane 7 contains tRNA. All four clones infected with U3βgeoSupF show strong bands at 501 nt corresponding to

---

**Figure 2.4** Analysis of RNA isolated from clones infected with U3βgeoSupF. RNA was prepared as in Figure 2.2. (A) Diagram of possible transcripts detected by a Neo probe. Stippled bars indicate fragments protected in RNase Protection assays. (B) Northern blots were hybridized to a Neo probe. Lanes 1-4 contain RNA from clones infected with βgeoΨ2 and lane 5 contains RNA from uninfected 3T3 cells. (C) RNase Protection analysis. Lanes 1-4 are the same as in (B). Lanes 5-7 contain the following controls: 5: blue ES clone, 6: white 3T3 clone, and 7: yeast tRNA.
Figure 2.4

A. 501 nt

B. 643nt -

C. free probe ---

1 2 3 4 5

18S

28S

1 2 3 4 5 6 7 8

643nt

501 nt
protection of a 5' fusion transcript. 3' protected fragments are also present, but are either equal to or weaker in intensity than the 5' transcripts.

These experiments indicated that U3βgeoSupF is an efficient gene trap vector in NIH 3T3 cells. The RNase Protection data obtained from U3βgeoSupF was much more clear-cut than data obtained with U3LacZpgkNeo. In addition, selection of clones with gene trap insertions in expressed loci was more simple than with the other gene trap. An unexpected result was that approximately 30% of the NeoR clones were white. I hypothesized that these clones had integrated into expressed genes, but that expression of the reporter gene was not sufficient to stain well with X-Gal. This gave us the opportunity to not only look at genes being shut off upon differentiation, but to also look at genes whose expression is increased.
Discussion

Before proceeding with the *in vitro* embryoid body screen, new gene trap technology had to be developed and tested. The requirements for the new trap were that it had to be an efficient targeting vector, have a relatively high titer in order to facilitate infection of many cells, and integrations which allowed expression of the reporter gene had to be easily isolated. I constructed and tested two new retrovirus gene trap vectors, U3βgeoSupF and U3LacZpgkNeo, which both employ *Neomycin* and *LacZ* reporter genes. Using an internal *Neo* gene driven by the pgk-1 promoter, U3LacZpgkNeo allows selection for cells that contain integrated proviruses. However, trapping events must be separately screened for with the live stain, FDG. U3βgeoSupF, on the other hand, allows direct selection for trapping events in G418 by having a *LacZ-neo* fusion gene as the reporter. Both vectors have their advantages, but for the purposes of this study, I felt that U3βgeoSupF was a more effective vector.

The data collected with U3LacZpgkNeo indicated that it met most of the requirements for a new gene trap. I was able to isolate high titer producer lines and efficiently isolate LacZ expressing clones using the live stain FDG. Southern analysis of individual clones showed that proviruses were integrating normally. The average number of activating integrations was approximately 1/200, which is the same as that seen with U3-LacZ (Reddy *et al.* 1991). This number might not be entirely accurate for several reason. Firstly, RNase Protection analysis on white clones indicated that readthrough transcripts might be present in these clones. Initially, this may have been due to minor contamination by βgal+ cells. However, clones subcloned by cell sorting still exhibited faint bands at 501 nt. It is possible that these clones had proviral insertions into weakly expressed genes. However, it is more likely that the band appearing at 501 nt is a break-down product of the larger, 643 nt protected fragment from the 3' LTR. The pgkNeo transcript is expressed at very high levels, such that even a small percentage of breakdown product might appear to be a legitimate band on an RNase Protection. One possible way around this would have been to place the pgk-Neo cassette in the opposite orientation to viral transcription. However, this may have disrupted viral transcription, resulting in lower virus titers.

By all molecular criteria, U3βgeoSupF appeared to be acting as a gene trap vector in NIH 3T3 cells. Proviruses were integrating normally, and all NeoR clones exhibited readthrough transcripts originating in cellular flanking sequences
5' of the provirus. Although these transcripts were difficult to see in Northern blots, they were obvious in RNase Protection assays. The intensity of the 501 nt protected fragment on autorads was equal to or greater than the intensity of the internal 643 nt protected fragment, indicating that there were at least equal amounts of fusion transcript compared to internal transcripts. Unlike clones containing U3LacZpgkNeo, there was a significant difference in the intensity of the 5' and 3' protected fragments. Although U3βgeoSupF producer lines tended to have lower titers than U3LacZpgkNeo, this was to be expected. The only way to obtain NeoR is for the retrovirus to integrate into 5' introns or exons of actively transcribed genes, as there is no internal drug resistance marker. The highest titer of 2 x 10^4 cfu/ml is reasonable when one assumes that the average activation frequency for LacZ gene trap vectors in our hands is 1/200, yielding a potential titer of 4 x 10^6 cfu/ml. As seen in Figure 2.3, the highest titer producer line, βgeoΨ8, produced aberrant proviruses. Since all of the proviruses were rearranged, it is probable that the plasmid underwent some mutation or rearrangement during transfection, causing a loss of at least one ClaI site. As the other high titer producer line, βgeoΨ2, did not produce any apparently abnormal proviruses in 3T3 cells, it was used in all subsequent infections, where it consistently had a high titer on NIH 3T3 cells.

Approximately 30% of the NeoR clones infected with U3βgeoSupF appeared white in X-Gal staining assays. There are at least two possible explanations for this. The first would be that NeoR was the result of translation of 3' LTR sequences from the transcript originating in U3 of the 5' LTR. This is unlikely, as white and light blue clones did show small amounts of cellular-proviral fusion transcripts in RNase Protection Assays (data not shown). These apparent fusion transcripts could not be the result of breakdown of the larger protected fragment, as there is little of this transcript present in cells infected with U3βgeoSupF. This would certainly not be a problem in ES cells, where the viral promoter is inactive (Jaenisch and Berns 1977; Teich et al. 1977). Another possibility is that white clones contain retroviral insertions into genes that are expressed at low levels that are sufficient to impart NeoR, but not sufficient to allow staining by X-Gal. Experiments with ES clones corroborate this hypothesis (see next chapter).

In summary, both U3LacZpgkNeo and U3βgeoSupF appeared to fulfill the requirements for new gene trap vectors. However, strong expression of the internal pgkNeo in clones infected with U3LacZpgkNeo caused a great deal of
background on RNase Protection Assays. This made it difficult to determine if white clones were the result of integrations into weakly expressed genes or just random integrations into non-expressed regions. In contrast, RNase Protections with RNA from clones infected with U3βgeoSupF showed a clear band at 501 nt indicative of cellular-viral fusion transcripts. As the in vitro screen required being able to show that genes were regulated, such unequivocal data was preferable. In addition, integrations of U3βgeoSupF into weakly expressed genes resulted in NeoR clones which did not stain with X-Gal. Thus, although we were unable to do a screen with a library of non-expressing integrations, as would have been possible with U3LacZpgkNeo, we were able to screen for genes whose expression was increased upon differentiation. For these reasons, it seemed appropriate to use the βgeo gene trap construct instead of U3LacZpgkNeo for the in vitro embryoid body screen.
Materials and Methods

Plasmid Construction

\textit{pU3LacZpgkNeo:} U3LacZpgkNeo was derived from the (BG2) plasmid pGgTKNeolacZen(-) (Reddy \textit{et al.} 1991). To remove the thymidine kinase promoter, BG2 was linearized with \textit{XhoI} and the overhang was filled in with DNA polymerase large (Klenow) fragment. BG2/XhoI was then cut with \textit{BglII} to excise the TKneo cassette and 5' phosphates were removed with calf intestine phosphatase (CIP). The phosphoglycerate kinase (pgk-1) promoter was kindly supplied by Betsy George in the Hynes lab. pPGK was linearized with \textit{XbaI} and overhangs were filled in with Klenow polymerase. The resulting fragment was cut with \textit{BglII} to excise the promoter region, which was ligated to BG2 in the proper orientation 5' of the neomycin transferase gene.

\textit{pU3\betageoSupF:} pU3\betageoSupF was derived from BG2 and p\betageo, the gift of P. Soriano (Friedrich and Soriano 1991). BG2 LacZ sequences from \textit{ClaI} to \textit{NheI} were replaced by the \textit{ClaI} to \textit{XhoI} \betageo fragment. To remove the internal TKneo gene, the resulting pU3\betageo-TKneo plasmid was partially digested with \textit{BamHI}, and the 1.4 kb Neo insert was replaced with the 200 bp \textit{Sau3AI} fragment of the bacterial amber suppressor gene, \textit{supF} (Reik \textit{et al.} 1985). The resulting vector contains the \betageo fusion gene in U3 with the translational signals in the 5' portion of LacZ identical to those in BG2.

In both vectors, LacZ was originally derived from pSDKLacZ (Darling and Rossant), which contains a Shine-Delgarno (Shine and Dalgarno 1974; Steitz and Jakes 1975) consensus ribosome binding site and a Kozak (Kozak 1986) consensus AUG sequence.

Construction and titering of \textit{Ψ2} Virus Producer Lines

\textit{Cell Culture:} NIH 3T3 cells and \textit{Ψ2} (Mann \textit{et al.} 1983) cells expressing a packaging-defective ecotropic helper virus were cultured in DME supplemented with 10% calf serum, 10 units of penicillin per ml and 10 \textmu g of streptomycin per ml. When appropriate, G418 (Geneticin, Gibco) was added to a concentration of 1 mg/ml. All cells were grown at 37°C in a humidified incubator containing 5% CO\textsubscript{2}.
Transfection of Ψ2 cells: DNA precipitates of either pU3LacZpgkNeo or pU3βgeoSupF were made using a standard calcium phosphate protocol (Maniatis et al. 1982). 5 x 10^5 Ψ2 cells were seeded the evening before transfection. The DNA precipitate was added to cells for 20 minutes, and the cells were "shocked" with media containing 100 μM chloroquin. After four hours, the cells were washed with PBS, normal medium was added, and the cells were cultured for 48 hours. Plates were split 1:10 into culture medium supplemented with G418. Individual Neo^R^ clones were picked after 10-14 days of selection and expanded for titering.

Titering of producer cell lines: 2 x 10^6 producer cells and 1 x 10^5 3T3 cells were seeded onto 10 cm tissue culture dishes the evening before infection. The following morning, the media on the producer cells was replaced with 2 ml of fresh media (no G418). Virus was collected for two hours at 37° with occasional rocking. Serial dilutions of the viral supernatant were filtered through a 0.45 micron syringe filter and 1 ml was added to each plate of 3T3 cells with 8 μg/ml polybrene. Infection was carried out for 1 hour at 37° with gentle rocking every 15 minutes. After 1 hour, 9 ml of fresh media (no G418) was added. The following day, G418-containing media was added, and plates were cultured for 10-14 days, with media changes every 3 days. To assess virus titer, the dishes were washed with PBS, fixed in 10% formaldehyde/PBS for 20 minutes, and stained with crystal violet (0.1%, in 50% EtOH: 50% H_2O) for 10 minutes. Stained colonies were carefully rinsed in cold H_2O and the plates were allowed to dry. The number of crystal violet stained colonies was then counted and titers were calculated as the average of 3 or 4 plates.

Infection of target cells

Producer lines with the highest titer (pgkΨ8 and βgeoΨ2 and Ψ8) were used to infect 3T3 cells for further analysis. 1 x 10^5 3T3 cells were infected as above with appropriately diluted viral supernatants at an estimated multiplicity of infection (moi) of 1. Neo^R^ colonies were selected for in G418-containing media for 10-14 days. At this point, individual colonies were picked for Southern and Northern analysis. The remaining colonies were washed with PBS, fixed in 0.5% glutaraldehyde/PBS for 10 minutes and assayed for β-galactosidase expression with 5-bromo-4-chloro-3-indolyl-b-D-galactopyranoside (X-Gal), which is cleaved by β–galactosidase to produce a blue precipitate. Cells were stained for 6 hours.
at 37°C; longer staining is not recommended because of the background staining which can result from endogenous lysosomal βgal activity. X-Gal solution was made up of 5 mM K$_3$Fe(CN)$_6$, 5 mM K$_4$Fe(CN)$_6$, 1 mM MgCl$_2$ and 1 mg/ml X-Gal in PBS. Subsequent to X-Gal staining, the plates were stained with crystal violet for 10 minutes to count the total number of colonies. The trapping frequency was calculated by dividing the total number of Neo$^R$ colonies by the number of X-Gal$^+$ colonies. In the case of U3βgeoSupF, the trapping frequency could not be accurately determined and was estimated to be similar to that of other LacZ trap vectors.

In order to pick LacZ positive and negative clones from cells infected with U3LacZpgkNeo, a live stain for β-galactosidase activity was used. Plates with well-separated colonies were washed with PBS and 150 μl of staining media (4% fetal calf serum, 10 mM HEPES pH 7.3 in PBS) was added. Fluorescein-di-galactosidase (FDG, Molecular Probes) was introduced via hypotonic shock by adding 150 μl of 20 μM FDG in dH$_2$O and incubating for 2 minutes at room temperature. The FDG solution was aspirated off and 2 ml of ice-cold staining media were added to quench the reaction. To prevent leaching of the fluorescent signal, the dishes were kept on ice in the dark until visualization on a Nikon inverted fluorescent microscope (Zarbl lab). Very little background staining was observed up to one hour after staining under these conditions (it took approximately 1 hour to screen 10 plates). βgal positive and negative colonies were marked and later picked and expanded for further analysis.

**Southern and Northern analysis of individual clones**

Southern analysis of genomic DNA was performed to assess the copy number and structure of integrated proviruses. 10 μg of genomic DNA was digested with appropriate restriction enzymes, fractionated on 0.8% agarose gels and transferred onto nitrocellulose membranes (Schleicher and Schuell). Cellular-proviral fusion transcripts were detected by Northern blot analysis. 10 μg of total RNA was fractionated on 1% formaldehyde-agarose gels and transferred onto nitrocellulose membranes. Probes were labeled with $^{32}$P dATP by the random prime labeling method (Feinberg and Vogelstein 1984). The Neo probe was prepared from a 1.4 kb TKneo fragment from pU3-gptTKneo (C. A. Scherer, unpublished vector). The LacZ probe (gift of Sita Reddy) was generated from a 3.0 kb EcoRI-HindIII fragment from LacZ cloned into Bluescript (KS,
Prehybridization took place at 42° in 50% formamide, 5x Denharts, 5x SSCPE and 500 μg/ml denatured salmon sperm DNA for 4 hours as described elsewhere (Maniatis et al. 1982). Hybridization took place overnight at 42° in 50% formamide, 10% dextran sulfate, 1x Denharts, 5x SSCPE and 100 μg/ml denatured salmon sperm DNA. Blots were washed in 2x SSC/ 0.1% SDS, 1x SSC/ 0.1% SDS, and 0.5x SSC/ 0.1% SDS at 65°C for approximately 30 minutes each and bands were visualized by autoradiography.

**RNase Protection Analysis**

RNase protection analysis was used to differentiate between internal viral transcripts and fusion transcripts originating in cellular flanking sequences. Radioactive riboprobes complementary to viral sequences were transcribed using either T3 or T7 RNA Polymerase. The probes were hybridized to total cellular RNA under conditions promoting RNA-RNA hybrid formation and single stranded RNA was digested with a combination of RNases. Protected fragments were separated on acrylamide gels and visualized by autoradiography.

**Materials:**

A probe complementary to the provirus coding strand was generated by using T3 RNA Polymerase (Pharmacia) to transcribe a 689 nt BamHI-Hpa1 fragment of BG2 cloned into Bluescript KS(−) (gift of Sita Reddy). This probe includes sequences from the viral envelope gene, U3 and LacZ and protects a 501 nt fragment from 5' fusion transcripts and a 643 nt fragment from 3' internal transcripts. Prior to use, the plasmid was linearized with BamHI, isolated on an agarose gel and purified using a Gene Clean Kit from Bio 101, Inc.

5x Transcription buffer was obtained from Pharmacia. This was composed of 200 mM Tris (pH 8.0), 40 mM MgCl₂, 20 mM spermidine and 200 μg/ml BSA.

RNasin RNase inhibitor and RQ1 RNase-free DNase were obtained from Promega.

5x Hybridization buffer was composed of 200 mM Pipes (pH 6.7), 2 M NaCl, and 5 mM EDTA. This was made in RNase free sterile dH₂O and aliquots were frozen at -20°C. Fresh buffer was prepared for each reaction by diluting the 5x stock in 100% formamide.
1x RNase Digestion buffer was made up of 10 mM Tris-CI (pH 7.5), 5 mM EDTA and 300 mM NaCl. This was stored at room temperature. To make up the final solution, 0.15 μl RNase A (10 mg/ml) and 1.7 μl RNase T1 (0.35 mg/ml, Boehringer Mannheim), were added per 300 μl of buffer.

Loading buffer stock was made of 1 mM EDTA, 0.1% Bromophenol Blue and 0.1% Xylene Cyanol. This was autoclaved and stored at room temperature. When ready to use, 100% formamide was added to a total volume of 80%.

**Protocol:**

**Note:** All solutions were RNase free until the RNA Digestion step.

**Labeling probe:** The following were mixed on ice:

- 4 μl 5x Riboprobe transcription buffer
- 2 μl 0.1M DTT
- 0.8 μl RNasin (25 units/μl)
- 3 μl 2.5 mM ATP, CTP, GTP mix
- 1 μl 0.125 mM UTP
- 1 μl probe DNA (0.5 μg)
- 10 μl α-32P UTP
- 1 μl T3 RNA Polymerase

The reaction was incubated at 37°C for 30 minutes. 1 μl of RQ1 DNase was added and the reaction was continued for another 15 minutes to remove the template DNA. Unincorporated nucleotides were removed by spinning through a G-50 RNA spin column (Quick Spin, Boehringer Mannheim). The purified probe was extracted twice with 1:1 phenol: chloroform, once with chloroform, and then Ethanol precipitated with 1/10 volume of 4M NaCl and 3 μg tRNA (as a carrier).

After storage for 10 minutes at -70°C, the precipitated probe was spun in an Eppendorf centrifuge for 30 minutes at 4°C. The pellet was air-dried for 10 minutes, and then resuspended in 100 μl of 1x Hybridization Buffer. 1 μl of probe was counted in a scintillation counter to assess radioactive nucleotide incorporation.

**Hybridization to RNA:** 1.5 μl of RNA (at 20 μg/μl) was added to 29 μl of 1x Hybridization buffer, and approximately 2 x 10^5 cpm of probe was added to each RNA sample. The samples were denatured for 5 minutes at 85°C and then immediately placed in a 55°C water bath and incubated overnight.
Digestion and processing of samples: The following morning, 300 µl of RNase digestion buffer was added to each sample, the tubes were vortexed and quickly transferred to a 37°C water bath for 1 hour. At this stage, only 3-4 samples were handled at a time to ensure quick processing. 2.5 µl of Proteinase K (20 µg/µl) and 10 µl of 20% SDS were then added (vortex after each addition) and the tubes were incubated at 37°C for another 10 minutes. The samples were extracted once with phenol:chloroform and ethanol precipitated with 20 µg of tRNA as a carrier. The pellets were washed two times with 70% EtOH to remove excess salt and air-dried for 10 minutes. They were then resuspended in 10 µl of loading buffer, denatured for 3 minutes at 85°C and immediately loaded onto a 6% denaturing acrylamide gel. Gels were run at constant wattage and bands were visualized by autoradiography.
Chapter 3:

The *in vitro* screen
Introduction

The search for novel genes involved in mammalian development can be quite costly and labor intensive using traditional methods. Therefore, in this study, I have developed an efficient \textit{in vitro} screen to facilitate the identification and cloning of novel genes involved in early mouse development. Embryonic stem (ES) cells closely resemble the totipotent cells of the early inner cell mass (ICM) from mouse blastocysts (Beddington and Robertson 1989). Upon \textit{in vitro} differentiation, ES cells can form many different cell types reminiscent of early embryo development (Evans and Kaufman 1981; Doetschman \textit{et al.} 1985). I have taken advantage of this feature in designing a screen for mutations that disrupt developmentally regulated genes. This screen utilizes a gene trap retrovirus, U3\(\beta\)geoSupF, as an insertional mutagen in ES cells. ES cells containing proviral insertions in active loci were differentiated in culture, and those clones exhibiting regulated expression of the \(\beta\)geo reporter gene were used to generate transgenic mice. This chapter describes the results of the \textit{in vitro} screen and the derivation of three mouse lines containing proviral insertions into developmentally regulated genes.

