NOTE TO USERS

The original manuscript received by UMI contains pages with slanted print and margins that exceed the guidelines. Pages were microfilmed as received.

This reproduction is the best copy available

UMI
Gene Trap Studies of the Mouse Genome and Embryonic Development

by

Mehran Sam

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy
Graduate Department of Molecular and Medical Genetics
University of Toronto

© Copyright by Mehran Sam 1998
The author has granted a non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of this thesis in microform, paper or electronic formats.

The author retains ownership of the copyright in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author’s permission.
ABSTRACT

In recent years, gene trap insertional mutagenesis strategy has been developed and successfully used to disrupt and study developmentally regulated genes in the mouse embryos. Using a gene trap approach, two independent mouse embryonic stem (ES) cell clones were identified and investigated for developmental expression pattern, disrupted gene sequences and phenotypic effects. The strategy made use of a promoterless lacZ reporter gene to disrupt and mark genomic loci and to determine the expression pattern of genes. I193 and R140 gene trapped ES cell clones were isolated in a retinoic acid (RA) response screen and were then characterized by myself as described in this dissertation.

The RA-induced Dr repetitive element and RA-induced Aquarius gene were cloned from the I193 clone. Dr elements are highly similar in sequence, show peculiar structural features and are expressed in a spatially restricted manner. The expression of Dr transcripts seems to be part of an ES cell differentiation program induced by RA. Aquarius is a novel gene with weak similarity to RNA-dependent RNA polymerases (RRP) of the murine hepatitis viruses and it contains an RRP motif. During embryonic development, Aquarius is expressed in mesoderm, neural crest and neural crest-derived tissues, and in neuroepithelium. Mice homozygous for the Aquarius gene trapped allele are viable, are normal in size and weight, and breed normally.

The gene trapped locus, designated myocyte maintenance (mym), was identified and characterized in the R140 clone. mym is expressed in heart, limbs, gut, meninges and choroid plexus. The disrupted mym locus was transmitted through the mouse germ line
on to different genetic backgrounds. Homozygous mutant embryos died during midgestation due to heart failure. Myocardial wall and trabeculae were found to be diminished in thickness and size, and were hypoplastic. In mutant myocytes the organization of actin-myosin filaments and cell junctions were disrupted and the expression of myosin light chain isoforms was deregulated. mym is required for the maintenance and organization of heart myocytes during midgestation development. The results presented in this thesis indicate the facility of gene trapping to identify and study developmentally regulated mammalian genes and genetic elements.
Acknowledgments

I would like to express my appreciation and thank my supervisor, Alan Bernstein, for providing a rich and fertile environment to learn and to do science. I also thank the members of my graduate committee, Henry Krause and Janet Rossant, for their help and guidance throughout this work. The members of the Bernstein lab, including postdocs, graduate students and technicians also provided me with guidance and assistance whenever I needed them. For this and for their friendship, I thank them all. I take with me many valuables from this period of my studies including many lasting memories. I dedicate this work to my dear wife, Azita, and to my parents.
Attribution of Data

I am responsible for all the experiments reported in this thesis except the following. The original retinoic acid gene trap screen and the isolation of gene trap clones were performed by Wolfgang Wurst and Lesley Forrester. The blastocyst injections and generation of chimeric animals were carried out by Sandra Tondat. FISH analyses were performed by Henry Heng. In Chapter 3, RNA in situ hybridization was performed by Michael Klüppel, and the whole mount in situ hybridization by Ou Jin. Transmission electron microscopy in Chapter 4, was done by Douglas Holmyard and myself.
# Table of Contents

ABSTRACT ........................................................................................................... ii
Acknowledgments ................................................................................................. iv
Attribution of Data ................................................................................................. v
Table of Contents ................................................................................................... vi
List of Tables and Figures ....................................................................................... viii

**Chapter 1**  
Historical Background ....................................................................................... 1  
Genetics of model organisms ............................................................................. 1  
The mouse as a model for mammalian development ...................................... 6  
Chimeric mice ....................................................................................................... 6  
Mouse embryonic stem cells ............................................................................... 8  
Transgenic mice .................................................................................................... 9  
Gene targeting by homologous recombination ................................................ 11  
Gene trap insertional mutagenesis ................................................................. 13  
Retinoic acid response gene trap screen ......................................................... 20

**Chapter 2**  
A Novel Family of Repeat Sequences in the Mouse Genome Responsive  
to Retinoic Acid ................................................................................................. 23  
ABSTRACT ........................................................................................................... 24  
INTRODUCTION .................................................................................................. 24  
MATERIALS AND METHODS .......................................................................... 26  
Gene trapped embryonic stem (ES) cell lines and tissue culture .................. 27  
5' RACE, PCR DNA amplification and cloning ........................................... 27  
cDNA cloning and sequencing ......................................................................... 28  
Retinoic acid induction, RNA preparation and Northern blot analyses .......... 28  
RNA in situ hybridization ............................................................................... 29  
Genomic screening and Southern blot analysis ............................................. 29  
FISH .................................................................................................................... 30  
RESULTS ............................................................................................................ 30  
5' RACE-PCR cloning of fusion Dr-1 transcripts from three  
independent gene-trapped ES cell lines induced with retinoic acid ............. 31  
Molecular cloning and nucleotide sequence analysis of Dr-2 and  
Dr-3 cDNAs ....................................................................................................... 32  
Dr-containing transcripts are induced by retinoic acid in ES cells ............. 37  
Developmental expression of Dr repeats ....................................................... 37  
Genomic representation and organization of Dr sequences ....................... 43  
DISCUSSION ...................................................................................................... 46