There are a number of possible approaches to studying mouse development. Classical embryological studies such as those used in \textit{Xenopus} have been very useful in elucidating cellular interactions and embryo morphogenesis (New 1991; McMahon 1993; Sive 1993) However, because of the intrauterine development of the mouse embryo, such studies are not very practical. Another method would be to map and identify known genetic mutations. Despite advances in mapping the mouse genome, this is slow and labor intensive; only a few genes have been cloned in recent years [for instance \textit{brachyury} and \textit{steel} (Copeland \textit{et al.} 1990; Herrmann \textit{et al.} 1990; Huang \textit{et al.} 1990; Zsebo \textit{et al.} 1990)]. Another approach involves targeted mutations of genes known to be involved in the development of other organisms, such as the highly conserved homeobox genes [reviewed in (Capecchi 1989a; Capecchi 1989b; Rossant 1991)]. However, this method is limited to known genes. A fourth alternative is to use insertional mutagenesis in ES cells to identify novel genes that are involved in development.

Insertional mutagenesis in early embryos or ES cells with exogenous DNA or retroviruses allows one to search for novel genes resulting in embryonic
phenotypes. However, the fraction of recessive phenotypes resulting from random DNA or provirus integration is low, only 10% and 5%, respectively (Gridley et al. 1987; Jaenisch 1988). To increase the efficiency of insertional mutagenesis, our lab and others (Gossler et al. 1989; von Melchner and Ruley 1989; von Melchner et al. 1990; Friedrich and Soriano 1991; Reddy et al. 1991; Reddy et al. 1992; Skarnes et al. 1992; von Melchner et al. 1992; Chen et al. 1994) have developed retrovirus gene trap vectors that select for integration into expressed genes. Our vectors contain promoterless coding sequences for selectable markers in the U3 region of the viral long terminal repeats. Proviral integration places the 5' U3 gene only 30 nucleotides from flanking cellular DNA, and selection for U3 gene expression gives rise to clones in which the proviruses have inserted into or near expressed exons of transcriptionally active genes (von Melchner et al. 1992; Chen et al. 1994).

As discussed in the Introduction (Chapter 1), I decided to use an in vitro screen to enrich for gene trap insertions into developmentally regulated genes. The embryoid body screen has several potential advantages over the use of chimeras. Specifically, the assay is fast, easy, and inexpensive, allowing a single investigator (such as myself) to analyze many clones simultaneously. In addition, a number of in vitro systems have been developed to induce ES cells to differentiate into specific cell types (discussed in Chapter 5). However, potential problems also existed. Firstly, it was possible that the embryoid bodies might not always differentiate properly (i.e. into all possible cell types). Secondly, at the time these studies were initiated, it was unclear if the gene expression seen in ES cells was a true indication of gene expression in the inner cell mass or just a tissue culture artifact. However, previous studies had indicated that, in a number of cases, gene expression of regulated genes in vitro accurately reflected in vivo expression. Mouse embryonic globin genes, for instance, are expressed in the correct temporal order in embryoid bodies, and further differentiation results in an appropriate switch to fetal/adult genes (Brown et al. 1987). Other examples of genes appropriately regulated in vitro include REX-1, murine H19, and gap junction genes (Fujiwara and Mizuuchi 1988; Hosler et al. 1989; Zack et al. 1992). Experiments by Sita Reddy had also shown that in three cases, in vitro expression of an activated U3-LacZ provirus was a good indicator of in vivo gene expression (Reddy et al. 1992). Taken together, this evidence indicated that an in vitro embryoid body screen utilizing the U3βgeoSupF gene trap would likely be an effective way to enrich for insertions in developmentally regulated genes.
Given the potential advantages offered by such a screen, then, the only way to address the possible problems was to actually test the system.

U3-βgeoSupF contains promoterless coding sequences for a β-galactosidase-neomycin fusion protein (βgeo) (Friedrich and Soriano 1991) such that integration of the provirus into an actively transcribed gene confers not only NeoR but also βgal activity. Selection in G418 is extremely sensitive, as I was able to identify integrations into weakly expressed genes which appear white in X-Gal staining assays. The existence of both white and blue clones allowed me to not only screen for genes that are restricted upon differentiation, but also those whose expression is increased. The in vitro screen is quite simple. ES cells were infected with U3βgeoSupF and selected in G418. NeoR clones were assayed for LacZ expression by X-Gal staining and were then allowed to differentiate in vitro. After 9 days, the embryoid bodies were stained with X-Gal and clones exhibiting differential expression of βgal were identified. Only clones which exhibited major changes in expression (i.e. blue to white or white to blue) were chosen for further study. This was to both prevent the possible effects of incomplete differentiation in the embryoid bodies and allow easy comparison of in vitro and in vivo expression (I assumed it would be easier to see major changes of expression than subtle ones in the embryos). Thus, clones which only showed minor changes in expression (i.e. white to white with a few blue cells and vice versa) were not selected. In future studies, however, it would be interesting to look at such clones in vivo, as it is quite possible that these clones contain insertions into genes that are only expressed in a subset of cell types.

Upon in vitro differentiation of 101 clones containing activated proviruses, approximately 20% exhibited changes in LacZ expression. Ten of these clones were injected into blastocysts to generate chimeric mice. Chimeras were assessed for germline transmission of the retrovirus; only those clones which were transmitted to the germline were studied further, eliminating the need to generate chimeras twice. When combined with data generated by Dr. Jin Chen (Vanderbilt), we were able to show that in all seven cases studied, expression patterns in vitro were an accurate indication of expression patterns in vivo. Thus, this screen, which combines gene trap technology and in vitro ES cell differentiation, is an efficient alternative to other methods available for isolating genes involved in early embryo development.

All of the molecular data (Southern, Northern, RNase Protections) described in this chapter were generated by me. Injections of ES cell clones into
blastocysts were performed by Dr. Jin Chen at Vanderbilt University. Initial analysis of germline contribution and maintenance of mouse colonies were done by Abdal Nachabeh and Jin Chen. Subsequent analysis was performed by myself, both at Vanderbilt and at MIT. Some of the pictures of X-Gal stained cells and embryos shown here are by courtesy of Jin Chen.
Results

Using a the newly constructed retroviral gene trap U3βgeoSupF, I utilized an
*in vitro* screen to identify insertions into developmentally regulated genes. ES-D3
cells were infected with the gene trap and NeoR clones containing activated
proviruses were selected. All clones were tested for β-galactosidase expression
before and after *in vitro* differentiation. Out of 101 NeoR clones tested, 20
exhibited regulated expression of the reporter gene. Ten of these clones were
selected for injection into mouse blastocysts to generate chimeric animals. Three
clones were transmitted to the germline, two showing blue to white *in vitro*
regulation, and one showing white to blue regulation. *In vitro* expression of LacZ
was consistent with the pattern observed *in vivo* in all three lines, as well as four
mouse lines generated by Jin Chen (Vanderbilt University).

*Selection and analysis of regulated clones*

The βgeoΨ2 producer line was used to infect low passage (4°-7°) ES-D3
cells and NeoR clones were selected for in G418 containing medium. Although
the average NeoR titer on NIH 3T3 cells was 2 x 10^4 NeoR cfu per ml per 10^7
cells, titers on ES cells were at least 100 fold less, approximately 10^2 NeoR cfu
per ml per 10^7 cells. Southern analysis of NeoR clones indicated the presence of
either 1 (66%), 2 (26%) or 3 (8%) proviruses (Fig. 3.2B and data not shown).
This is consistent with an approximate multiplicity of infection (m.o.i.) of 1,
indicating that ES cells are as susceptible to infection as NIH 3T3 cells; however,
expression of integrated proviruses is much lower than in NIH 3T3 cells. This
drop has been seen with other gene trap vectors as well (Soriano *et al.* 1991;
Reddy *et al.* 1992; von Melchner *et al.* 1992) and could be due to the stem cell-
specific transcriptional silencer located in the provirus as will be discussed later.
Because of the low titer on ES cells, several rounds of infection were performed
in order to generate enough clones for analysis. Individual clones were
transferred to 24 well dishes and expanded. Although there was little chance of
picking sister clones from the same plates (as there was no trypsinization after
infection), no more than five clones were picked per infected plate. A total of 101
NeoR clones were tested in this study. Initial staining data and the results of the
*in vitro* screen are summarized in Table 3.1.
Table 3.1
Results of Screen

<table>
<thead>
<tr>
<th>Total clones screened: 101</th>
</tr>
</thead>
<tbody>
<tr>
<td>#Blue: 30</td>
</tr>
<tr>
<td>#Mixed (blue and white): 30</td>
</tr>
<tr>
<td>#White: 41</td>
</tr>
</tbody>
</table>

Clones showing regulated expression: 20  20% of total

Blue ---> white: 15
White ---> blue: 5

Integration of the U3-βgeoSupF provirus places the βgeo reading frame just 30 nt away from flanking cellular DNA. In order to be activated, the provirus must integrate into expressed genes (von Melchner and Ruley 1989; von Melchner et al. 1990; Reddy et al. 1991; von Melchner et al. 1992; Chen et al. 1994). Because βgeo is a fusion protein, all NeoR clones should also be expressing β-galactosidase. However, as had been observed in 3T3 cells, when β-galactosidase activity was assessed in undifferentiated cells by X-gal staining, only 30% of the clones were uniformly stained blue. The remainder of the clones exhibited either no staining at all (white, 40%) or mixed blue/white staining (30%). "Patchy" staining has been observed before with LacZ transgenes (MacGregor et al. 1987; Friedrich and Soriano 1991). This expression is a characteristic of certain insertions which is maintained even after subcloning. Many clones which exhibited patchy staining were not scored in the in vitro differentiation screen, as it was difficult to determine how much expression had changed.

After testing undifferentiated cells for β-gal activity, all the clones were induced to differentiate in vitro by removing differentiation inhibiting factors (LIF and embryonic fibroblast feeder cells). Lightly trypsinized ES cells were grown at high density (approximately 1:10 dilutions of 80% confluent plates) in suspension culture in bacterial petri dishes in feeder cell medium. After 5 days, the early embryoid bodies were replated on gelatinized tissue culture plates and allowed to differentiate an additional 4 days. The resultant embryoid bodies were tested for β-gal activity by X-Gal staining. To ensure that changes in expression weren't due to random changes in some cells, my criteria for differential expression were
fairly stringent; only clones which showed significant differences in expression were chosen. Therefore, a number of clones which remained mostly white but had small patches of blue cells were not counted, and vice versa. Figure 3.1 indicates some typical staining patterns observed in undifferentiated and differentiated clones. Panels A and B show clones which do not exhibit regulated expression, i.e. blue to blue (A) and white to white (B). Panels C and D show examples of a blue to white clone (C) and a white/light blue to blue clone (D). Twenty clones exhibited changes in \( \beta \text{geo} \) expression in the first round of screening; 15 clones decreased expression (blue to white), and 5 clones increased expression (white to blue). All twenty clones were retested to ensure the accuracy of the screen, and ten clones were chosen for further study and injection into mouse blastocysts. Table 3.2 summarizes the data on these clones, which will be discussed in detail.

Data from Southern analysis of the 10 clones is shown in Figure 3.2. Unlike earlier studies using NIH 3T3 cells (Chapter 2), the majority of ES cell clones contained deleted proviruses. The expected structure of a properly integrated provirus is depicted in (A). Digestion with \( \text{HindIII} \) (which does not cut within the provirus) and hybridization to a viral-specific probe will indicate the total number

---

**Table 3.2**

<table>
<thead>
<tr>
<th>Clone (# Proviruses):</th>
<th>Regulation:</th>
<th>Structure:</th>
<th>Germline:</th>
</tr>
</thead>
<tbody>
<tr>
<td>B2-3 (1)</td>
<td>Blue to white</td>
<td>Deleted</td>
<td>Yes</td>
</tr>
<tr>
<td>D1-2 (1)</td>
<td>Blue to white</td>
<td>Deleted</td>
<td>No</td>
</tr>
<tr>
<td>D4-2 (3)</td>
<td>White to blue</td>
<td>Normal</td>
<td>No</td>
</tr>
<tr>
<td>E1-1 (2)</td>
<td>Blue to white</td>
<td>1 deleted</td>
<td>No</td>
</tr>
<tr>
<td>F3-5 (1)</td>
<td>White to blue</td>
<td>Deleted</td>
<td>No</td>
</tr>
<tr>
<td>L2-2 (2)</td>
<td>Blue to white</td>
<td>1 deleted</td>
<td>No</td>
</tr>
<tr>
<td>1.5 (1)</td>
<td>Blue to white</td>
<td>Deleted</td>
<td>No</td>
</tr>
<tr>
<td>2.4 (1)</td>
<td>Blue to white</td>
<td>Deleted</td>
<td>Yes</td>
</tr>
<tr>
<td>5.5.2 (1)</td>
<td>White to blue</td>
<td>Deleted</td>
<td>No</td>
</tr>
<tr>
<td>7.4.2 (2)</td>
<td>White to blue</td>
<td>Normal</td>
<td>Yes</td>
</tr>
</tbody>
</table>
Figure 3.1 Examples of X-Gal staining patterns seen in undifferentiated and differentiated ES cells infected with U3βgeoSupF. Undifferentiated ES cells are shown in the left panels and embryoid bodies from the same clones are shown on the right. The clones show the following regulation: (A) constitutive expression of the transgene (blue to blue), (B) little to no staining in either ES cells or embryoid bodies (white to white), (C) repression of proviral transcription (blue to white), and (D) up-regulation of proviral transcription (white to blue). Photos in Panels A and C are courtesy of Jin Chen.
Figure 3.1
of proviruses present in the clone. The genomic fragments containing the provirus should be at least as large as the total size of the provirus, 10.5 kb. In order to analyze the structure of the provirus, genomic DNA was also digested with *BamHl*, which cuts three times within the provirus. Probes specific for LacZ or Neo sequences should hybridize to not only an internal fragment (3.4 kb and 2.7 kb, respectively) but also to one flanking fragment. (B) shows the results of a Southern in which genomic DNA was digested with *HindIII* and hybridized to a Neo-specific probe. Most clones contained 1 or 2 integrated proviruses, with the exception of clone D4-2, which contained 3 proviruses (lane 3). A number of proviruses are located in DNA fragments that are less than 10.5 kb, the minimum size for a properly integrated provirus. This indicated that the proviruses contained deletions. To analyze where the deletions had occurred, Southern were performed on genomic DNA digested with *BamHl*. When hybridized to LacZ or Neo specific probes, only 4 clones (lanes 3, 4, 6 and 10) contained the expected internal bands (Panel C and data not shown). A number of clones showed hybridization to a single band on the Southern. In addition, only DNA from clones containing the internal fragments hybridized to a supF probe (data not shown). Together, this data suggests that the aberrant proviruses consist of a single LTR lacking the internal sequences which encompass *gag*, *pol*, *supF* and *env*. These deletions do not affect the ability of U3βgeoSupF to act as a gene trap, as the single LTR still positions the βgeo AUG only 30 nucleotides away from cellular DNA. However, they might affect the mutagenic potential of the gene trap (to be discussed in Chapter 4).

Figure 3.2 Southern analysis of 10 clones chosen in the *in vitro* screen. Approximately 10 μg of genomic DNA was digested with appropriate restriction enzymes, fractionated on 0.8% agarose gels and blotted onto nitrocellulose membranes. Blots were hybridized to probes specific for either Neo or LacZ. Expected fragment sizes are indicated in Panel (A). Southern blot of DNA digested with *HindIII* and hybridized to a Neo probe is shown in Panel (B). Southern blot of DNA digested with *BamHl* and hybridized to a LacZ probe is shown in Panel (C). Lanes contain the following clones: 1, B2-3; 2, D1-2; 3, D4-2; 4, E1-1; 5, F3-5; 6, L2-2; 7, 1.5; 8, 2.4; 9, 5.5.2; 10, 7.4.2.
Figure 3.2

A. 

B. 

C. 

10.5 kb

3.4 kb
β-galactosidase activity accurately reflects the level of transcription.

When the Neo^R^ ES clones were tested for βgal expression, it was found that 40% of the clones did not stain. There are at least two possible explanations for this. The first is that the βgeo AUG may not have been in a favorable context for optimal translation of the cellular-proviral fusion transcript. Another possibility is that these white clones contained proviral insertions into weakly expressed genes, where transcription was sufficient to confer Neo^R^, but not enough to stain with X-gal. Northern blots and RNase protection assays corroborating this hypothesis are shown in Figure 3.3. Expected proviral transcripts are depicted in Panel A. As the internal viral promoter is inactive in ES cells, the only transcripts present should be those initiating in 5' flanking sequences. Northern analysis on total RNA (Panel B) showed the presence of a single fusion transcript in most blue cell lines (lanes 6, 7, 8, 9, 11 and 12). However, transcripts were not detected in white (lanes 1, 3, 4, 5 and 13) or light blue (lane 10) cell lines, with the exception of clone F3-5 (lane 2). The transcript around 1.4 kb seen in all lanes is most likely due to the presence of RNA from Neo^R^ MEFs (Mouse Embryonic Fibroblast feeder cells) in the RNA preparation. It is present in all samples, including uninfected ES cells (data not shown).

The same cell lines which did not show bands on the Northern had very weak fusion transcripts on RNase Protections. Analysis on 30μg of total RNA indicated that blue clones expressed more fusion transcript than white cell lines (Panel C). RNA from all clones protected a 501-nt fragment that is consistent with the size expected for hybrid cell-virus transcripts initiating in the cellular DNA and extending through the 5' LTR. This fragment is not seen with control tRNA and uninfected ES cell RNA (data not shown). The levels of the 501 nt protected fragment were consistently higher using RNA from blue clones (lanes 4-10) than RNA from white clones (lanes 1-3 and 11), indicating that less of the fusion transcript is present in white clones. For comparison, all clones exhibited similar levels of a 289 nt protected fragment which corresponds to transcripts of the large ribosomal subunit L32. Lane 11 contains RNA from clone 7.4.2, which exhibited an increase in β-gal expression upon differentiation. In undifferentiated cells, most of the cells stain white; however, a few cells stain very pale blue in a patchy manner. Thus, although no transcript is visible on Northern blots, the RNase Protection shows that there is more transcript than other white clones. One aberrant white clone (F3-5, lane 2), which also showed a transcript on the
Figure 3.3  Analysis of cellular-proviral fusion transcripts in clones infected with U3bgeoSupf. Total RNA was isolated from ES cells and analyzed by Northern blot or RNase Protection Assays. Panel (A) indicates potential transcripts extending through the provirus and the fragments which would be protected by the LacZ riboprobe in RNase Protection Assays (5' RT and 3' RT). (B) Northern blot of 10 μg RNA from ES cell clones hybridized to a Neo probe. Cellular-proviral fusion transcripts are only evident in blue clones (lanes 6, 8, 9, 11, and 12). Fusion transcripts from white (lanes 1, 2, 3, 4, 5, and 13) and light blue (lanes 7 and 10) are not visible by Northern blot. Lanes contain the following samples: 1, C3-1 (white control); 2, D4-2; 3, F3-5; 4, 5.5.2; 5, J2B4; 6, B2-3; 7, D1-2; 8, E1-1; 9, F3-2 (blue control); 10, L2-2; 11, 1.5; 12, 2.4; 13, 7.4.2. (C) RNase Protection Analysis of fusion transcripts. 30 μg of total RNA was simultaneously hybridized to a 689 nt antisense riboprobe complementary to sequences in env, U3 and LacZ and a 280 nt riboprobe complementary to the ribosomal L32 transcript (internal control). Transcripts protected from the 3' LTR will be 643 nt long, whereas transcripts protected from the 5' LTR will be 501 nt long. Lanes contain the following samples: 1, D4-2; 2, F3-5; 3, 5.5.2; 4, B2-3; 5, D1-2; 6, E1-1; 7, F3-2; 8, L2-2; 9, 1.5; 10, 2.4; 11, 7.4.2; 12, yeast tRNA.
Figure 3.3

A.

B.

1 2 3 4 5 6 7 8 9 10 11 12 13

28S

18S

C.

1 2 3 4 5 6 7 8 9 10 11 12

Probe

5' RT

3' RT

L32

πgeo

U3

RU5

SupF

U3

RU5

>4.3 kb
Northern blot, exhibits a strong band around 460 nt. This clone has undergone some internal rearrangement or deletion which was not determined. RNA from clones containing full length proviruses also protected small amounts of a 643 nt fragment corresponding to transcripts extending through the 3' LTR (lanes 1, 6, 8 and 11). This is probably the result of a transcript which did not terminate in the 5' LTR, since the internal proviral promoter is inactive in ES cells. Taken together, the RNA data suggest that the intensity of LacZ staining (as measured in X-gal assays) is a general indication of transcription levels. In addition, it confirms that there is no bias against insertion of U3βgeoSupF into weakly expressed genes.

To test if the changes in the expression of β-geo in differentiated cells reflected changes in the amount of the cellular proviral fusion transcripts, RNase protection analysis was performed on total RNA isolated from undifferentiated cells and embryoid bodies. Figure 3.4 shows data from seven clones (four from Jin Chen) which were transmitted to the germline. Lanes labeled "U" contain RNA from undifferentiated cells and lanes labeled "D" contain RNA from embryoid bodies. There is a decrease in the levels of the 501 nt fusion transcript expressed in all of the clones which restrict β-geo expression upon differentiation (B2-3, 2.4, J1D4, J3A3 and J5C1), whereas the L32 transcript levels remained similar in each set of clones. Clones 7.4.2 and J2B4, which increased β-geo expression upon differentiation, show increased levels of the fusion transcript. An increase in the amount of 3' (689 nt) protected fragment is also seen in clone 7.4.2, the only clone shown which contains an intact provirus. This increase is most likely due to the activation of the viral promoter upon differentiation. In summary, all 7 regulated clones exhibit changes in the levels of cellular-proviral fusion transcripts which matched the observed X-Gal staining pattern.

Figure 3.4 RNase Protection Analysis of RNA from undifferentiated and differentiated ES cells. Total RNA was obtained from undifferentiated ("U") ES cells and embryoid bodies cultured for 9 days ("D") and 30 μg was used for RNase Protection Analysis. Protected fragments are the same as those diagrammed in Figure 3.3. Clones 7.4.2 and J2B4 were classified as white to blue in vitro; the rest of the clones were classified as blue to white. Clones J2B4, J1D4, J3A3 and J5C1 were isolated by Jin Chen.
Expression of βgal in transgenic mice is accurately predicted by in vitro gene expression.

The 10 clones in my study were injected into C57/BL6 blastocysts in pools of three by Jin Chen at Vanderbilt University. The blastocysts were implanted into pseudo-pregnant CD1 recipients and chimeras were identified by the presence of the agouti coat color (for which ES-D3 cells are homozygous). Male chimeras were bred to C57/BL6 females to assess germline contribution (i.e. transmission of agouti coat color). Tail DNA from the progeny of germline chimeras was isolated to determine which of the injected clones had contributed to the germline in each mouse line. It was hoped that all three injected clones would be able to contribute to the germline, as had been seen previously (Friedrich and Soriano 1991). DNA from each of the three appropriate ES cell lines and tail DNA samples were digested with BamHI and fractionated on agarose gels. BamHI was chosen because of the clearly differentiated restriction digest patterns that the clones showed in previous Southern blots. Figure 3.5 shows sample Southern blots from the three mouse lines. Only one clone from each pool was found to contribute to the germline. It is uncertain why this occurred. One possibility is that, since the three lines were cultured together for at least a few days, one of the cell lines grew faster than the other two and dominated the culture. Growing the cell lines separately and mixing equal quantities of each cell line in the injection mixture may have alleviated this problem. However, for this thesis, three mouse lines sufficed, and no effort was made to generate additional ones. Two of the founder mice, Peabody and Mike,

---

**Figure 3.5** Southern Blots to determine which ES cell clones had contributed to the germline of chimeric founder mice. Approximately 10 μg of genomic DNA isolated from tail biopsies and control DNA isolated from the three ES cell clones injected in each pool were digested with BamHI and processed for Southern analysis as previously described. All blots were hybridized to a Neo probe. Germline clones are indicated by an asterisk (*). Numbered lanes represent individual progeny mice; appropriate clones are shown in the three right lines in each panel. (A) "Mike" progeny (ES cell clone 2.4). (B) "Peabody" progeny (ES clone B2-3). (C) "O3" progeny (ES clone 7.4.2). 7.4.2 contains two provirus (virus 1 and virus 2) which segregated separately. The 2.7 kb internal Neo fragment is also indicated.
were generated from blue to white clones B2-3 and 2.4, respectively (Panels A and B). The third founder mouse O3, was generated from the white to blue clone 7.4.2 (Panel C). The 7.4.2 cell line contained two proviral insertions, indicated in Figure 3.5 as "Virus 1" and "Virus 2".

Heterozygous F1 males were outbred to C57/BL6 females and expression of the βgal transgene was assayed in blastocysts and post-implantation embryos by X-Gal staining. Jin Chen and I observed three categories of gene expression in vivo: (i) blue to white, (ii) white (or very weak staining) to blue, and (iii) blue to restricted blue. The first category included clones B2-3 and 2.4, which exhibited blue or light blue staining in blastocysts and no staining in embryos aged 6.5 through 12.5 days (Figs. 3.6 and 3.7, and data not shown). It is possible that expression might resume later in development in these clones, but this was not investigated. The second category, showing increased expression, included clone 7.4.2. Consistent with in vitro staining and RNase Protection results, 7.4.2 did not show staining at the blastocyst stage but showed staining as early as 6.5 days p.c. (Fig. 3.8, and data not shown). This early staining was absent from extraembryonic portions of the embryo (Panel B). Day 8.5 embryos exhibited widespread staining which appeared to be stronger in the head process (especially in the forebrain), brachial arches, and neural tube (8C and D). Clone 7.4.2 contained two proviral insertions which were identified as Virus 1 and 2 on the basis of BamHI Southerns (see Figure 3.5 C). Only one provirus, #2, was associated with βgal activity (Virus 1 staining data not shown). The two proviruses were separated by breeding, and the mice carrying the inactive provirus were not further analyzed. The third category consisted of genes which are expressed in blastocysts but show restricted staining patterns later in development. This category included two mouse lines generated by Jin Chen, and will not be discussed here.

Collectively, these data show that an in vitro screen using differentiation of ES cells infected with a gene trap virus is an effective way to identify developmentally regulated genes. Approximately 20% of the screened clones showed regulated LacZ expression upon differentiation in vitro. These clones can be transmitted to the mouse germline at a high frequency and analyzed in vivo. Together, Jin Chen and I investigated seven regulated genes in transgenic mice. Not only was in vivo expression accurately predicted by in vitro LacZ expression, but three clones, including clone 7.4.2, expressed LacZ in interesting patterns in vivo.
Figure 3.6: X-Gal staining of heterozygous Mike progeny. Embryos were dissected from the uterus, fixed in 2% paraformaldehyde, 0.2% glutaraldehyde (in PBS) and stained overnight in X-Gal solution. Photos of the following stages are included: (A) blastocysts (day 3.5), (B) egg cylinder (day 6.5), and (C) day 8.5 p.c.

Figure 3.7: X-Gal staining of heterozygous Peabody progeny. Embryos were prepared as previously described. The following stages are shown: (A) blastocysts, (B) day 7.5, and (C) day 9.5.

Figure 3.8: X-Gal staining of heterozygous O3 progeny. Embryos were prepared as previously described. The following stages are shown: (A) blastocysts, (B) day 7.5, (C) day 8.5, and (D) day 8.5 heterozygous and wild type litter mates.
DISCUSSION

I have developed a new gene trap retrovirus to isolate insertional mutations in developmentally regulated genes via an in vitro screen. Embryonic stem cell clones containing U3βgeoSupF proviruses integrated into expressed loci were selected in G418 and then differentiated in culture. Approximately 20% of the trapped genes exhibited regulated expression of the proviral reporter gene upon differentiation. Together with Jin Chen, seven clones were transmitted to the germline and heterozygous progeny were analyzed for βgal expression. All seven clones exhibited in vivo staining patterns that were consistent with the regulatory changes seen in vitro. As discussed in Chapter 1, this screen provides an efficient alternative to cDNA subtraction analysis and other screens which utilize chimeric animals. Large numbers of clones can be screened with relative ease, and only those insertions which appear interesting need be transmitted into the mouse germline. Once in the germline, temporal and spatial LacZ staining patterns, as well as the phenotypes of mice homozygous for the proviral insertions, can be investigated.

U3βgeoSupF accurately reports endogenous gene expression

The expression of LacZ from retroviral gene traps is a reliable indication of wild type promoter activity. When LacZ transgenes are placed under the control of cloned promoter sequences, they are often either sporadically expressed, or not expressed at all [reviewed in (Jaenisch 1988) and (Palmiter and Brinster 1986)]. Tissue specific expression of transgenes via cloned promoters can be quite difficult; it often requires all of the regulatory sequences, which may be many kilobases upstream or downstream (Flenniken and Williams 1990; Püschel et al. 1990; Behringer et al. 1993). In addition, adjacent cellular sequences sometimes influence the expression of inserted transgenes (Bonnerot et al. 1990; Cohen-Tannoudji et al. 1992), which can result in either ectopic expression or no expression at all. However, when LacZ-containing vectors are used in homologous recombination experiments to knock out specific genes, LacZ expression is a reliable indicator of normal expression for that gene (Lé Mouellic et al. 1990; Mansour et al. 1990; Lé Mouellic et al. 1992). Because proviruses in gene trap events are controlled by cellular promoter sequences in their normal
chromosomal location, their expression should reflect the actual regulation of endogenous genes, instead of external influences (Skarnes et al. 1992). Therefore, expression of LacZ in U3βgeoSupF trapping events would be a reliable indicator of the genes' normal expression patterns. This is corroborated by X-Gal and in situ staining patterns in three of the mouse lines for which we have identified insertion loci: REX-1, ECK, and O3 (Rogers et al. 1991; Chen and Ruley 1994; Scherer and Ruley 1994).

The screen we employed is an in vitro embryoid body assay. Although a number of labs have reported accurate expression of regulated genes in cultured ES cells and EBs (Lindenbaum and Grosveld 1990; Simon 1993), it was conceivable that the differential expression of the gene trap reporter gene was due to tissue culture artifacts. It is now clear that this is not the case. βgeo expression in undifferentiated ES cells and differentiated embryoid bodies predicted in vivo reporter gene expression in all seven clones reported here, as well as in three clones (two constitutive and one repressed) analyzed by Sita Reddy in a previous study (Reddy et al. 1992). RNase Protection Analysis of RNA isolated from ES cells and embryoid bodies indicated that the differential regulation was due to changes in the level of cellular-proviral transcripts in differentiated cells. Because the βgeo transcription unit is basically the same in each clone (except for the 5' end), this is most likely due to changes in transcription levels rather than changes in RNA processing or stability. If cellular differentiation modified the processing or stability of that transcript, then all clones should exhibit the same changes in βgeo expression.

Negative selection against the viral primer binding site may cause proviral deletions in ES cells

Approximately 80% of the ES cell clones contained single LTR structures, with all internal sequences (gag, pol, env and supF) deleted. No deletions were observed in over 20 NIH 3T3 clones that were analyzed (Scherer and Ruley 1994). Although the mechanism of deletion in unclear at this time, an investigation of this phenomenon must consider several issues. Firstly, the deletions appear to be occurring at a higher frequency in ES cells, as they are rarely observed in NIH 3T3 cells. This implies that the deletions are not a result of some event intrinsic to the replication of the vector; if this were the case, all infected cells types would contain deleted proviruses at the same frequency.
However, we cannot rule out that the deletions are occurring in every cell type but are being preferentially selected for in ES cells (as discussed later).

Secondly, when clones contain multiple proviruses, expression of the reporter gene has been linked to the deleted provirus (J. Chen, S. Reddy and H. E. Ruley, unpublished results). Thus, deleted proviruses appear to be preferentially expressed. Thirdly, the timing of the deletion is an issue. In principle, if the deletions were the result of homologous recombination once the provirus had integrated, then one would expect to find mixed clones containing cells with either deleted or wild-type proviruses. This might be visualized by the "patchy" staining seen in some clones, where not all cells stain to the same extent. However, all of the clones analyzed in this study appeared to be homogenous in Southern blots. In addition, patchy staining has been observed with other LacZ transgenes as well, and is characteristic of certain insertions (MacGregor et al. 1987; Friedrich and Soriano 1991). Thus, if the deletions are occurring after integration, they must either occur before cell division or there must be a great selective advantage for clones carrying deleted proviruses such that they outgrow cells with wild-type proviral insertions.

One possible mechanism involves the retroviral tRNA primer binding site (PBS). This site normally initiates reverse transcription of the retrovirus by binding a proline tRNA primer to allow synthesis of the "strong-stop" DNA, which is a copy of the 5' end of the retroviral RNA molecule containing the PBS, R and U5. In EC cells, this site has been shown to act as a transcriptional silencer of viral genes controlled by either the LTR or internal promoters (Weiher et al. 1987; Feuer et al. 1989). This repression occurs at the DNA level and is position-independent. A cellular factor ("binding factor A"), which is present only in undifferentiated (ES and EC) cells and capable of binding to the PBS site, is thought to mediate the repression (Loh et al. 1990; Petersen et al. 1991; Kempler et al. 1993). Binding Factor A is hypothesized to be a helicase, and may cause transcriptional repression by changing the structure of the proviral DNA (E. Barklis, personal communication to H. E. R.) It is possible that this effect could also extend to cellular sequences adjoining the proviral insertion.

Transcriptional silencing mediated by the PBS could explain the selective advantage of clones containing deleted proviruses. If the silencer is capable of repressing transcription of the genes into which an intact retrovirus has integrated, then these genes might not express enough of the reporter gene to confer Neo^R, and therefore would not survive selection. It has also been shown
that retroviruses containing a wild-type PBS tend to become methylated; this might be an indirect mechanism due to lack of expression. When expression of wild type proviruses is compared to that of PBS-mutant proviruses located in the identical chromosomal position, it is found to be approximately 100 fold less (Berwin and Barklis 1993). Thus, a combination of silencing and methylation might prevent a number of wild-type proviral insertions from surviving G418 selection. In addition to potentially lowering the overall titer of virus infection, this might result in an apparent skewing of results, as viruses which have deleted the PBS would dominate the population of selected clones. In support of this is the fact that in the cases where more than one provirus integrated, either all proviruses were normal (D4-2 and 7.4.2; 3 and 2 proviruses, respectively) or only one of the two proviruses were deleted (E1-1 and L2-2). Since most proviruses do not integrate into genes, one would expect the deleted provirus to be associated with the expressed fusion gene; this has been shown to be true in two clones (J. Chen and S. Reddy). Interestingly, both D4-2 and 7.4.2 expressed cellular-proviral transcripts at very low levels in ES cells. It is possible that this could be an effect of PBS-mediated silencing, and that these genes are normally expressed at somewhat higher levels in embryonic cells.

The main argument against the silencing theory is that the silencing effect appears to decrease with distance (Kempler et al. 1993). U3βgeoSupF contains approximately 4 kb of sequence between the genomic sequences and the RBS. In addition, the cellular promoter sequences can be an additional 100 nucleotides to several kb from the viral integration site. In theory, then, the cellular promoter should not be affected by a transcriptional silencer at least 4 kb away. However, if the putative helicase can affect the proviral DNA structure, it may also affect upstream promoter regions.

A retrovirus backbone which utilizes a mutant primer binding site is presently being constructed to test if the deletion event is linked to the presence of the PBS (G. Hicks and H. E. Ruley). If the PBS is associated with the transcriptional repression of genes adjacent to wild-type proviruses, then the titers of the mutant proviruses on ES cells should be higher than those observed with wild-type gene trap vectors. In addition, the majority of clones isolated after selection for expression of the proviral reporter gene should contain full-length proviruses instead of deleted ones.

Although selection against the PBS explains why deletions might be selected for, it does not explain how the deletions occur. Since the βgeo
sequences inserted into the LTR are unusually long (approximately 4kb), it is possible that the defective proviruses are the product of post-integration homologous recombination events between the LTRs. This is supported by the fact that as the length of the LTR insert increases in our constructs, the percentage of deleted proviruses observed also increases [Sita Reddy, personal communication, and (von Melchner et al. 1992)]. However, unless the selective advantage for clones containing deleted proviruses is quite strong, the probability of this occurring at such a high frequency is low. In "hit and run" gene targeting techniques, homologous recombination events are selected for to generate subtle mutations in cellular genes (Bautista and Shulman 1991; Hasty et al. 1991b). The mutations are introduced into the chromosomal locus using insertion vectors which create duplications of the targeted genomic DNA. Excision of the vector results in the mutation being retained in the chromosomal target. The insertion vector contains sequences homologous to the target gene, but with a small mutation. In addition, selectable markers (such as thymidine kinase, Neo, or guanine phosphotransferase, gpt) are present in the vector sequences outside the region of homology. In the "hit" step, the vector is linearized in the homologous sequences and integration into the gene of interest is selected for. If properly integrated, the insertion vector will have created a duplication of genomic sequences upstream and downstream of the selectable markers. The "run" step involves selection for an intrachromosomal recombination event which results in the loss of vector sequences accompanied by either the reversion of the targeted gene to wild type or formation of the mutant gene, depending on how the recombination occurs. The frequency of these events varies between 1/1000 to 1/100,000 per cell generation (Bautista and Shulman 1991; Hasty et al. 1991b; Ramírez-Solis et al. 1993; Serwe and Sablitzky 1993). Properly integrated proviruses consist of two identical LTRs separated by approximately 2kb of non-homologous sequence, much like the situation used for hit and run mutations. If the single LTR phenomenon was due to intrachromosomal recombination, one would expect the frequency to be similar to the frequency of "run" events described above. Post-integration recombination events, although they might occur, can not explain the extremely high frequency of deletions observed in our system.

If the deletion event is occurring before integration, it should be possible to isolate single LTR structures from infected cells. ES and 3T3 cells can be infected at a high moi (approximately 10) and Hirt extractions of low molecular
weight DNA can be isolated at different time points before integration (which usually occurs within 24 hours). Southern analysis of this DNA can then be used to analyze the structure of the viral DNA intermediates. If Binding Factor A has a direct effect upon the reverse transcription process that results in the production of integration-competent single LTRs, then one should see bands indicative of that structure in ES but not 3T3 cells. On the other hand, the deletions might be due to a homologous recombination event during reverse transcription (perhaps during plus strand synthesis, when one complete double-stranded LTR has been formed) and thus could occur in both cell types at an equal frequency. In this case, one should see single LTR structures in both populations of infected cells at an equal frequency.

Summary

The in vitro screen that I have described here is a powerful tool for the isolation of differentially regulated genes. Not only is the vector an efficient insertional mutagen which accurately represents cellular transcription, but the in vitro screen itself is also efficient and accurate. With the development of more sophisticated cell biological techniques, one could imagine a number of different positive-negative selection schemes involving this vector. We only investigated spontaneous differentiation of ES cells, however, it is possible to induce ES cell differentiation into specific developmental pathways (discussed in Chapter 5). Variations of the screen used here could thus be used to identify many new genes involved in mouse development.
Materials and Methods

Culture of ES Cells

Materials:

**Media:** Mouse Embryo Fibroblast (MEF) feeder cells were grown in DME supplemented with 10% fetal calf serum, 10 units of penicillin per ml, and 10 μg of streptomycin per ml. ES cells were grown in DME supplemented with 15% fetal bovine serum (Rehautin F.S., Armour Pharmaceutical, heat inactivated at 55°C for 30 min.), 100mM nonessential amino acids (Gibco), 0.1mM 2-mercaptoethanol, 10 units of penicillin per ml, 10 μg of streptomycin per ml, and 1000 units of leukemia inhibitory factor (LIF, ESGRO, Gibco) per ml.

**Gelatin:** 0.1% gelatin in dH$_2$O was autoclaved and stored sterily.

**Trypsin:** 1 ml of 2.5% trypsin (Gibco, mycoplasma free) was added to 40 ml of "Versine", 0.02% EDTA in PBS plus Phenol red, and used within one week.