**Chapter 3**  
Aquarius, a novel gene isolated by gene trapping with a RNA-dependent  
RNA polymerase motif .................................................................................... 50  
ABSTRACT ........................................................................................................... 51  
INTRODUCTION .................................................................................................. 51  
MATERIALS AND METHODS ........................................................................ 54  
Gene trapped embryonic stem (ES) cell lines, lacZ staining and  
germ line transmission ...................................................................................... 54  
Mouse genotyping and F2 breeding in different backgrounds ...................... 55  

vi
Chapter 4
The Novel Gene Trapped mym Locus Is Required for Midgestation Heart Development

DISCUSSION

Chapter 5
Concluding Remarks
Summary and General Discussion
Dr repeats
Aquarius
mym
Conclusions

Literature Cited

vii
List of Tables and Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>15</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>22</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>33</td>
</tr>
<tr>
<td>Figure 2.2</td>
<td>35</td>
</tr>
<tr>
<td>Figure 2.3</td>
<td>39</td>
</tr>
<tr>
<td>Figure 2.4</td>
<td>41</td>
</tr>
<tr>
<td>Figure 2.5</td>
<td>44</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>62</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>68</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>72</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>74</td>
</tr>
<tr>
<td>Figure 3.5</td>
<td>78</td>
</tr>
<tr>
<td>Figure 3.6</td>
<td>80</td>
</tr>
<tr>
<td>Figure 3.7</td>
<td>84</td>
</tr>
<tr>
<td>Figure 3.8</td>
<td>86</td>
</tr>
<tr>
<td>Figure 4.1</td>
<td>105</td>
</tr>
<tr>
<td>Figure 4.2</td>
<td>107</td>
</tr>
<tr>
<td>Figure 4.3</td>
<td>109</td>
</tr>
<tr>
<td>Figure 4.4</td>
<td>112</td>
</tr>
<tr>
<td>Figure 4.5</td>
<td>114</td>
</tr>
<tr>
<td>Figure 4.6</td>
<td>119</td>
</tr>
<tr>
<td>Figure 4.7</td>
<td>123</td>
</tr>
<tr>
<td>Figure 4.8</td>
<td>125</td>
</tr>
<tr>
<td>Figure 4.9</td>
<td>129</td>
</tr>
<tr>
<td>Figure 4.10</td>
<td>131</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>70</td>
</tr>
<tr>
<td>Table 4.1</td>
<td>117</td>
</tr>
</tbody>
</table>
Chapter 1: Historical Background

As biologists, we want to understand the living world in all its splendor and wonder. It is perhaps our awe for the breathtaking beauty and variety of life forms that drives us in our desire to know more. Or it is perhaps our fortitude that we live just a short while, yet our appetite is to fathom so much. Every time we learn something about a living organism, we have learned something about ourselves. This thesis is about using genetics and gene trapping in the mouse to understand something about mammalian development. Chapter One provides a short overview of the developmental genetics of model organisms and then briefly describes the historical development of the techniques and fields of studies that have made gene trapping in the mouse useful and possible. This chapter is not intended to be comprehensive, rather only to provide a general overview for an appreciation of the scope of the field.

Genetics of model organisms

The most powerful tool in the arsenal of genetic analyses is the study of mutations. By mutational analyses we gain knowledge about the function of genes even before we have any information about the genes involved. The recent explosion in our understanding of developmental processes at the molecular level is largely due to the extensive mutational analyses, performed or in progress, in several model organisms, notably *Saccharomyces, Arabidopsis, Caenorhabditis, Drosophila*, laboratory mice and most recently in the vertebrate *Danio rerio* (zebrafish).

The most widely used eukaryotic microorganism for genetic analyses is the yeast *Saccharomyces cerevisiae* (Guthrie and Fink, 1991). The rapid growth and ease of mutant
isolation makes *Saccharomyces* an ideal genetic system in which to unravel cell biology, cell division, metabolism and chromosome structure and function. In addition, the stable existence of yeast in both haploid and diploid states allows for facilitated analysis of many mutations that may not be otherwise possible in other organisms. Mutagenesis techniques such as gene targeting, disruption, and replacement, as well as allele rescue, complementation and transformation are well established in yeast and make this organism the most powerful system for biological studies of eukaryotes. In the late 1970s, reproducible transformation of yeast was achieved by introducing purified exogenous DNA. Today, yeast remains the only eukaryotic organism that can be transformed conveniently with synthetic oligonucleotides, allowing direct production of numerous gene alleles which can in turn be studied in the whole organism. This is a powerful molecular tool and a highly desirable one in any organism since it makes it feasible and convenient to investigate the multiple domains and functions of any gene product. In addition, yeast development of mating types, polarized morphogenesis and their response to mating pheromones have been investigated as a model to understand eukaryotic development at the molecular level (Chant 1996; Leberer *et al*., 1997). The complete yeast genome sequence is now available as a first complete eukaryotic genome which will be used as a reference point for other eukaryotic genomes (http://speedy.mips.biochem.mpg.de/mips/yeast/). The complete set of expressed genes, the transcriptome, from the yeast genome is also now under investigation and has made feasible genome-wide expression studies in eukaryotes (Velculescu *et al*., 1997).