**Mouse Embryo Fibroblasts:** (Protocol from the Hynes lab) A fourteen day pregnant mouse was sacrificed by cervical dislocation and the embryos were dissected from the uterine horns. Embryos were dissected one at a time; fetal membranes and placenta, head, and soft tissue (liver, heart and other pigmented areas) were removed and the remaining embryo was rinsed in fresh PBS. The embryo was placed in a fresh dish, minced with two scalpels in 2 ml of trypsin and incubated at 37°C for 5 minutes. 8 ml of feeder cell medium was added and the solution was then allowed to settle in a conical tube for a few minutes. The supernatant was plated into a 10cm tissue culture dish and incubated for 24 hours, at which point the medium was changed. The cells were cultured until confluent (about 1-2 days) and then split 1 to 10 into new dishes and cultured until confluent (about 2 days). The cells were then trypsinized and frozen in 1 ml of ice-cold freezing medium (feeder cell medium, 20% FCS and 10% DMSO). To expand MEFs for irradiation, 4-5 vials were thawed out and cultured until confluent. Each dish was split 1:3 two times. All dishes were then trypsinized, resuspended in approximately 50 ml of medium, and an aliquot was counted. The cells were irradiated in the gamma cell irradiator for 30 minutes (approximately 3000 rads) and then frozen at a concentration of 2.5 x 10$^6$ cells/vial. One vial was used per 10 cm dish for growth of ES cells. All batches of
feeder cells were tested for mycoplasma contamination using the "Rapid Detection System" kit from Gen-Probe®.

Methods:
D3-ES cells were the generous gift of Janet Rossant and Rudolf Jaenisch. ES cells were cultured on gelatinized plates with irradiated mouse embryo fibroblast layers (MEFs). To infect ES cells, 1x10^5 cells were seeded in 6 cm plates and incubated overnight. Cells were incubated at 37°C with 1 ml of appropriately diluted and filtered viral supernatant from producer line βgeoΨ² (titer on NIH 3T3 cells, 2 x 10⁴ neo^R cfu/ml per 10⁷ producer cells) in 8 μg/ml Polybrene for 1 hour with occasional rocking. Neo^R Clones were selected in ES medium containing 0.3 μg/ml G418 for 10-14 days, at which point individual colonies were picked. In order to pick colonies, a small dissecting microscope was placed in the hood and culture dishes were marked to identify the positions of colonies. Individual colonies were pulled up in a drawn-out Pasteur pipette (controlled by a mouth piece) and placed into a gelatinized 24-well tissue culture plate containing 100 μl trypsin. Colonies were trypsinized for 5 minutes and gently disrupted with a Pipetteman. 1 ml of fresh selection medium was added and the cells were cultured undisturbed. Fresh media was added after 24 hours, and as needed thereafter (every 1-2 days). Colonies were expanded to 10 cm dishes, whereupon aliquots were either frozen, stained for βgal activity, or expanded for embryoid body formation.

In vitro differentiation of ES cells.
Embryoid bodies were generated as described in detail in (Robertson 1987). Nearly (80%) confluent 10cm plates of ES cells were trypsinized for 2-3 minutes and diluted without disaggregation into petri dishes for suspension culture in feeder cell medium. After 5 days, the mixture of simple and cystic embryoid bodies was plated onto gelatinized tissue culture plates and incubated for a further 4 days in the same medium. Embryoid bodies were fixed and stained with X-Gal as described in Chapter 2.

RNase protection analysis
In order to analyze cellular-proviral fusion transcripts, cellular RNA was isolated from undifferentiated ES cells and differentiated embryoid bodies. The
LacZ probe is exactly as described in Chapter 2. In addition, a 280 nt NotI-EcoRV fragment of the L32 large ribosomal subunit (generous gift of Michael Shen and Phil Leder, (Shen and Leder 1992)) was transcribed at 10% the specific activity of the LacZ probe to use as an internal control. 30μg of cellular RNA was hybridized to 32P-labeled probes overnight at 55°C. After hybridization to both probes simultaneously, samples were digested with 5μg/ml RNaseA and 2μg/ml RNaseT1 and processed for electrophoresis as previously described (Chapter 2).

Construction of germline chimeras

Both individual NeoR ES cell clones and pools of three NeoR clones were injected by Jin Chen into preimplantation C57/BL6 blastocysts as described elsewhere (Robertson 1987). Chimeric mice were identified by the presence of agouti coat color and males were outbred to C57/BL6 females to assess germline transmission. The presence of specific transgenes in F1 progeny was assayed by Southern analysis of DNA isolated from tails (Laird et al. 1991). A 1 cm piece of tail was obtained from anaesthetized mice and was digested overnight at 55° in 0.5 ml lysis buffer containing 100 mM Tris-HCl pH 8.5, 5 mM EDTA, 0.2% SDS, 200mM NaCl and 100μg Proteinase K. Continuous agitation during incubation was essential for obtaining complete lysis. Following lysis, the tubes were vortexed and then spun in an Eppendorf centrifuge for 10 minutes to pellet undigested tail remnants. The supernatant was poured into fresh tubes containing 0.5 ml isopropanol, and the tubes were inverted approximately 10 times to precipitate the DNA. The DNA "blob" was taken out of the tube with a yellow tip, transferred to a fresh yellow tip (to get rid of excess liquid), and placed in a clean tube. Uncapped tubes were placed at 37° for about 30 minutes to dry, and the pellets were then resuspended in 100μl TE either overnight at room temperature or for 30 minutes to 1 hour at 55°. 10μl aliquots were digested for Southern analysis.

Embryonic Expression of LacZ Fusion Genes

Males heterozygous for proviral insertions were mated to C57/BL6 females. The presence of mating plugs the following morning was determined, and 12 Noon of that day was defined as day 0.5 of development. Embryos from
various stages of development were isolated from pregnant females, fixed in PBS/ 2% paraformaldehyde/ 0.2% glutaraldehyde for 10-30 minutes at 4°C and rinsed in PBS for at least 1 hour on ice before staining. 3.5 to 8.5 day embryos were fixed for 10 minutes, and embryos aged 9.5 days and older were fixed for 30 minutes. Staining was performed overnight in PBS/ 0.02% Nonidet P-40/ 0.01% SDS/ 2mM MgCl₂/ 5mM K₃Fe(CN)₆/ 5mM K₄Fe(CN)₆/ and 1 mg /ml X-Gal, pH 7.2. The staining solution for day 9.5 and older embryos was prepared in 1mM Tris-buffered saline (TBS: 50mM Tris, 0.8% NaCl, 0.02% KCl, pH 8.0), instead of PBS, to reduce background staining due to endogenous lysosomal β-gal activity. After staining, the embryos were rinsed in PBS and visualized by dark field microscopy. Photographs were taken with either Kodak Tungsten-64 slide film or Kodak 400 speed print film.
Chapter 4:

Analysis of the Mike, Peabody and O3 Mouse Lines
Introduction

The study of early mammalian development has been hindered by the lack of efficient methods for the identification of genes responsible for recessive embryonic phenotypes. Insertional mutagenesis with retroviral vectors has resulted in a number of novel mutations (Schnieke et al. 1983; Soriano et al. 1987; Kratochwil et al. 1989; Spence et al. 1989); the advantage here is that genes can be cloned with relative ease by using the retrovirus as a tag to isolate adjacent genomic DNA sequences. However, random insertional mutagenesis is also inefficient, as only 5% of insertions result in obvious phenotypes (Gridley et al. 1987; Jaenisch 1988). Therefore, a number of labs have developed gene trap retroviruses that enrich for insertional mutations in ES cells (Gossler et al. 1989; Friedrich and Soriano 1991; Reddy et al. 1992; Skarnes et al. 1992; von Melchner et al. 1992). These vectors vary in design but generally contain a promoterless reporter gene which confers a selectable phenotype (i.e. drug resistance or LacZ expression) when integrated into expressed genes. Clones which have been selected in vitro can be injected into blastocysts to generate transgenic mice, where the mutagenic effects of proviral integration can be determined. As discussed in the previous chapter, I have added an additional level of selection to the gene trap scheme. In order to identify genes involved in early developmental processes such as gastrulation and cell lineage specification, an in vitro embryoid body assay was used to identify gene trap insertions that disrupt genes regulated during early development.

Characterization of gene trap insertions involves several steps. First, sequences flanking the provirus must be cloned to identify the integration site. Genomic flanking sequences can be cloned by inverse PCR (von Melchner and Ruley 1989); alternatively, cellular transcripts appended to the proviral transcript can be cloned via 5' RACE (Frohman et al. 1988). In either event, the flanking sequences can be used to probe cDNA libraries in the hopes of identifying the endogenous gene. In addition, flanking sequences can be used to genotype F2 progeny of heterozygous parents to assess whether the gene trap insertion results in a recessive phenotype in homozygous animals. Very few gene trap insertions involving embryonic phenotypes have been characterized to date. Disruption of the transcription factor TEF1 with ROSAβ-geo resulted in heart defects and embryonic lethality (Chen et al. 1994). Disruption of another gene,
*fug1*, a mouse homologue of the yeast RNA1 gene, resulted in growth arrest of homozygous embryos on or about day 6 of embryonic development (DeGregori et al. 1994). Recently, additional evidence indicates that this gene acts as a GTPase Activating Protein (GAP) to the G-Protein Ran (personal communication to H.E.R.). In this chapter, I describe the characterization of three additional gene trap lines which were generated as a result of the *in vitro* screen described in the previous chapter.

Three ES clones containing U3βgeoSupF insertions into developmentally regulated genes were transmitted to the germline of chimeric mice. ES cell clones B2-3 and 2.4 were classified as blue to white on the basis of their *in vitro* differentiation characteristics. Analysis of sequences upstream of the proviral integration sites in ES cells did not reveal any homologies to known genes. Analysis of F2 progeny indicated that the proviral insertions did not cause any obvious phenotypes in homozygotes; therefore, these mouse lines were not studied in detail. Embryos homozygous for the 7.4.2 insertion, on the other hand, died soon after implantation. Embryos heterozygous for this clone, which was classified as white to blue *in vitro*, showed no observable X-Gal staining in blastocysts but significant staining in post-implantation embryos. Cellular-proviral fusion transcripts cloned by 5' RACE were used to probe a cDNA library and several cDNAs were sequenced. The gene has significant homologies to an open reading frame (ORF) of unknown function present on chromosome II of *Saccharomyces cerevisiae* (Smits et al. 1994).

Most of the data in this chapter were generated by me. Abudi Nachabeh (Vanderbilt) helped with the sequencing of the final two cDNAs and also took very good care of the mice at Vanderbilt. Heterozygous mating pairs from each line were maintained at MIT. Marjorie Kummiskey helped me obtain tail biopsies from all of my progeny (she anaesthetized and ear-clipped, I chopped and soldered). Dr. Andy McMahon was very helpful in analyzing LacZ expression patterns in O3 embryos. Thanks also go to Kate Willett of the Steiner lab, who let me use their microscope on numerous occasions, and Patricia Reilly of the EM lab, who taught me how to section.
Results

Three germline chimeras were generated from the screen described in the previous chapter. Here, I describe the cloning of 5' flanking sequences from all three insertion sites and an analysis of F2 progeny from heterozygous mating pairs. Two of the mouse lines (Mike and Peabody) did not exhibit any obvious phenotypes in mice homozygous for the proviral insertions. No viable homozygous progeny were recovered from the third line, O3, which contained an insertion into a gene with homology to a yeast open reading frame of unknown function. I will first describe results from the two blue to white mouse lines, Peabody and Mike, which were not studied in detail. The rest of the results will then concentrate on the white to blue line, O3.

Peabody

Two germline chimeras, Peabody and Peter, were generated from the B2-3 cell line. Progeny from these two founder mice were characterized together, and will be referred to as Peabody for the sake of simplicity. The B2-3 cell line exhibited blue to white regulation in vitro; ES cells stained dark blue, and EBs were white in X-Gal assays. As shown in Chapter 3, the only embryo stage examined that showed staining was the blastocyst stage, which is consistent with the ES-specific staining seen. RNase protection assays showed that this regulation occurred on the RNA level, as the amount of cellular-proviral fusion transcript decreased to almost undetectable levels in embryoid bodies.

Several different attempts were made to clone 5' cellular sequences flanking the provirus in B2-3 cells. Inverse PCR using HindIII to cut the genomic DNA was performed twice. The first time, a 300 nucleotide product was cloned after secondary PCR using the U3H and MseI primers (described in Methods). When this fragment was used as a probe on southern blots of tail DNA digested with StuI, a single band was observed in all lanes at about 2.4 kb (data not shown). This indicated that the fragment cloned was not an actual flanking sequence, as a true flanking sequence should have hybridized to both an endogenous fragment and a shifted band equal in size to the endogenous fragment plus the viral insert. This fragment did turn out to be useful after all, though, as it was used as an internal control band in the southern blots used to
genotype Mike F2 progeny. The second time inverse PCR was performed, a product approximately 200 base pairs in length was cloned. Sequencing of this fragment showed that it was completely different from the first PCR product. Southern analysis on genomic DNA isolated from the B2-3 cell line and heterozygous Peabody and Mike progeny showed two bands at approximately 5.5 kb and 9.5 kb (data not shown). The lower band was present in all lanes, whereas the upper band was only present in lanes containing DNA with the B2-3 proviral insertion. When this blot was stripped and rehybridized to a Neo probe, the upper band also hybridized to that probe (data not shown). This data is consistent with the shifted band containing the 4 kb single LTR insert and indicates that the second iPCR product cloned does indeed represent cellular-proviral 5' genomic flanking sequences. Numerous attempts were also made to clone cellular fusion transcripts via 5' RACE. However, only two PCR products were cloned, and neither was the proper fragment. In the first case, the sequences upstream of the LacZ2 primer used in the PCR reaction were not U3 sequences, indicating that the PCR product was an artifact. In the second case, a PCR product was amplified regardless of whether reverse transcriptase was included in the 5' RACE reaction, indicating that this was again an artifact.

The B2-3 genomic flanking sequence is shown in Figure 4.1. This fragment is 191 nt long from the genomic Hinfl site to the provirus. Most striking about this sequence is a potential 3' splice site immediately preceding the provirus. The genomic sequence is: CCCTTCCCCTCAGG, which is quite similar to the consensus 3' splice site sequence: (T or C)\textsubscript{10} NCAGG (Senapathy \textit{et al.} 1990). In addition, there are several possible branch sites upstream of the potential acceptor site (CTGAG, three times, and ATGAC, one time). No potential 5' donor sequences are present. Without RNA sequences and because of the uncertainty associated with identifying real splice sites (ibid), it is difficult to know for certain whether or not this is an active 3' splice site; if it is, then the provirus has integrated into the 5' end of an exon. GCG sequence analysis of this sequence on the "Blast" server did not identify any homology to known sequences (last checked on 12/15/94).

In order to analyze the effect of this insertion on homozygous progeny, heterozygous mating pairs were set up. Pups were weaned at 20-21 days, and tail biopsies were taken 5-7 days after weaning. Genomic DNA was isolated from tails as described in (Laird \textit{et al.} 1991), digested with \textit{Stul}, and prepared for Southern analysis. A typical phosphorimager scan of a southern blot probed
Figure 4.1

B2-3 Flanking Sequence

5' ATTCCTGTCTTCCCCACACCAACTCTTTCCACTGA 40
ACCAACCATCTGACGTGAGGCTACTGAGGAATTGCAA 80
TGACTTTCCAGACCTTGGGACTCCACCCGACAGGATTGTT 120
TGCCGTTGGAGTTGTCGCTCACCCCCGATGATTGGACGAG 160
GGGCGGGCATTTAAGGCAGTCCTCCCTTCAAG 3' 191

Figure 4.1  Sequence of the B2-3 genomic 5' flanking sequence cloned by inverse PCR. The total length of the PCR product is 191 nt. The underlined region indicates a potential splice acceptor site.
Figure 4.2

A.

B.

Total number of mice: 45
Wild type: 6 13%
Heterozygote: 32 71%
Homozygote: 7 16%

Figure 4.2  Southern analysis of Peabody F2 progeny. Approximately 10μg of tail DNA was digested with Stu1, fractionated on 0.8% agarose gels and transferred to nitrocellulose membranes. Blots were hybridized to a probe specific for the B2-3 flanking sequence. (A) Phosphorimager scan of a representative Southern blot. Lanes 2, 8, 9 and 11 contain DNA from wild type mice (wt). Lanes 7 and 10 contain DNA from mice homozygous for the retroviral insertion (m). The total number of genotyped mice and the percentage of each genotype is shown in (B).
with the iPCR flanking fragment is shown in Figure 4.2 (A). Lanes 2, 8, 9 and 11, which show only the lower endogenous band (wt), contain DNA from wild type mice. Lanes 7 and 10, which show only the shifted mutant band (m), contain DNA from mice homozygous for the proviral insertion. The remainder of the lanes, which show both bands, contain DNA from heterozygous mice. A total of 45 F2 progeny were genotyped. The percentage of each genotype is shown in (B): 71% of the progeny were heterozygous, and only 16% and 13% were homozygous and wild type, respectively. This apparent skewing towards heterozygous mice is probably just random, as the sample size is quite small. The most important conclusion from this experiment is that homozygous animals are viable at birth and do not exhibit any obvious defects. Two homozygous breeding pairs were set up to examine their fertility. So far, each breeding pair has had two litters; thus, the B2-3 insertion does not seem to have any effect on either embryo viability or the reproductive capacities of adult mice. However, homozygous embryos will be analyzed in further detail to assess if there are subtle phenotypes caused by the proviral insertion.

Mike

The chimeric founder mouse "Mike" was generated from the cell line 2.4, which exhibited blue to white regulation \textit{in vitro} and \textit{in vivo}. Undifferentiated ES cells stained blue in X-Gal assays but very little staining was present in EBs. \textit{In vivo}, blastocysts stained light blue and no staining was detected in any other stages (up to day 12.5) investigated. RNase Protection analysis confirmed that this regulation occurred at the RNA level, as there was a large decrease in the amount of cellular-proviral fusion transcript detected in differentiated cells.

Genomic flanking sequences and RNA fusion transcripts were cloned from the 2.4 cell line via iPCR and 5' RACE. These sequences are summarized in Figure 4.3. The iPCR fragment, which begins at the \textit{HinfI} site shown in bold type, is contained within the 5' RACE sequence. A number of ATG codons are present in the sequence; however, each one has an in-frame stop codon located downstream. Therefore, none of these ATGs could be the actual start codon. No obvious 5' or 3' splice sites are apparent in the sequence either. Since the inverse PCR product was cloned first, it is possible that the 5' RACE product was actually amplified from contaminating genomic sequences. Unfortunately, as the proper controls were not performed in this case, this question can only be
### Figure 4.3

#### 2.4 Flanking Sequence

<table>
<thead>
<tr>
<th>5'</th>
<th>Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CCGGGGCTTGGACAGCAATGCACAGTCAGTGTCTGTTCA</td>
<td>40</td>
</tr>
<tr>
<td>ATGGCATCAGTCAGATGATTCAGGCTGATGTGCGCC</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>ACAAAGGCCATGGGGCTTGCTTTGAGAGCATGGGGGCTGAT</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>ATGTTAACATGCCCCAGTGGG</td>
<td>142</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 4.3** Sequence of the 5' flanking sequence cloned from the 2.4 proviral insertion. The 5'RACE product encompasses the entire sequence, whereas the inverse PCR product begins at the Hinf1 site (GATTC).
answered by attempting to reclone cellular flanking sequences via 5' RACE. Sequence searches using the Blast server revealed no homologies to other known genes (last checked on 12/15/94).

Tail biopsies from F2 progeny of heterozygous mating were prepared for Southern analysis as described above. When either of the cloned flanking sequences was used as a probe on Southern blots, a large number of bands were observed on the autorad, indicating that the fragments contain repetitive sequences. It was impossible to genotype F2 progeny using the flanking sequences. Therefore, southern blots of tail DNA digested with StuI were simultaneously hybridized to a Neo probe and an internal control probe (the first B2-3 iPCR product described above). The internal probe was used to control for differences in sample loading and hybridization. The ratio of the intensity of the band of interest (i.e. Neo) and the internal band (Int) were calculated for each sample using a phosphorimager. The Neo: Int ratio in homozygotes should be approximately twice that observed in heterozygotes, if the amount of hybridization is linear. A sample Southern is shown in Figure 4.4. Wild type progeny were obvious by the lack of the Neo-specific upper band (lane 4). To determine if the other lanes contained DNA from heterozygous or homozygous mice, the Neo: Int ratio was calculated using the Image Quant program on a Molecular Dynamics phosphorimager. Grids were drawn around the neo band, the internal band, and an area exhibiting background hybridization in each lane. The phosphorimager was then used to calculate the volume of each box, and the resultant values were used as a measure of band intensity. The calculated Neo: Int ratios for all animals genotyped are shown in Table 4.1. A heterozygous control mouse in the first experiment had a Neo: Int ratio of 2.1. The other ratios varied between 0.3 (wild type) and 8.7. Sample numbers 4 and 9 have Neo: Int ratios that are 4 and 2.5 fold greater than the heterozygous control ratio, respectively. These two mice (both males) are most likely homozygotes. Several other mice were borderline cases; samples 2, 5 and 6 have ratios that are 1.5-, 1.6-, and 1.8-fold greater than the control ratio. If the ratio in homozygotes has to be at least 2-fold greater than the heterozygote ratio, then these mice are probably not homozygotes. However, if there is some variation (which could have only been calculated by having a number of heterozygous control mice in each experiment to calculate the variance), then it is possible that these could be homozygotes. In the second experiment, presumptive heterozygous mice have Neo: Int. ratios ranging from 1.6 to 2.3. Samples having
Figure 4.4 Southern analysis of DNA obtained from F2 progeny of heterozygous "Mike" mice. Approximately 10 μg of tail DNA was digested with Stu1, fractionated on 0.8% agarose gels and transferred to nitrocellulose membranes. The membranes were simultaneously hybridized to a Neo probe and an internal control probe (see text). Lane 4 contains DNA from a wild type mouse. To determine the genotypes of the other progeny, the ratio of the Neo: Int signal was calculated. Results are shown in Table 4.1.
Table 4.1
Probable Genotypes of "Mike" F2 Offspring

<table>
<thead>
<tr>
<th>Sample</th>
<th>Neo: Internal Ratio</th>
<th>Probable Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>wild type control</td>
<td>0.5</td>
<td>wild type</td>
</tr>
<tr>
<td>heterozygous control</td>
<td>2.1</td>
<td>heterozygous</td>
</tr>
<tr>
<td>1</td>
<td>2.2</td>
<td>heterozygous</td>
</tr>
<tr>
<td>2</td>
<td>3.3</td>
<td>heterozygous?</td>
</tr>
<tr>
<td>3</td>
<td>2.3</td>
<td>heterozygous</td>
</tr>
<tr>
<td>4</td>
<td>8.7</td>
<td>homozygous</td>
</tr>
<tr>
<td>5</td>
<td>3.4</td>
<td>heterozygous?</td>
</tr>
<tr>
<td>6</td>
<td>3.9</td>
<td>heterozygous?</td>
</tr>
<tr>
<td>7</td>
<td>2.1</td>
<td>heterozygous</td>
</tr>
<tr>
<td>8</td>
<td>2.2</td>
<td>heterozygous</td>
</tr>
<tr>
<td>9</td>
<td>5.3</td>
<td>homozygous</td>
</tr>
<tr>
<td>10</td>
<td>0.9</td>
<td>wild type</td>
</tr>
<tr>
<td>11</td>
<td>2.6</td>
<td>heterozygous</td>
</tr>
<tr>
<td>12</td>
<td>2.6</td>
<td>heterozygous</td>
</tr>
<tr>
<td>13</td>
<td>1.9</td>
<td>heterozygous</td>
</tr>
<tr>
<td>14</td>
<td>2.1</td>
<td>heterozygous</td>
</tr>
<tr>
<td>15</td>
<td>2.4</td>
<td>heterozygous</td>
</tr>
<tr>
<td>16</td>
<td>0.3</td>
<td>wild type</td>
</tr>
<tr>
<td>17</td>
<td>2.1</td>
<td>heterozygous</td>
</tr>
</tbody>
</table>

Experiment 2:

<table>
<thead>
<tr>
<th>Sample</th>
<th>Neo: Internal Ratio</th>
<th>Probable Genotype</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.8</td>
<td>heterozygous</td>
</tr>
<tr>
<td>2</td>
<td>1.7</td>
<td>heterozygous</td>
</tr>
<tr>
<td>3</td>
<td>1.7</td>
<td>heterozygous</td>
</tr>
<tr>
<td>4</td>
<td>0.3</td>
<td>wild type</td>
</tr>
<tr>
<td>5</td>
<td>3.4</td>
<td>homozygous?</td>
</tr>
<tr>
<td>6</td>
<td>3.5</td>
<td>homozygous</td>
</tr>
<tr>
<td>7</td>
<td>2.3</td>
<td>heterozygous</td>
</tr>
<tr>
<td>8</td>
<td>3.2</td>
<td>heterozygous</td>
</tr>
<tr>
<td>9</td>
<td>3.4</td>
<td>homozygous?</td>
</tr>
<tr>
<td>10</td>
<td>3.6</td>
<td>homozygous</td>
</tr>
<tr>
<td>11</td>
<td>1.6</td>
<td>heterozygous</td>
</tr>
</tbody>
</table>

Total # of F2 offspring: 28
a signal ratio greater than 3.4 were classified as probable homozygotes (i.e. samples 5, 6, 9 and 10). This is not a very accurate method for determining genotype; depending on how one categorizes the borderline cases, the ratio of wild type to heterozygous to homozygous progeny varies. If the borderline cases are classified as heterozygotes, then the ratio equals 4: 19: 5, or 14%: 68%: 18%. If, on the other hand, the borderline cases are classified as homozygotes, the ratio changes to 4: 16: 8, or 14%: 58%: 28%. Most likely, the actual ratio lies somewhere in between. Although the latter ratio reflects a Mendelian ratio of inheritance, the sample size is again very small, and the method of genotyping is not precise. Therefore, the only conclusion that can be drawn is that some homozygous progeny are viable at birth.

In order to determine if the putative homozygous animals are fertile, four different mating pairs were set up. Two mating pairs consisted of homozygous males and either a heterozygous female or a potential homozygous female. Both of these mating pairs have had litters, indicating that the male homozygotes are fertile. Two other mating pairs with homozygous males and probable homozygous females (samples 5 and 6 in the second experiment in Table 4.1) were also set up. Both of the females have had litters; therefore, there is probably no obvious phenotype associated with this insertion either. However, homozygous progeny will be further analyzed for subtle phenotypes.

O3

The O3 germline founder was generated from the 7.4.2 cell line. In X-Gal assays, undifferentiated 7.4.2 ES cell colonies stained mostly white with a few pale blue cells. Embryoid bodies generally showed an increase in LacZ expression, although some variability was observed (cell staining shown in previous chapter, Figure 3.1 D). This increase in expression was also seen in vivo. There was no detectable X-Gal staining in blastocyst stage embryos, whereas post-implantation embryos showed wide-spread staining with some regions staining more strongly. The X-Gal staining patterns were briefly summarized in Chapter 3, and will be examined in more detail here (Figure 4.5).
Figure 4.5  X-Gal staining of heterozygous O3 embryos. (A). 8.5 day p.c. embryo. Staining is diffuse all over, but particularly strong in the head process, brachial arches (b), and notochord / neural tube region (n). Light staining is also visible around the primitive heart (h). (B). 9.5 day p.c. embryo. Again, diffuse staining is visible throughout the embryo. However, stronger staining is seen in the head, first and second brachial arches (1 and 2), neuropore (p), and forelimb bud (f). (C). 12.5 day p.c. embryo. Diffuse staining has decreased throughout the embryo. Darker staining is still visible in the notochord, brain (especially telencephalon (t)), head, and proximal forelimb and hindlimb buds (f and h).
LacZ expression in embryos

In the following section, two reference texts were used to identify structures in the developing embryo: "The House Mouse" and "The Mouse" (Theiler 1989; Rugh 1990). Day 6.5 embryos exhibited weak staining throughout the egg cylinder. By day 7.5, however, staining is evident in the embryonic portion of the embryo (no staining is visible in extra-embryonic regions). Staining appears to be strongest along the midline of the neural plate and the head fold. Diffuse staining is evident throughout the rest of the neural plate and the primitive streak (see picture in Figure 3.8, Panel B).

Figure 4.5 shows whole-mount embryos of various stages stained with X-Gal. Diffuse staining is evident throughout day 8.5-12.5 embryos. However, it appears to be stronger in some tissues, particularly neural epithelia. Panel A in Figure 4.5 shows an 8.5 day embryo, focusing on the posterior neural tube. Staining is seen in the head region, brachial arches (b), neural tube and notochord (n). Light staining is also seen around the developing heart (h). It is possible that the apparently localized staining is due to a greater density of cells in those areas. Therefore, sagittal sections of an 8.5 day embryo were cut to investigate this possibility. However, the method of sectioning used (plastic imbedded embryos, cut with a glass knife on a microtome) produced very thin sections (1-2 microns) in which it was difficult to see the blue precipitate. A few of the thicker sections (2-4 microns) did verify the staining seen in the forebrain, neural tube and primitive heart in whole mounts (data not shown). These areas also stained darkly with methylene blue, indicating that they do contain a greater density of cells. Thus, I can not rule out that the darker staining seen in some regions is not simply due to the presence of more cells.

Day 9.5 staining (Panel B) is similar to that seen in day 8.5. Darker staining is again seen in the head region, along the neural tube, and the heart. In addition, staining is seen in the first and second brachial arches (1 and 2), the neuropore (p) and in the developing forelimb bud (f). The only major change seen at day 10.5 is a more pronounced staining in the forebrain (data not shown). General (diffuse) staining in day 12.5 embryos is decreased compared to earlier embryos but darker staining is still evident in neural tissue (Panel C). Most striking is the dark staining visible in the telencephalon (t). In addition, both the forelimb (f) and hindlimb (h) show staining in proximal but not distal regions of the buds.
Cloning and analysis of the O3 gene

Cellular transcripts appended to the 7.4.2 fusion transcripts were cloned from total cell RNA by 5' RACE. Attempts were also made to clone genomic flanking sequences by inverse PCR, but these were unsuccessful. The sequence of the 5' RACE product is shown in Figure 4.6. A potential AUG translational start codon is boxed. In addition, a probable splice donor site is underlined and in bold type. When this sequence was analyzed using the "Blast" server, no homologies to known genes were detected. The presence of the 5' splice site indicates that the provirus might have integrated into an intron. To see if the cloned PCR product contained any 5' genomic flanking sequences, a Southern blot of O3 tail DNA digested with StuI (prepared by Abudi Nachabeh) was hybridized to a 5' RACE probe. The probe hybridized to two endogenous bands (approximately 1.2 kb and 2.8 kb) in all lanes (data not shown). In addition, a band at approximately 11 kb was present in lanes containing DNA from only those animals which had inherited Virus 2 (data not shown). As animals with the Virus 1 insertion did not stain with X-Gal (tested by Jin Chen, data not shown), this indicated that the 5' RACE product was generated from the activated provirus (#2). The 5' RACE product was also used to probe a Northern blot of total RNA from ES cells. A transcript of about 1.2 kb was present in all lanes, but no fusion transcript was apparent in the 7.4.2 lane (data not shown).

An 8.5 day embryonic cDNA library (gift of Brigid Hogan) was probed with the 5' RACE product. 1 x 10^6 plaques were screened on duplicate filters in the first round, from which five potential positive plaques were identified. Secondary screening of these plaques indicated that two out of the five were positive. Phage DNA from these two plaques was isolated and the cDNA inserts were subcloned into Bluescript (Stratagene) for sequencing. The first clone, 5A2, was 600 nucleotides long, and the second clone, 14C1, was 1.1 kb long. Both clones contained sequences identical to the 5' RACE product. 14C1 matched the 5' RACE product from nucleotide 6 through the putative splice donor site at nucleotide 112. 5A2 began homology at nucleotide 2 but only matched the 5' RACE product through nucleotide 56. The sequences of the two cDNAs then converged at the splice junction and were identical until nucleotide 550 of 14C1. The 55 nucleotides missing from clone 5A2 (nucleotides 56-112) could comprise an alternatively spliced exon (indicated by asterisks in Figure 4.6). When
Figure 4.6

7.4.2 Flanking Sequence

5' AGGCCCAGGGACTCCGGGTGAAGATGCCCACGAGGCCC 40
GAACGCTCATGAG*AAATTTTGAGCCACCTTGGCTAAT 80
GGGATGAGCTCCAGCCGCCCTCTTGAAGAAGT*AAATAATCC 120
TTTTGTGAGACCCCCACCCAGCCTTCCACATAAGCTT 160
ACTTGCTATACATCTGACTCCCCAAAGAGG 3' 190

**Figure 4.6** Sequence of the 5' RACE product cloned from 7.4.2 RNA. Asterisks flank an alternatively spliced exon not present in the 5A2 cDNA sequence. Sequences downstream of the second asterisk are intron sequences, and the probable splice donor site is underlined and in bold type. A potential ATG initiator codon is boxed.
analyzed by the "Blast" server, the cDNA sequences were identified to be
homologous to a yeast open reading frame, YBR0320, of unknown function
(Smits et al. 1994). Because this gene appears to be expressed at higher levels
in neural tissue, I have named this gene Neural Regionalized, or Nrd.

To verify the Nrd cDNA sequence, additional cDNAs were cloned and
sequenced. Eighteen positive plaques were identified when the 14C1 cDNA was
used to reprobe the original cDNA library. Phage DNA was isolated from four
plaques and the cDNA inserts, which ranged from 0.9 kb to 1.2 kb, were
subcloned into Bluescript. The two longest clones (2 and 19), which were
approximately 1.2 and 1.1 kb long, respectively, were sequenced (with help from
Abudi Nachabeh). The composite cDNA sequence and putative open reading
frame are shown in Figure 4.7. Interestingly, cDNA 19 was missing 91
nucleotides (compared to 14C1) starting at position 659. This missing sequence
seems to contain coding sequences which are not present in the yeast
sequences. When the sequence is present, the open reading frame is
prematurely terminated (in the exon). However, if the exon is eliminated, an
open reading frame of 310 amino acids (starting at nucleotide 46 and ending at
nucleotide 976) is present. Therefore, this sequence is not included as a part of
the open reading frame; the sequence and stop codon (*) are shown between
nucleotides 659 and 660 flanked by two † symbols. Approximately 300
nucleotides of additional 3' sequence was obtained from the two new cDNAs
using primers to Bluescript. These sequences did not include the 3' sequences
obtained from the 14C1 cDNA; therefore about 50-100 nucleotides of 3'
untranslated sequence are missing from this cDNA. This can be easily
sequenced using internal primers.

Figure 4.8 shows an alignment of the putative Nrd open reading frame and
the yeast YBR0320 open reading frame (ODP1 gene product). These two gene
products are 64% conserved on the amino acid level. Identities are indicated by
vertical lines, whereas highly (and less highly) conserved residues are indicated
by colons and periods. The Nrd protein sequence was analyzed in a number of
programs in ggc, including the motif search and the "peptidestructure" program.
No obvious protein motifs are present in the sequence. The protein also seems
to be neither strongly hydrophilic nor hydrophobic. As none of the hydrophobic
stretches are longer than about 18 amino acids, it is unlikely that this is a
membrane protein (hydrophobic sequences of transmembrane proteins tend to
Figure 4.7 Sequence of the Nrd cDNA. An alternatively spliced exon that would cause premature termination is inserted between nucleotides 659-660, flanked by ‡ symbols. The sequence of a 310 amino acid open reading frame is indicated underneath the nucleotide sequence. In addition, 3' untranslated sequences (note gap) are shown at the end. Approximately 50-100 nucleotides are missing from the 3' untranslated region.
Figure 4.7
Nrd cDNA sequence

GCGGACTGGGGTGAAGATGGGAGGCGCCGGAACCTGATCTGGNGAGAAATTGTGA 60
MENFV
GCGACCTGGCTAAAGGAGTGACGCTCCAGCCGCTCTTGGAGAGTTCTCTGTGGCQA 120
ATLANGMSLQQPLEEVSFCGQ
GCGAAGACTCTGAGAACCCAAAGGTCAGATCAGACTCAAAGAGCAGACTCTTCTGAC 180
AESSKEPNAMEDMTSKDYFYD
TCTATGCCCACTTTGGCAGCTACAGGATGCTGAAGGATGAGGCTCCACCTCACA 240
SYAHFGHEEMLKEDEVRTLT
TACOCGAATCGCTATTTGTCTCAACATCGCCATCTCTCTAAAGACAAAGGTGTTGTGGATGTG 300
YRNSMFHNRHFLFKDKKVVLVDV
GCCTGAGGCCACTCCTCCTGATTTGCTCACAAGGCGGGGCGCGCGAAATTTATT 360
GSPTGILCMFAAASKAGARKVI
GGGATGTGAGTGTCCATATCTCGATTTGCTGAGATGGTCAAAAGGCACAAGATTA 420
GIECSSISDYAVKIVKANKL
GACCATGTGTGAGCATCAGGAAAGGTGAAGTTGAGGCTCCACCTGGCAAG
DHVVTIIKVEEEVELPVEK
GTGACCATCGCTCAAGGAGGATGATTTTCTGGCTCATGGCATGTCACAC 540
VDIIISEWMLGFCFLYESMLN
ACCGTGCTGAGCTGGCAAGAAGGGCAGCGTGCCAGGCTGTTCTCCAGACGCG
TQLHLARAHGKAPDLIFPDPR
GCGACCTGGTACAGGACAGCCATTGGAGCCAAATAAAGACTGACAGATCCACTG 600
ATLYVTAIEDRQYKDYKIH

#TGAGCTGGGGTTATGTGGTGAGGGGCAAGATGGGAGGCGCCGAGGCCCTGG
CELG*L
ATCCACAGCTCATCTGACCCCTGTCTATCCAGG*

GTGGGAGAAAGCTGATTGCTGATGTGCTCTGCTTATTAAGGATGGCCATCAAGGACCC 720
WENVYGDMSCIKDVAIKEP
CTGTTGAGCTGGAGACCACAGCTGTCACCCATGCTCCTCTAAAGGAGGTTG
LVVDVPQLVVTNAACLIEKV
GACATCTACACGATCCAGTGAGGAGCTACCTCCACCTGCACCCTGCTGGCAATG 840
DIYTVKDELTFSTSFCLQV
AAAGGAAAGACTCACGCTGGTCGCCTGCTCTACAGGCGCTGCTGTTGGATGAAG
KRDNYVHALVAYFKAVRGKL
ACCAGGCCCAGCTATGAGAATGAGAGGTGGTAGTATGGTGTATCATCCAAAGTGCGATTGCC 900
TPSPYMRASHVWVIKSMSG
AAAACACTGTAGCTTGGAAAAAGGCCATTGGCCTTGAAGCTAAGCAT.... 1008
KTLM*T

....CCAGTCTTCTAGTCCCGTACACACACTGAAAGCAGACTGTGTTCTCTAATGAGACTACC
TAACAGTGAGACTGGCGAGGAAGACTTTGGCAGCAATTGGGAATGACCCAAATGCCAAA
AACAGTGAGACTGGACTCTTACCATGGAACAGCTGACTCTACAGGCTAGCTGTCGACTCTTCT
CTTGGTCCACGAGACTACCGAGATGCCGCTGAGAGACGCGCTGCCCTTCGAGAG
GGGGCTGCCGGGGGATGGCGCTTGGGGGATGGGGGATGACATGTGCTGACTGTGTTTTCTACATAA....
### Figure 4.8

**Alignment of Nrd and YBR0320 (ODP1) protein sequences**

<table>
<thead>
<tr>
<th></th>
<th>Nrd</th>
<th>ODP1</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>VSOQQAESSEKPNAEDMTSDKYFDSYAHFGIHEEMLKDEVRTLTYRNSM 70</td>
<td></td>
</tr>
<tr>
<td>71</td>
<td>FHNHRHELKDVKVLVDVSGTGILOMQFAAKAGAKVGIECSSISDYAVKIV 120</td>
<td></td>
</tr>
<tr>
<td>121</td>
<td>KANKLDHVVTIIKGVVEVELPVKVIDIISVMGYCLFYESMLNVTLHA 170</td>
<td></td>
</tr>
<tr>
<td>171</td>
<td>RDKWLAPDGFLIFPDRTLYVTAEKDQYKYKHMENVYGFMDSCIKDV 220</td>
<td></td>
</tr>
<tr>
<td>221</td>
<td>AIKEPLVDVFDPQQLVTNACLIKEVDITYVKVEDLTFSPCLQVKRNDY 270</td>
<td></td>
</tr>
<tr>
<td>271</td>
<td>VHALAYFKAHRQKTLSCPYMSAGSWVIIKSGMSKTLMT*KRPIALTVRH 321</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>MSKTAVKDSATEKKLSESEQHYDNSYDHGIHEEMLQDTVRLSYRRAI 50</td>
<td></td>
</tr>
<tr>
<td>101</td>
<td>ELNGFSKQITLLRKGLEDVHLPSKVIDIISVMGYFLLYESMNTLVA 150</td>
<td></td>
</tr>
<tr>
<td>151</td>
<td>FDHYLVEGGLIFDPDKSIIYLAGLEDQYKDEKLYQDVYGFYDSPVPL 200</td>
<td></td>
</tr>
<tr>
<td>201</td>
<td>VLHEPITVIVERNNTSSKIVFEDNTKISDLAFKSNKLTAKRQOM 250</td>
<td></td>
</tr>
<tr>
<td>251</td>
<td>INGIVTFVDIV.......FPAPKGRVPEFSTCPHAPYTHKWKITYFYPD 293</td>
<td></td>
</tr>
</tbody>
</table>
be longer than 20 amino acids). This information does not provide any more clues as to the structure or function of the Nrd gene product.

Several experiments confirmed that the cDNAs cloned did represent the gene in which the 7.4.2 provirus had integrated. First, Southern blots were performed with genomic DNA from the 7.4.2 cell line and uninfected ES cells. 10μg of genomic DNA was digested with three restriction enzymes which do not cut within the provirus, BglII, HindIII and StuI. The blots were then hybridized to a Neo probe, the 5A2 fragment, and the 14C1 fragment. Both of the cDNA probes hybridized to multiple bands on the Southern; however, with each enzyme tested, an additional band was present in lanes containing 7.4.2 DNA but not in lanes containing uninfected ES cell DNA (data not shown). These "shifted" bands also hybridized to the Neo probe, confirming that the shift was due to integration of the provirus. To further confirm that the cDNAs were correct, in situ hybridizations were performed on 8-8.5 day embryos using the 5A2 (short) cDNA. The results are shown in Figure 4.9. Sense and anti-sense probes were synthesized with digoxigenin-labeled UTP and hybridized to fixed embryos. Antibody conjugated to alkaline phosphatase was then bound to the probes. The color reaction was started by adding a solution of nitroblue tetrazolium chloride (NBT) and 5-bromo-4-chloro-3-indolyl phosphate, toluidine salt (BCIP), which forms a purple precipitate. The staining pattern with the anti-sense probe (Panel A) is quite similar to that seen in X-Gal assays (i.e. strong in the head region, brachial arches and neural tube), whereas the sense probe (Panel B) shows only background staining levels.

Analysis of F2 progeny

Heterozygous 03 mating pairs were set up to analyze if the 7.4.2 insertion produced a phenotype in homozygous progeny. DNA from tail biopsies was processed as described before and southern blots were hybridized to a probe synthesized from the 5' RACE product. Depending on how the probe was

Figure 4.9 In situ hybridization of wild type embryos (FVB strain) with the 5A2 cDNA. The color reaction was allowed to proceed for 2 hours. (A) Hybridization to the anti-sense probe. (B) Hybridization to the sense probe.
Figure 4.9
prepared (either by excising with $XhoI$ and $XbaI$ or with $EcoRI$ and $HindIII$), either two or three endogenous bands were visible on autorads. A representative Southern (hybridized to the $EcoRI$/HindIII probe) from one F2 litter is shown in Figure 4.10. The smallest endogenous band (about 1.2 kb) represents the wild type allele (wt) and the top band represents the mutant allele (mut). DNA from wild type progeny should show hybridization to the three endogenous bands but not the shifted band, whereas DNA from heterozygous progeny will hybridize to all four bands. DNA from homozygous progeny should hybridize to the shifted band but not to the 1.2 kb endogenous band. Lane 4 is the only sample that does not hybridize to the shifted band, indicating that this was the sole wild type pup from this litter. In the other 8 lanes, the probe hybridized to all four possible bands, indicating that all of these progeny are heterozygous. Out of 48 offspring genotyped, no homozygous progeny were observed. Thirty four heterozygotes and fourteen wild type pups were recovered, a ratio of approximately 2.5:1. This indicated that the Nrd gene product is required for embryonic development.

In order to determine when homozygous embryos were dying, timed matings of heterozygous mice were set up, and females were sacrificed at specific stages. Over half of the females, despite the fact that they exhibited mating plugs, were not pregnant when sacrificed. Unfortunately, due to the limiting numbers of heterozygous females and time constraints, additional matings were not possible. Thus, these data are not as complete as I would have liked. At day 12.5, one litter was obtained. Four out of the ten decidual sites were in an advanced state of decay, indicating that these embryos had been resorbed. The remainder of the embryos were stained with X-Gal; three did not stain (wild-type) and three stained blue. One litter was also obtained at day 10.5. Again, four out of ten decidual sites were in the process of decaying. No embryonic material could be extracted from these, indicating that the embryos had been resorbed for quite some time. The other six embryos were fixed for future $in situ$ analysis and the yolk sacs were used to isolate genomic DNA, four of which actually yielded DNA. These samples were genotyped by Southern blot; two were wild type, and two were heterozygous (data not shown). No day 8.5 embryos were obtained, but one litter of day 7.5 embryos was obtained. No decidual decay was observed, but only 7 embryos were successfully dissected out from a possible total of 9 embryos. One of these appeared much smaller than the others (data not shown). Attempts to further dissect this embryo away from maternal tissue for genotyping failed, as the embryo fell apart. The rest of
Figure 4.10

Southern analysis of O3 F2 progeny. Tail DNA from O3 F2 progeny was processed for southern analysis as described before. Blots were hybridized to a probe made from the EcoR1-HindIII fragment of the 5' RACE product. Lane 4 contains DNA from a wild type mouse, whereas the rest of the lanes contain DNA from heterozygous progeny.
the embryos were fixed and dehydrated for future in situ analysis. Two day 6.5 litters were obtained (from a total of 5 females!). One female had only two implantation sites, both of which contained embryos (one smaller than the other). The other female exhibited eight implantation sites, from which 5 normal looking embryos were obtained. Because of the difficulty in dissecting out day 6.5 embryos, I can not say whether the other 3 embryos were actually missing from the decidua or whether I simply missed them. It is possible that if they were extremely small, I would not have found them. No resorbed embryos were ever observed in numerous heterozygote (male) by wild type (female) matings performed for LacZ expression analysis. This data points towards an early death of homozygous O3 mice, possibly within days after implantation. However, additional litters between the ages of 6.5 and 8.5 days must be analyzed in detail to confirm this hypothesis.
Discussion

The three mouse lines described here were the product of an *in vitro* screen to isolate gene trap-induced mutations in developmentally regulated genes. The embryoid body screen was an accurate and efficient way of identifying both novel and known developmentally regulated genes. A total of seven ES cell clones containing proviruses integrated into developmentally regulated genes were transmitted to the germline. Two of the proviral insertions (both isolated by Jin Chen) occurred into the previously identified genes *REX-1* (Hosler *et al.* 1989) and *ECK* (epithelial cell kinase) (Lindberg and Hunter 1990; Ganju *et al.* 1994; Ruiz and Robertson 1994). It is interesting to note that this is the second time our laboratory has isolated a gene trap insertion into *REX-1* (von Melchner *et al.* 1992). Although this might indicate that the *REX-1* locus is a favorable site for activating U3 gene expression, it is quite possible that this was just coincidence. Over 500 proviral insertions sites were recently cloned and sequenced using the U3-Neo shuttle vector, none of which corresponded to the *REX-1* locus (Hicks *et al.* 1994).

Two of the three mouse lines discussed in this chapter exhibited no X-Gal staining in post-implantation embryos. In addition, neither proviral insertion was associated with an obvious phenotype in homozygous mice. The third mouse line, O3, exhibited wide-spread X-Gal staining in post-implantation embryos, but little staining in blastocysts. Cloning of the proviral insertion locus revealed a novel mouse gene (*Nrd*) with significant homology to an open reading frame found on yeast chromosome II (Smits *et al.* 1994). Embryos homozygous for this insertion die around 6.5 days p.c., shortly after implantation into the uterus.

*The mystery of the missing phenotypes*

One unexpected outcome of this study was that only one of the seven insertions isolated by Jin Chen and myself caused an observable phenotype in homozygous mice, compared to the high frequency of embryonic lethal phenotypes (30-40%) seen in other gene trap studies (Friedrich and Soriano 1991; Skarnes *et al.* 1992; von Melchner *et al.* 1992). Although we predicted that pre-selecting for developmentally regulated genes would increase our yield of embryonic phenotypes, this was not the case. Several factors could have
contributed to this result, including the functions of the genes themselves, the possible effect of functional redundancy, and the circumstances surrounding proviral integration.

Three of the seven clones showed no staining in post-implantation embryos. It is possible that we selected for integrations that are not required for development by selecting genes which are repressed upon differentiation. Alternatively, it is possible that these clones constitute the background for this screen; their expression in ES cells and blastocysts could just be an effect of a general loosening of transcriptional control, perhaps due to decreased methylation in these cells (Monk et al. 1987; Kafri et al. 1992). These possibilities will be difficult to assess without cloning some genes from this category and analyzing their functions in vivo and in vitro.

An alternative explanation could be that some of the genes we have mutated have overlapping functions with other gene products. For instance, in the case of ECK, other receptor kinases from the elk/eph family might be able to compensate for the loss of ECK. No phenotypes have been associated with knockouts of ECK or other members of this receptor kinase family yet (E. Robertson, personal communication to J. Chen). Functional redundancy of eukaryotic genes has been observed in a number of systems [reviewed in (Thomas 1993)] and is defined as the ability of two or more genes to compensate for each other in loss of function experiments. The genetic test for redundancy between two gene products is the difference between mutations in a single gene product and mutations in both (or more, in some cases). If the double mutant has a more severe phenotype than that observed in either single mutant, those genes are considered at least partially redundant. This can result in gene products not being assigned the proper role in a pathway. For instance, in Drosophila, the sloppy paired gene products (slp) appear to be biochemically equivalent (Cadigan et al. 1994). Early genetic screens did not identify any mutations in slp2, but did identify mutations in slp1. When both genes were inactivated in later studies, a much more severe phenotype than that observed in either slp1 or slp2 mutants resulted. Although these two genes can partially compensate for each other in functional assays, slight differences in temporal or spatial regulation may allow them to mediate different processes in wild-type embryos.

In vertebrates, targeted mutagenesis studies have revealed numerous cases of possible functional redundancy (Rossant 1991; Soriano et al. 1991;
Targeted mutations of genes thought to be involved in normal embryonic development often did not result in the expected phenotypes. For instance, a knockout of *Wnt-1* (the homologue to the *Drosophila* segment polarity gene *wingless*), which is expressed in the developing nervous system, caused severe defects in the mid-brain and cerebellum which resulted in death shortly after birth. It was expected that mutations in *en-2*, a homologue of the *Drosophila engrailed* gene which interacts with *Wnt-1* and is expressed in the same tissues, would result in a similar phenotype. However, a mutation that deleted the homeobox domain in *en-2* resulted in viable mice with very mild cerebellar defects. It has been hypothesized that the defect in *en-2* is compensated for by the related gene *en-1*, which shares the same expression pattern (Rossant 1991). This needs to be investigated by examining the development of mice deficient in both genes.

Another aspect of functional redundancy is that it prevents one from making any assumptions about gene function based on expression patterns. As mentioned above, targeted mutagenesis studies often result in unexpected phenotypes. An example of this is the knockout of the proto-oncogene *c-src*, a tyrosine kinase which may be involved in the control of normal cell growth and differentiation (Soriano *et al.* 1991). *Src* is expressed in all cells of the mouse, but its highest level of expression is seen in platelets and neurons. However, embryos homozygous for a disruption in *c-src* did not show any detectable abnormalities in either the brain or platelets. Instead, these embryos died within the first few weeks after birth due to osteopetrosis (a defect in osteoclasts characterized by a decrease in the rate of bone resorption with little or no change in the rate of bone formation). This pointed towards an unexpected role in bone formation for the *src* gene product. *Src* shows a high degree of similarity to two other widely expressed members of the same tyrosine kinase family, *yes* and *fyn*. It is possible that an overlap in function by these three proteins prevented a more severe phenotype in cells which normally express high levels of *src*. Targeted disruptions of several *Hox* genes had similar results [reviewed in (McGinnis and Krumlauf 1992)]. Targeted knockouts of *Hox-1.5* and *1.6* resulted in abnormalities in some but not all structures that normally express these genes. Because *Hox* genes from other clusters have similar domains of expression, they might be able to partially compensate for the loss of these genes.

How might the phenomenon of functional redundancy affect this screen? Unlike traditional genetic screens which select for interesting mutant phenotypes,
our screen does not discriminate against integrations into genes which may have functional homologues. Because the genes we identify are often novel, it may be difficult for us to interpret negative (i.e. no phenotype) results when analyzing homozygous mutants in vivo. Often, very subtle tests are required to identify specific functions of genes which do not exhibit obvious functions [discussed in (Wolpert 1992; Nüsslein-Volhard 1994)]. Without knowing anything about a gene, it is extremely difficult to guess where to look for such subtle phenotypes, especially if the gene is not expressed during post-implantation embryogenesis. In addition, the targeted mutagenesis data remind us that gene expression patterns do not necessarily predict phenotypes. This problem is not specific to the screen described here; all groups using LacZ as an insertional mutagen must deal with this possibility.

A third possibility involves the integration of deleted versus wild-type proviruses. In theory, insertions of a wild-type provirus into active chromosomal loci need not always be mutagenic. For instance, integration into a 5' non-coding alternatively spliced exon would allow expression of the transgene but may not necessarily disrupt gene function. Alternatively, insertion of the provirus into intron sequences might allow expression of the provirus, but wild-type gene transcripts could also be produced by splicing around the provirus. However, deletions in the provirus which result in the insertion of one anomalous polyA site into the genome instead of two poly A sites might decrease the mutagenic potential of the gene trap. The following situations have been observed in our laboratory:

1. Wild-type provirus in 5' non-coding exon (fug1): mutagenic.
2. Wild-type provirus in intron (Nrd): mutagenic.

At this point, it is difficult to formulate any rules for when a provirus insertion disrupts gene function. The insertion loci of other deleted proviruses must be cloned in order to determine whether or not these have disrupted gene function. Thus, although it is possible that the deleted proviruses are less mutagenic than wild-type proviruses, this cannot be concluded until a larger sample size is analyzed.
The O3 gene product: Form, function and further experiments

The 7.4.2 ES clone contained a proviral integration into a gene displaying significant homology to a yeast gene of unknown function. According to X-Gal staining, this gene is expressed at very low levels in ES cells and in blastocysts, but its expression increases greatly after implantation. The gene is widely expressed in the developing embryo, but appears to be expressed at higher levels in some tissues, particularly neural. Because of its predominant expression in the brain, I have suggested that this gene be called "Nrd", for Neural RegionalizeD. The analysis of Nrd presented here is quite incomplete; therefore, a number of experiments which would help to elucidate this gene's function are also proposed.

The sequence of the 5' RACE product indicates that the provirus has integrated into an intron. Genomic sequences flanking the provirus should be cloned to confirm the sequence of the 5' RACE product and verify where in the gene the provirus has integrated. As neither of the two enzymes used here (Hinfl and Mse I) were suitable for the amplification of genomic sequences by inverse PCR, a larger panel of enzymes should be utilized. In addition, the use of alternative splice sites should be examined. The cDNA sequence indicates the possibility of at least two differentially spliced exons. To determine whether or not these exons are alternatively spliced in undifferentiated and differentiated ES cells, one could use a Reverse Transcriptase PCR assay. Primers complementary to sequences flanking the exons in question can be used to determine which exons are utilized in either cell type.

The X-Gal staining patterns seen in heterozygous O3 embryos indicated that the Nrd gene product was up-regulated upon differentiation. In blastocysts and ES cells, very little LacZ expression is detected by X-Gal staining. Proviral transcripts could not be detected in clone 7.4.2 by Northern blot (when hybridized to a Neo probe), although they could be seen by RNase Protection. When a Northern blot was hybridized to a 5' RACE probe, however, an endogenous transcript of approximately 1.2 kb was detected. This indicates that the cellular-proviral fusion transcript is expressed at lower levels than the endogenous cellular transcript. Although it is possible that transcription of the proviral reporter gene is being repressed by the RBS-mediated silencing mechanism described in Chapter 3, it is more likely that the βgeo transcript is less stable than the
endogenous transcript. *LacZ* and *βgeo* transcripts have consistently been more difficult to detect by Northern analysis than *Neo* transcripts in our laboratory.

In post-implantation embryos, comparison with *in situ* staining indicated that X-Gal staining of heterozygous embryos correctly reflects endogenous gene expression. *In situ* hybridization of an 8.5 day embryo showed diffuse staining all over that was stronger in the head region, brachial arches, and neural tube. However, this should be verified by examining sections of embryos. Because the X-Gal staining is so diffuse, it is not easily visible in thin sections such as those produced by paraffin and plastic sectioning. Vibratome sectioning, however, can produce sections between 40 and 50 microns in width, in which one could more easily locate X-Gal stained tissues. Alternatively, X-Gal staining or *in situ* hybridizations could be performed on cryostat sections of fixed, unstained embryos. These are more sensitive methods for the detection of transgene expression.

In addition, we need to prove that the phenotype associated with homozygous O3 mice is due to the proviral insertion into *Nrd* and not some other random mutation in the ES cell clone. In order to do this, we first need to determine if there are any wild-type *Nrd* transcripts in homozygous embryos. This can be done by doing *in situ* hybridizations of early post-implantation embryos with a riboprobe complementary to *Nrd* transcripts downstream of the viral integration site. If the lack of wild type *Nrd* transcripts is responsible for embryonic death, then one should be able to recover abnormal-looking embryos which do not hybridize to the *in situ* probe. If post-implantation homozygotes can not be obtained for *in situ* analysis, then homozygous ES cells could be isolated from blastocysts and studied in culture. RNase protection assays and Northern blots of RNA from the homozygous cells should provide the desired information.

Embryos homozygous for the proviral insertion into *Nrd* probably die around day 6 of development, shortly after expression of the *βgeo* reporter gene becomes visible by X-Gal staining. Mutations in several kinds of genes could cause an early embryonic lethal phenotype. For instance, mutations in genes required for embryo implantation might either prevent implantation altogether or result in inefficient formation of the placenta, which mediates metabolite exchange between the embryo and the mother. As we do see evidence of resorbed embryos later in development, it is unlikely that the *Nrd* gene is involved in embryo implantation. In addition, no expression is evident in extraembryonic tissues, indicating that the *Nrd* gene is involved in the development of the embryo.
Mutations in structural components of a cell might also result in early embryonic death. For example, a knockout of the anti-oncogene NF2 results in growth arrest around day 6 p.c.; developmentally delayed embryos are seen at days 6.5 and 7.5 before resorption occurs (A. McClatchey and T. Jacks, personal communication). Gene products involved in signal transduction might also produce a similar phenotype. For instance, the fug1 gene, which also results in death around day 6 p.c., is thought to encode a GTPase Activating Protein (GAP) for the G-Protein Ran (personal comm. to H.E.R.).

The data presented here do not allow one to predict any function for the Nrd gene product at this time. Additional analysis of early embryos (days 5.5 to 8.5) will be required to determine not only the time of death (or growth arrest) but also any phenotypes associated with death. As it might be difficult to determine the gene function in vivo during that period of development, several other approaches could be considered. One possibility would be to knock out the yeast ODP1 gene product, which shares considerable homology to the Nrd gene. The gene function might be easier to dissect in yeast, as one could also screen for suppressor mutations that compensate for any phenotype associated with a null mutation. Another possibility would be to study the growth and differentiation characteristics of homozygous ES cell lines in vitro. These cells could also be used to generate chimeric embryos to determine if they can contribute to any embryonic structures. Similar experiments with ES cells containing a targeted mutation in GATA-1 indicated that the mutant ES cells could not contribute to hematopoietic lineages in chimeras (Pevny et al. 1991).

In conclusion, the in vitro screen which we have developed has great potential for identifying genes involved in early embryogenesis. Two of the seven clones transmitted to the germline of chimeric mice exhibited interesting X-Gal staining patterns, one of which was the homozygous lethal insertion into the Nrd gene. Although the rest of the clones (excluding REX-1) did not result in any obvious phenotypes or specific transgene expression patterns, we can not rule out the possibility that these are not interesting genes. Further analysis of the insertion loci is required to determine the circumstances surrounding expression of the proviral reporter gene and the possible functions of the endogenous genes.
Material and Methods

PCR Cloning Methods

Two methods were used to clone cellular flanking sequences from regulated clones. Inverse PCR, which is based on the protocol developed by Harald von Melchner in the Ruley laboratory (von Melchner and Ruley 1989), is diagrammed in Figure 4.11. The method involves cutting genomic DNA with a common restriction enzyme, circularizing the restriction fragment, and recutting the circular fragments with \textit{PvuII}, which cuts inside the LTR and in \textit{env}. This prevents the amplification of viral sequences flanking the 3' LTR. Primers from the 5' LTR are then used to amplify internal sequences, which happen to be cellular flanking sequences. This method will not work if the flanking sequences restriction site is either too close or too far from the provirus. Therefore, a number of different enzymes are often tried, but mostly \textit{Hinfl} or \textit{MseI}. In the cases described here, \textit{Hinfl} was the only enzyme which resulted in amplified flanking sequences. 5'RACE (diagrammed in Figure 4.12) is based on a protocol from Frohman and colleagues (Frohman et al. 1988) which involves generating cDNAs from total RNA with a primer homologous to the LTR. These cDNAs are then tailed with CTP, and PCR is used to amplify the cDNA sequences using a universal primer for the cDNA tail and another internal LTR primer. If 2° PCR is necessary, one can perform (half) nested PCR using the U3H primer, which is complementary to the first 21 nucleotides of the viral LTR. The sequences of all the PCR primers used is given below:

\begin{itemize}
  \item \textbf{Mse1}: 5' ATCATCGCGAGCCATGGTGCCCTCC 3'
  \item \textbf{Hinfl}: 5' GTAAGCTATTACGGTCAATCCGCCCTTTG 3'
  \item \textbf{U3H}: 5' CCTACAGGTGGGTCTTTTCAT 3'
  \item \textbf{LacZ 3 (RACE)}: 5' CCTCTTCGCTATTACGCGAAG 5'
  \item \textbf{LacZ 2 (RACE)}: 5' CTGCAAGCGATTAAGTTGGG 3'
  \item \textbf{Anchor Primer}: 5' CUACUACUACUAGCCACGCGT...
  \item \textbf{...CGACTAGTACGGGGGIGGGGIGGGIIGG} 3'
  \item \textbf{UAP}: 5' CUACUACUACUAGGCCACGCGTACTAGTAC 3'
\end{itemize}
**Inverse PCR:**

Inverse PCR was used to clone cellular flanking sequences from clones 2.4 and B2-3. 15 μg of genomic DNA from cell lines was digested overnight with *Hinfl* (New England Biolabs). Digested DNA was extracted twice with 1:1 phenol: chloroform and once with chloroform. The DNA was precipitated with 0.5 μg glycogen as a carrier and resuspended in 50μl TE. 2μg of digested DNA was ligated overnight at 15°C in 400μl total volume to encourage intramolecular ligations. Ligation products were purified using the Wizard™ DNA Clean-up System from Promega and digested with 40 units of *PvuII* (NEB) for at least 4 hours. The digestion products were also purified via Wizard™ and resuspended in 50μl dH2O. PCR was performed using the following reaction conditions:

10 μl target DNA  
5 μl Mse1 primer (20 μM)  
5 μl Hinfl primer (20 μM)  
52.5 μl sterile dH2O  
16 μl dNTPs (200 μM each)  
10 μl 10x PCR buffer *  
1 μl formamide  
0.5 μl Taq Polymerase

PCR cycles:  
1x: 94°C for 2 minutes  
40x: 94°C for 1 minute  
55°C for 2 minutes  
72°C for 3 minutes  
1x: 72°C for 10 minutes

* 10x buffer contained 100 mM Tris-HCl pH 8.3, 500 mM KCl, and 15 mM MgCl₂. Aliquots of buffer were stored at -20°C.

---

**Figure 4.11** Schematic diagram of the inverse PCR protocol used to clone genomic flanking sequences upstream of proviral integration sites. Restriction enzyme sites are noted as *Hinfl* (H) or *PvuII* (P). The protocol is described in the text.
Figure 4.11
Inverse PCR

Cut genomic DNA with HinfI

Ligate fragments

Cut with PvuII and amplify
5' RACE (Rapid Amplification of cDNA Ends)

5' RACE was used to clone cellular fusion transcripts from clones 2.4 and 7.4.2. All materials for the reaction (except for Taq DNA polymerase) are contained in the 5' RACE kit from BRL.

A. First strand Synthesis (all materials are RNase free)

Combine the following and incubate 5-10 min. at 70°C:

- 2 μl 1 acZ primer
- 3 μg RNA
- dH2O to 15 μl

Chill 1 min. on ice, add the following and incubate 2 min. at 42°C:

- 10x buffer
- 25 mM MgCl2
- 10 mM dNTPs
- 0.1 M DTT

Add 1 μl Superscript Reverse Transcriptase and incubate 30 min. at 42°C.

Heat inactivate the RT for 15 min. at 65-70°C, spin down

Equilibrate the reaction at 55°C for 1-2 min., then add 1 μl RNaseH and incubate for 10 min. at 55°C.

Collect the reaction and place on ice.

B. Glass Max Isolation of cDNA

Equilibrate approximately 100 μl dH2O per sample at 65°C

Equilibrate binding solution to room temperature

Add 120 μl to the above reaction and transfer to a spin cartridge. Spin for 20 sec. in a microfuge at top speed.

Wash the cartridge 3x with 0.4 ml of cold 1x wash buffer by adding the buffer and then spinning for 20 sec.

Wash the cartridge 3x with 0.4 ml of cold 70% EtOH.

Transfer the cartridge to a fresh sample recovery tube and add 50 μl pre-heated H2O. Spin for 20 sec. to elute DNA.

Figure 4.12  Schematic diagram of the 5' RACE protocol used to clone cellular transcripts appended to the proviral reporter gene transcript. Primers and protocol are described in the text.
Figure 4.12
5' RACE (Rapid Amplification of cDNA Ends)

1. Copy mRNA to cDNA (RT)

2. Degrade mRNA template (RNase H)

3. Purify cDNA and tail with dCTP and TdT

4. Amplify dC tailed CDNA by PCR with Anchor Primer and Nested LacZ primer

mRNA 5' \(\rightarrow\) (A)n-3' (LacZ primer 1)

cDNA 5' \(\leftarrow\) (A)n-3'

3'-(dC)n

Anchor Primer

3'-(dC)n

Nested LacZ primer 2
C. TdT tailing

Incubate the following 2-3 min. at 94°C:
- 2.5 µl dH2O
- 2.5 µl 10x buffer
- 1.5 µl 25 mM MgCl2
- 2.5 µl 2 mM dCTP
- 15 µl purified cDNA

Chill 1 min. on ice, add 1 µl TdT and incubate 10 min. at 37°C. Heat inactivate TdT at 65-70°C, spin down and keep on ice.

D. PCR Reaction

Incubate the following at 94°C for 5 min., then 80°C for about 5-10 min (programmed into PCR cycles):
- 22.5 µl dH2O
- 4.5 µl 10x buffer
- 3 µl 25 mM MgCl2
- 1 µl 10 mM dNTPs
- 2 µl LacZ Primer 2
- 2 µl Anchor Primer
- 5 µl tailed cDNA

Add 5µl Taq DNA Polymerase, previously diluted 1:10 in 1x buffer.

PCR cycles:
- 40x: 95°C for 1 min.
  57°C for 1 min.
  72°C for 2 min.
- 1x: 72°C for 10 min.

Cloning of PCR Products:

10 µl aliquots from iPCR and 5'RACE PCR reactions were electrophoresed on 1.2% agarose gels. If products were visible, 20-30 µl aliquots were electrophoresed on 1% low melting point agarose gels and purified using Wizard™. Purified PCR products were subcloned into a T-tailed Bluescript (KS) vector (see below). If no products were visible, the gel was blotted for Southern analysis (see below). 2° PCR was performed on gel isolates corresponding to bands on the Southern. For iPCR, the primers Mse1 and U3H were used, whereas for 5' RACE, U3H and the UAP primer supplied in the BRL kit were
After the 2° PCR reaction, bands were clearly visible on agarose gels. Products were subcloned into KS and sequenced.

**T-tailed cloning vector:**

As Taq DNA Polymerase often adds an extra adenosine nucleotide at the 3’ ends of PCR products, an efficient way of cloning products without engineered restriction enzyme sites is to ligate them into a vector with thymidine "tails". This procedure is a slight modification of the protocol in Current Protocols of Molecular Biology (Ausubel et al. 1993). Note: high concentrations of Taq Polymerase, for some reason, do not create efficient T-vectors. For this reason, tailing reactions were performed with various dilutions (1x, 1/2, 1/4, 1/8 and 1/16) of Taq. Vectors from all five dilutions were tested and the three most efficient vectors (from 1/4, 1/8, and 1/16 dilutions) were pooled and used.

**Incubate the following for 2 hours at 75°C.** Tailed vector was stored at -20°C.

- 5 μg KS vector, digested with *EcoRV*
- 20 μl 5x Taq DNA polymerase buffer*
- 20 μl 5mM dTTP
- 1 μl Taq DNA polymerase
- dH2O to 100 μl

*5x buffer consisted of 200 mM NaCl, 100 mM Tris-HCl pH 8.9, 25 mM MgSO₄, 0.05% gelatin and 0.5% Triton X-100. Aliquots were stored at -20°C.

**Southern analysis of PCR products:**

Agarose gels containing PCR products were prepared for blotting by treating with Denaturing buffer (0.5 M NaOH, 1.5 M NaCl) for 30 min. and Neutralization buffer (1.0 M Tris-Cl, 1.5 M NaCl, pH 7.5) twice for 20-25 min. DNA was blotted onto Hybond-N membranes (Dupont) overnight in 20x SSC. The DNA was cross linked to the membranes using a Stratagene UV cross linker.

Pre-hybridization and hybridization took place in the same buffer at 48°C. The buffer consists of 0.2 M NaCl, 2x Denharts, 10 mM Tris pH 7.6, 10 mM Hepes pH 7.9, 5 mM EDTA and 0.2% SDS. Pre-hyb for about 2 hours, while labeling the probe. Hybridization occurs overnight.

The probe should be a short oligo which will recognize internal sequences in the PCR products. I used U3H, an oligo (also used as a PCR primer)
complementary to the first 20 nucleotides of the proviral LTR. To label the probe, the following were incubated for 1 hour at 37°C:

- 2 µl probe DNA (50 ng/µl)
- 6 µl 5x TdT buffer (from Promega)
- 11 µl dH2O
- 10 µl α32P-dATP (NEN)
- 1 µl TdT enzyme (Promega)

The probe was then purified, boiled for 5 minutes, and added to the pre-hyb.

The blots were rinsed once with 2x SSC, 0.1% SDS at room temperature. The rest of the washes were at 50°C. Wash once with 2x SSC, 0.1% SDS, once with 1x SSC, 0.1% SDS, and once with 0.5x SSC, 0.1% SDS for approximately 15 min. each. Bands were visualized by autoradiography. If the blot was very hot, a 2-3 hour exposure at room temperature was sufficient.

**Sequencing of PCR products:**

Miniprep DNA was prepared using the rapid boil-lysis method (Maniatis et al. 1982). Double stranded sequencing was performed with α35S-dATP using the Sequenase kit from US Biochemical Corporation, with the following modifications. The GTP labeling mix was diluted 1:5 before use, and the Sequenase enzyme was diluted 1:8 before use. Flanking sequences were sequenced from both strands using the M13, Reverse, KS and SK primers homologous to the Bluescript vector.

**Screening of cDNA library**

The 7.4.2 fusion transcript isolated via 5' RACE was used to screen an 8.5 day mouse embryonic cDNA library, which was obtained from Brigid Hogan. The library contained cDNAs cloned into λgt10 phage using the EcoRI site. The initial screen resulted in the cloning of two cDNAs, one long (14C1) and one short (5A2). A second round of screening using the longer cDNA as a probe was performed to isolate additional clones.

**Materials:** All materials were made as described elsewhere (Maniatis et al. 1982)

- NZY media: 10 g/l NZ amine, 5 g/l NaCl, 5 g/l bacto-yeast extract, and 2 g/l MgSO4·7H2O. Sterilize by autoclaving.
- Top agar: NZYM plus 0.7% agarose
SM: 5.8 g/l NaCl, 2 g/l MgSO\textsubscript{4}·7H\textsubscript{2}O, 50 mM Tris-Cl pH 7.5, and 2% gelatin. Sterilize by autoclaving.

Procedure:

50,000 plaque forming units (pfu) were plated with 0.6 ml OD\textsubscript{600} = 0.5 C600 host cells on 150 mm NZY plates. A total of 20 plates were used to screen 1 x 10\textsuperscript{6} plaques. Plates were grown at 37°C overnight and then chilled at 4°C for 2 hours. Plaques were transferred onto nitrocellulose membranes for 2 min., and duplicate filters were transferred for 4 min. Filters were denatured in Soak 1 (1.5 M NaCl, 0.5 M NaOH) for 2 minutes, and neutralized in Soak 2 (1.5 M NaCl, 1.0 M Tris-HCl, pH 7.5) twice for 5 minutes each. The filters were then baked in a vacuum oven for 1.5-2 hours at 80°C. Filters were prehybridized and hybridized at 42°C in the following solution: 50% formamide, 6X SSC, 5X Denharts, 0.5 %SDS and 100 μg/ml salmon sperm DNA. (Note: all Southern blots in this chapter were hybridized in the same solution) Washes were as described previously.

Five putative positive plaques were identified from the first round of screening. An inverted yellow tip was used to core the region around these putative plaques and the phage was eluted out in 1 ml of SM plus 20 μl chloroform for secondary screening. Secondary screening of these plaques revealed that two out of the five were indeed positive. In order to pick individual plaques, Tertiary screening was performed on plates with well-separated plaques. High titer phage stocks were generated from four individual positive plaques (two from each phage line) and large-scale liquid cultures were grown for phage DNA preps. Protocols for generation of high titer stocks and liquid culture were obtained from (Maniatis et al. 1982), and the lambda DNA isolation was performed using the Qiagen Lambda Maxi prep kit. cDNA inserts were isolated by digesting phage DNA with EcoRI and subcloning the fragment into KS (digested with EcoRI and treated with CIP) for sequencing purposes. cDNAs were sequenced using the same primers used for sequencing PCR products. For sequencing internal sequences of cDNA 14C1 (long cDNA), internal primers were constructed with the following sequence:

14C1 5' primer: 5' CGAGGAGATGCTGAA 3'
14C1 3' primer: 5' CCACCTCCTTTATGA 3'
In order to isolate more cDNAs, the original filters were re-screened using 14C1 as a probe. Out of 18 additional positive plaques, 4 were isolated and two of these were sequenced.

*Whole mount in situ hybridization:*

This is based on a protocol provided by Jill McMahon, (then at the Roche Institute for Molecular Biology, Nutley, New Jersey). Modifications were made in Brigid Hogan's lab at Vanderbilt University. See also *In Situ Hybridization: A Practical Approach*, (Wilkinson 1992).

*Materials:* (RNase free materials designated by a *)
1. *Fix:* 4% paraformaldehyde in PBS. Make fresh! Dissolve at 65°C with 20 μl 1 M NaOH; cool on ice and readjust pH with 20 μl 1 M HCl when dissolved. For second fixing*, add 0.25 ml EM-grade glutaraldehyde (8%, Sigma) to cooled solution.
2. *PBT:* PBS with 0.1% Tween-20.
3. *Prehybridization buffer:* 50% formamide, 5x SSC pH 5 (adjust pH with 1 M citric acid*), 50 μg/ml yeast total RNA, 1% SDS and 50 μg/ml heparin. For hybridization, the riboprobe was added to 1 μg/ml, after boiling for 5 minutes.
4. TBST: Tris Buffered Saline (0.14 M NaCl, 2.7 mM KCl, 25 mM Tris HCl, pH 7.5) plus 0.1% %Tween-20.
5. Solution 1: 50% formamide, 5X SSC pH 4.5-5, 1% SDS
6. Solution 2: 0.5 M NaCl, 10 mM Tris HCl pH 7.5, 0.1% Tween-20
7. Solution 3: 50% formamide, 2X SSC pH 4.5-5
8. Sheep anti-digoxigenin Fab conjugated to calf intestinal alkaline phosphatase. Before use, the antibody was preadsorbed by diluting it 1:2000 in cold TBST, adding 1% heat inactivated sheep serum and 5 mgs of heat inactivated embryo powder, and mixing thoroughly. The tube was rocked for 30 minutes- 1 hour at 4°C. The mixture was then centrifuged at 10,000 X g for 10 minutes at 4°C and the supernatant was used as the preadsorbed antibody. The antibody can be stored for several months at 4°C.
9. Sheep serum: Heat inactivated sheep serum was used for diluting the antiserum and blocking non-specific sites in the embryo. To inactivate
endogenous alkaline phosphatase activity, the serum was heated to 60°C for 30 minutes. 1 ml aliquots were then frozen at -20°C.

10. Embryo powder: Homogenize one litter (approximately 12) of day 12.5 embryos (I used FVB mice) in a minimum volume of ice-cold PBS. Add 4 volumes of ice-cold acetone, mix and incubate on ice for 30 minutes. Centrifuge at 10,000 X g for 10 minutes, remove the supernatant and wash the pellet with ice-cold acetone. Centrifuge again (10 min.). Spread the pellet out on filter paper or weigh paper and grind into a fine powder. Mine seemed to be somewhat more mushy than powdery, but after air drying and storing at 4°C, it seemed to be fine.

11. 0.5 M levamisole was made fresh before use.

12. NTMT (alkaline phosphatase buffer): NTM (100 mM NaCl, 100 mM Tris pH 9.5, 10 mM MgCl₂) plus 0.1% Tween-20. Make this fresh from stocks because a precipitate tends to form on standing.

13. NBT (nitroblue tetrazolium chloride), Boehringer Mannheim. Stock is 75 mg/ml in 70% dimethylformamide. Store at -20°C.

14. BCIP (5-bromo-4-chloro-3-indolyl phosphate, toluidine salt), Boehringer Mannheim. Stock is 50 mg/ml in 100% dimethylformamide. Store at -20°C.

**Methods:**

*Probe synthesis:*

This protocol was obtained from Miguel Allende in the Hopkins lab. RNA probes were synthesized from linearized DNA templates using T3 and T7 RNA Polymerases (Promega). To generate the anti-sense probe for 5A2, DNA was linearized with *Xba*I and transcribed using T3 RNA Polymerase. The 5A2 sense probe was synthesized by T7 RNA Polymerase from DNA linearized with *Xho*I. All solutions were RNase free. The following were incubated at 37°C for 2 hrs:

- 1 µg linearized DNA
- 10 µl 2.5 mM NTP mix: (2.5 mM ATP, CTP, GTP; 1.625 mM UTP, 0.875 mM dig-11-UTP)
- 5 µl 10x Transcription buffer (Boehringer Mannheim)
- 1 µl RNasin RNase Inhibitor (Promega)
- 90 units T7 or T3 polymerase
- H₂O to 50 µl

Take a 5 µl aliquot and save. Add 5 µl RQ1 DNase (Promega) and incubate 15 minutes at 37°C. Take another 5 µl aliquot and run both aliquots on an agarose
gel to quantitate the yield from the labeling reaction. The RNA band should be approximately 10x the intensity of the DNA band for a yield of approximately 10 μg riboprobe. To purify the probe, add the following and store at -20°C for about 2 hours:

- 2 μl 0.5 M EDTA
- 2 μl 10 M LiCl
- 200 μl EtOH

Spin the tubes at 4°C for 20 minutes in an Eppendorf centrifuge (top speed), and resuspend the pellet in 1 ml Hybridization Buffer by boiling for 5 minutes. Store at -20°C. The probe can be used as a 10x stock- i.e. add 10 μl for every 100 μl of hybridization solution used.

**Preparation of Embryos:**

8.5 pc embryos were obtained from pregnant FVB females. The embryos were dissected from the decidua in PBS containing 1 mg/ml BSA (to reduce embryo stickiness). The yolk sac and amnion were removed to prevent trapping of the probe, and the amniotic cavity was punctured to minimize background caused by trapping of antibody and/ or probe.

The embryos were washed in PBS and fixed in 1 ml of fixative (4% paraformaldehyde in PBS) at 4°C for about 2 hours. The embryos were washed three times on ice with 10 ml of PBS containing 0.1% Tween-20 (PBT). For storage, the embryos were dehydrated through a series of methanols in PBT (25%, 50%, 75% and 2X 100%, 5 minutes each, on ice). Embryos were stored in 100% methanol at 20°C (can be stored for up to 2 months).

**In situ Hybridization:**

Note: All steps were done in 5 ml glass vials. The vials were siliconized with Sigmacote (Sigma), rinsed with sterile water, and baked for 4 hours before use. Embryos were rocked at all steps on an Orbitron rotating platform, unless otherwise indicated. Up until step 16, embryos were rocked upright and liquid contact with the cap was avoided. After the first three 5 min. washes at step 16, the vials were placed on their sides during the washes to allow thorough mixing. Washes were carefully removed using a long Pasteur pipette (baked), and when indicated by "remove all", using a 200 μl pipetor as well. Volumes of washes are indicated in parentheses.
1. Stored, fixed embryos were bleached in 4:1 methanol/30% hydrogen peroxide for 1 hour at room temperature, and then washed with PBT 3 times for 5 minutes (3 ml).
2. The embryos were then treated with 15 mg/ml proteinase K in PBT for 5 minutes (no longer) at room temperature (1 ml).
3. To stop the proteinase K reaction, the embryos were washed in freshly prepared 2 mg/ml glycine /PBT and then washed twice with PBT for a few minutes each (3 ml).
4. The embryos were refixed in fresh 0.2% glutaraldehyde, 4% paraformaldehyde in PBS at room temperature for 20 minutes (2 ml).
5. After fixing, the embryos were washed 2x with PBT for 5 min. each (3 ml).
6. Prehybridization solution was added, mixed, and removed. Fresh prehyb solution was added and the embryos were incubated at 70°C for 1 hour (0.5 ml). Remove all.
7. Fresh prehyb solution was added, mixed and completely removed. 0.5 ml fresh hybridization solution with the riboprobe was added and incubated with gentle rocking overnight at 70°C.
Washes:
8. 2x with prewarmed Solution 1 for 30 min. at 70°C (3 ml, remove all).
9. 1x with prewarmed 1:1 mixture of Solutions 1 and 2 for 10 min. at 70°C (3 ml).
10. 3x with Solution 2 for 5 min. at room temperature (3 ml, remove all).
11. 2x with 100μg/ml RNase A in Solution 2 for 30 min. at 37°C (2 ml).
12. 2x with Solution 3 for 30 min. at 65°C (2 ml). (Start preadsorbing antibody)
13. 3x with TBST (4 ml, remove all)
14. Embryos were rocked in TBST and 10% heat-inactivated non-specific serum for 1 hour at room temperature (2 ml, remove all).
15. The embryos were incubated with the preadsorbed antibody overnight at 4°C (500 μl).
16. After antibody binding, the embryos were washed 3x with TBST containing 2 mM fresh levamisole. In order to attain low background staining, the embryos were then washed 6 times, one hour each, at room temperature in the same buffer. End over end rocking is very important at this stage. Levamisole was added to inhibit endogenous alkaline phosphatase activity. (4.5 ml, remove all after each wash).
17. The embryos were washed twice with freshly prepared NTMT containing fresh 2 mM levamisole for 20 min. each at room temperature (4.5 ml, remove all).

18. For the color reaction, the embryos were transferred to a 24 well dish precoated with Sigmacote. The color reaction was started by adding 4.5 μl NBT and 3.5 μl BCIP to 1 ml NTMT/2 mM levamisole. The dishes were rocked in the dark (covered in foil) for the first 5 minutes of the reaction, then left undisturbed for 2 hours. After 2 hours, the staining with the anti-sense 5A2 probe was quite dark, so the reactions were halted. This was done by washing with two changes of PBT/1 mM EDTA, pH 8, and then two changes of PBT. The embryos were stored in the dark at 4°C in PBT. Photography of whole embryos was carried out using 400 speed color print film and a Leica stereo microscope (Steiner lab).

Sectioning of LacZ stained embryos:

To determine if dark staining patterns were because of more dense cell patterns or if they were truly indicative of higher expression levels, X-Gal stained embryos were sectioned. Embryos were refixed and then exposed to propylene oxide, which removes lipids so that the plastic solution (Epon) can replace them. 50 ml of Epon was made up of 12 g Epon 812 (Poly (bisphenol A-co-epichlorohydrin)), 24.7 g DDSA (Dodecenyl succinic anhydride) and 0.5 μl DMP-30 (2,4,6-tris (dimethylamino-methyl) phenol), in water.

X-Gal-stained embryos were refixed in 4% paraformaldehyde/PBS for about 30 minutes at 4°C (kept dark). The embryos were then quickly dehydrated in a series of MeOH/PBS washes (50%, 70%, 95% and two times in 100%, 2-3 minutes each, on ice). The MeOH was then replaced with two changes of propylene oxide, for 5 minutes each. The propylene oxide was replaced with a 1:1 solution of Epon and prop. oxide, in which the embryos remained for about 4-5 hours. This solution was replaced with a 3:1 solution of Epon and prop. oxide overnight at room temperature. The following morning, the embryos were placed into fresh Epon, where they remained overnight. The embryos were then imbedded in fresh Epon in forming trays and the plastic was baked for 24 hours at 70°C.

Sectioning was performed in the electron microscopy lab with help from Patricia Reilly. A Duport Porter Blum 5000 microtome was used with glass knives to cut sections ranging between 1 and 3 microns by using both the automatic advance (set at 0.997 microns) and the manual advance. Some
sections were stained with a solution of 1% methylene blue and 1% azure II in 1% borate.
Chapter 5

Summary and General Discussion
Summary

Events in the first few days of post-implantation development of the mouse are not well understood at a molecular level. In an effort to gain an understanding of events surrounding gastrulation and the specification of cell lineages, it was the goal of this thesis to develop methods to disrupt genes expressed in embryonic stem cells and to screen the resulting cell clones for mutations in developmentally regulated genes. This screen exploits the fact that ES cells can form embryoid bodies in culture, which contain a wide variety of differentiated cell types. At the outset of this project, the relationship between in vitro and in vivo gene expression was unclear. However, one of the major conclusions was that changes in gene expression seen upon in vitro differentiation of ES cells do indeed predict gene activity in vivo. For instance, genes expressed in ES cells but repressed in EBs were expressed in blastocysts but repressed in post-implantation embryos. Conversely, genes expressed at low levels in ES cells but induced upon differentiation showed wide-spread expression in post-implantation embryos, but little or no expression in blastocysts. Genes which are constitutively expressed during in vitro differentiation are also expressed constitutively in mice (Reddy et al. 1992).

Both novel and known genes were identified in this screen, emphasizing its usefulness in future developmental studies. Proviral insertions into two known genes, REX-1 and ECK, were identified by Jin Chen. Both of these genes could be involved in cell-signaling events during early development. REX-1 is a zinc finger protein expressed predominantly in primitive cell lineages (Hosler et al. 1989; Rogers et al. 1991). Its function is unclear at this time, but the presence of zinc fingers indicate that it might be a transcription factor. ECK is a receptor-protein tyrosine kinase (R-PTK) from the eph/elk family of R-PTKs (Lindberg and Hunter 1990) which is thought to play a role in early hindbrain patterning because of its expression in the node and hindbrain (Ruiz and Robertson 1994). The fact that a high percentage of mutations identified by the in vitro screen involve regulatory molecules whose expression is restricted in early development confirms the efficacy of our screen.

In this thesis, insertions into three novel loci were characterized. Two of the clones, B2-3 and 2.4, contained insertions into genes which were repressed completely upon differentiation. Homozygous progeny were viable and did not exhibit any obvious abnormalities. The third clone, 7.4.2, which contained an
insertion in a gene (Nrd) whose expression was up-regulated upon differentiation, caused an embryonic lethal phenotype. Embryos homozygous for this insertion appear to die a few days after implantation. Cloning of the Nrd cDNA revealed significant homology (64%) to a yeast open reading frame of unknown function present on the Saccharomyces cerevisiae chromosome II (Smits et al. 1994). No obvious protein motifs are present in this 310 amino acid reading frame; therefore no definite function can be assigned to this gene at this time.

The in vitro screen developed in this thesis allows one to identify genes based on two criteria: patterns of gene expression in early embryos and phenotypes associated with provirus integration. By design, all of the clones transmitted to the germline of mice contain insertional mutations in developmentally regulated genes. Using the retrovirus as a "tag", we can identify genomic sequences adjacent to the integrated provirus (via inverse PCR). In some cases, these flanking sequences hybridize to single copy genes on Southern blots as well as to endogenous transcripts on Northern blots (von Melchner et al. 1992). It is then relatively trivial to screen a cDNA library and clone the gene of interest. However, despite the advantage provided by the retroviral tag, some genes are more difficult to clone than others. For instance, flanking sequences containing repetitive sequences may not be suitable for use as probes. Alternatively, a flanking sequence may not contain any readily identifiable exons. In these cases, we have used 5' RACE to clone cellular transcripts appended to the proviral transcripts or exon trapping to identify potential exons in the flanking regions (Buckler et al. 1991). For practical reasons, genes which prove difficult to clone will only be pursued if they exhibit an expression pattern suggestive of a specific role in development (i.e. ECK expression in the node) or an obvious phenotype in homozygous mutant mice.

Although priority is given to genes with distinctive patterns of expression or obvious phenotypes, other significant mutations may not fulfill either criteria. For instance, due to functional redundancy in the genome, there are a number of genes (such as MyoD and vimentin) that do not result in obvious phenotypes when disrupted (Rudnicki et al. 1992; Colucci-Guyon et al. 1994). In addition, expression patterns do not necessarily predict where the gene is required for biological function, as discussed in Chapter 4. Thus, when no obvious phenotypes are apparent in homozygous progeny, it would still be desirable to clone those genes, in order to assess if the insertion resulted in disrupted gene
function. This will tell us much about the circumstances which allow the provirus to disrupt gene function.

The work described in this thesis comprised a pilot study to assess whether an in vitro screen could be used to identify genes involved in early developmental processes. The results indicate that this screen is a valuable tool for the study of mouse post-implantation development. In the future, more comprehensive screens will increase our understanding of not only the kinds of genes which can be isolated via in vitro screens but also the circumstances which allow gene trap mutagenesis in mouse ES cells.

**Problems with this study**

Although the in vitro screen was extremely successful in predicting the expression of genes in vivo, two problems (which may or may not be related) were encountered. The first was that approximately 80% of the NeoR ES clones contained proviruses consisting of single LTR structures. These deletions could be the result of homologous recombination between the two LTRs. Although this is supported by the fact that the percentage of deleted proviruses increases with the size of the LTR insert (Reddy et al. 1992; von Melchner et al. 1992; Hicks et al. 1994; Scherer and Ruley 1994), spontaneous recombination alone can not account for the high percentage of deletions. Rather, negative selection against the primer binding site (PBS) on the wild type retrovirus is probably also playing a role. This site is associated with a stem-cell specific silencing effect on proviral sequences, probably mediated by a cellular repressor (Feuer et al. 1989; Loh et al. 1990; Petersen et al. 1991). Thus, deleted proviruses integrated into active chromosomal loci may be preferentially activated compared to wild type proviruses because of higher transcription levels. At this point, the phenomenon is still under investigation. A retroviral vector containing a mutant PBS has been constructed to evaluate the impact of the putative silencer on the percentage of observed deletions (Hicks et al. 1994).

Although they do not affect the performance of U3βgeoSupF as a gene trap, it is possible that the proviral deletions had an adverse effect on its function as an insertional mutagen. In this study, only one out of seven germ-line clones was associated with an embryonic phenotype in homozygous animals (compared to 30-40% in other studies). This clone (7.4.2) was also the only clone containing an insertion of a full length provirus. Although the lack of phenotypes could be
produced by a number of variables, including functional redundancy and the site of proviral integration, the deletions may be a contributing factor (as discussed in Chapter 4). Therefore, some modifications in the structure of U3βgeo might be advisable. One modification would be to insert βgeo into a retroviral backbone containing the mutant PBS, which might eliminate the negative selection against full-length proviruses. Another possibility would be to replace SupF with a selectable marker such as Hygro and select for double resistant (Hygro and Neo) ES clones. Two potential problems with this scheme are (a), the retrovirus may be too long for efficient packaging and (b), the already low virus titer on ES cells may become impossibly low and prevent efficient selection of "trapped" clones. As primarily activated proviruses appear to be deleted, selecting for intact proviruses would decrease the total number of clones isolated. A third possibility would be to use the U3LacZpgkNeo gene trap instead, where all NeoR clones contain intact proviruses. The disadvantage of using this gene trap vector is that LacZ-expressing clones would have to be isolated in an additional selective step. On the other hand, gene trap integrations into non-expressed loci can also be isolated, which allows one to screen for genes which become activated upon differentiation.

Future applications of the in vitro screen

This thesis has shown that an in vitro embryoid body screen can be an efficient way to identify genes regulated during early development. In the future, the general protocol described here could be applied to many specific developmental processes. Since ES cells can give rise to most differentiated cell types, this implies that ES cells and their descendants are competent to respond to most, if not all, inductive signals. In principle, then, one should be able to supply all signals required for specific differentiation protocols in trans. In principle, our ability to study specific differentiation events in vitro depends only on our ability to supply the proper external signals in the proper temporal order. Significant progress has been made in this respect; several in vitro systems have been recently described in which ES cells preferentially differentiate into specific cell lineages. Insertional mutagenesis with any gene trap allowing the selection of both expressed and non-expressed clones (i.e. U3βgeoSupF or U3LacZpgkNeo), combined with these in vitro differentiation protocols, could be used to identify genes regulated during differentiation into specific lineages. For
instance, ES cells infected with gene trap constructs could be induced to differentiate into hematopoietic lineages to identify genes involved in this pathway.

Upon spontaneous differentiation in culture, ES cells will form embryoid bodies, many of which produce blood islands similar to those seen in the visceral yolk sac in early embryos (Doetschman et al. 1985). Therefore, several labs have investigated the possibility of using the in vitro differentiation of ES cells as a model for the earliest stages of hematopoeisis observed in the yolk sac. ES cells are capable of differentiating into hematopoietic precursors of most of the colony forming cells in the bone marrow. Hematopoietic precursors are formed by differentiating EBs in an ordered manner which reflects their appearance in vivo (Schmitt et al. 1991; Keller et al. 1993). This is accompanied by the ordered activation of many hematopoietically relevant genes, including the early genes erythropoietin (Epo), the Steel factor and c-kit (Snodgrass et al. 1992). To enrich for cells of the hematopoietic lineages, ES cells can be grown in a semi-solid methyl cellulose culture matrix (Burkert et al. 1991; Wiles and Keller 1991). Alternatively, exogenous promoting factors, such as human cord serum or a combination of erythropoietin and IL3, can be added to embryoid body cultures (Snodgrass et al. 1992). Myeloid and B cell lineages have also been observed by coculturing ES cells with a stromal cell line lacking the macrophage colony stimulating factor (M-CSF) (Nakano et al. 1994). Conspicuously missing from the lineages formed in vitro is the pluripotent hematopoietic stem cell precursor, which can differentiate into all the members of the hematopoietic system (Snodgrass et al. 1992). Thus, although the in vitro system can be used to study the factors involved in the earliest stages of hematopoietic development in the embryo, it can not presently be used to produce cells with in vivo repopulating ability. Further experimentation is likely to eliminate this problem.

Other developmental processes could also be potentially studied in vitro. When undergoing spontaneous differentiation, many embryoid bodies will form muscle cells and even beating structures reminiscent of primitive hearts (Doetschman et al. 1985). Several groups have therefore developed in vitro systems using ES cells to study early events in myogenesis and cardiogenesis. ES-D3 cells transfected with MyoD1 will preferentially differentiate into contracting skeletal muscle fibers when grown in medium containing 10% horse serum and insulin (but not in 15% FCS) (Dekel et al. 1992). Embryoid bodies formed from another ES cell line, BLC6, efficiently differentiate into skeletal
muscle cells (Rohwedel et al. 1994). In the latter case, activation of muscle-specific determination genes was observed in the correct temporal order (myf5, myogenin, myoD, myf6). Risau et al. have induced vasculogenesis and angiogenesis in differentiated ES cells by growing the cells either in the mouse peritoneum or on quail chorioallantoic membrane (Risau et al. 1988). Even the patterning of the heart tube during cardiogenesis could potentially be studied, as ventricular specific markers (like the myosin light chain molecule) are expressed in embryoid bodies, despite the fact that no intact heart tube is present (Miller-Hance et al. 1993).

Another promising application of the in vitro differentiation of ES cells offers a potential way to study inductive events in the early embryo. There have been several reports that subsets of differentiated ES or EC cells include mesodermal derivatives (Pruitt 1994; Vidricaire et al. 1994; Yamada et al. 1994). These cell lineages can also respond to various inducing agents; for instance, brachyury expression (indicative of mesodermal cell lineages) is induced by treating ES cells with members of the TGFβ family such as activin and BMP (Vidricaire et al. 1994). Other labs have reported that ES cells can be induced to form mesodermal cells at a high frequency (H. von Melchner, personal communication). A screen to isolate genes regulated by a morphogen would involve treating ES cells infected with LacZ gene trap vectors with that particular factor and looking for changes in LacZ expression. Preliminary results in several labs indicate that this could be a fruitful approach. The Gossler lab has reported the identification of gene trap lines regulated by retinoic acid (Forrester et al. 1994), which is implicated in anterior-posterior patterning of embryos (Conlon and Rossant 1992). Members of the TGF-β and FGF families have also been implicated in inductive events (Kengaku and Okamoto 1993; Kessler and Melton 1994). Preliminary evidence from our lab indicated that one could also isolate ES gene trap clones which are differentially regulated by TGFβ and basic FGF (S. Reddy, C.A. Scherer and H. E. Ruley, unpublished results).

In conclusion, I have developed an in vitro embryoid body assay which accurately selects for insertional mutations in genes regulated during early mouse development. In the future, a wide variety of in vitro assays could be used in concert with gene-trapping techniques to identify genes regulated during specific developmental processes. Mutational analysis of genes identified in
these and other genetic screens will greatly add to our understanding of early mouse development.
Bibliography


Darling, S. and J. Rossant “Unpublished data.”:


Smits, P. H. M., M. De Haan, C. Maat and L. A. Grivell (1994). “The complete sequence of a 33kb fragment on the right arm of chromosome II from Saccharomycyes cerevisiae reveals 16 open reading frames, including ten new open reading frames, five previously identified genes and a homologue of the SCO1 gene.” Yeast 10: S75-S80.


DISCLAIMER OF QUALITY

Due to the condition of the original material, there are unavoidable flaws in this reproduction. We have made every effort possible to provide you with the best copy available. If you are dissatisfied with this product and find it unusable, please contact Document Services as soon as possible.

Thank you.

Some pages in the original document contain color pictures or graphics that will not scan or reproduce well.