*Arabidopsis thaliana* is a small flowering plant. Its diploid genetics, rapid growth cycle and small genome size has made it the model for a plant genome project and for developmental molecular genetic studies. Large scale insertional mutagenesis and cDNA and genomic mapping and sequencing are underway (Azpiroz-Leehan and Feldmann, 1997; Delseny *et al*., 1997). Approximately 4000 mutants have been so far identified and more than 40 genes isolated. These efforts have been instrumental in advancing our
knowledge of the physiology, biochemistry and development of plants. An efficient insertional mutagenesis enhancer and gene trap system has also been developed in Arabidopsis that permits gene identification based on expression patterns during development (Sundaresan et al., 1995).

Since the 1960s, the nematode worm Caenorhabditis elegans has been utilized as a model organism to study animal development and behavior (Riddle et al., 1997). Rapid life cycle, small size, ease of cultivation, inbreeding by the self-fertilizing hermaphrodite and crossing with males all offers great potential for genetic analysis in this organism. Anatomical simplicity (less than 1000 cells, the developmental lineages of all of which are now known), large number of progeny and small genome (60% of the genome sequence complete, http://www.sanger.ac.uk/~sij/C.elegans_Home.html) has contributed to the astonishing progress we have made in understanding the genetics and development of C. elegans. The ultimate goal is to determine the role of each gene in the development and function of this organism, a goal that now seems within reach. Induced and target-selected mutagenesis in C. elegans are now routine and have helped dissect genetic pathways and functional units in the worm genome. To date, 10% of the 15,000 total genes estimated by the worm genome sequencing project have been genetically identified. Mutations in about half of these genes display visible phenotypes, and about half are essential genes (have lethal alleles). Lethality may be due to a developmental blockage from egg to larval stages, sterility or maternal-effect. It is estimated that null mutations in up to 50% of C. elegans genes may yield no obvious phenotype either due to redundant genes or pathways, or lack of specific environmental conditions necessary to reveal a phenotype. On the other hand, while some essential genes have 'housekeeping' functions, i.e. are necessary for general cell processes such as intermediary metabolism, mosaic screens indicate the majority of essential genes function in only specific cell lineages. A promoter trap screen has been performed in C. elegans in which a worm genomic DNA library was cloned upstream of a lacZ reporter gene and used to create
transgenic worms (Hope 1991). Tissue specific expression of \( \text{lacZ} \) would then inform of the presence of functional regulatory elements upstream.

Earlier this century, work with \textit{Drosophila melanogaster} established many of the major principles of genetics such as the chromosome theory of heredity, multiple allelism, non-disjunction, chromosome aberrations, chemical and radiation mutagenicity, and gene mapping (Ashburner 1989). Wilkins has divided the history of developmental genetics into three fairly distinct phases based on the prevailing ideas and approaches in each phase (Wilkins 1993). The 'classical' period from the first decade of this century to about 1960, was dominated by the study of mutant phenotypes to deduce the roles of genes in development. In the second period, starting in 1961 and lasting approximately two decades, molecular concepts and various techniques of clonal analysis (tracking cell lineages by the presence of distinctive genetic markers) gained influence. The new era of developmental genetics, i.e. determining at the molecular level all the genes involved in a particular developmental process, started with the classic saturation mutagenesis approach in \textit{Drosophila} (Nusslein-Volhard and Weischaus, 1980), and was transformed by recombinant DNA technology and other molecular biological techniques. One of the main resources of the developmental biologists today is the large mutant collections of fruit flies collected during the period of classical developmental genetics.

\textit{Drosophila} was one of the first eukaryotic organisms in which vectors with dominant drug-resistant markers were used to genetically select for transformants (Steller and Pirrotta, 1985). The selectable marker used in this experiment was the \textit{neomycin} \( f \) (\textit{neo}) gene of bacterial Tn5 which confers resistance to the powerful inhibitor of protein synthesis, the aminoglycoside G418 (Jorgensen \textit{et al.}, 1979). Flies transformed with the \textit{neo} gene, under the control of a suitable promoter were resistant to G418. The \textit{lacZ} gene of \textit{E. coli}, coding for the enzyme \( \beta \)-galactosidase (\( \beta \)-gal) and assayed by histochemistry, was also one of the first markers used in \textit{Drosophila} as a reporter of gene activity (Lis \textit{et al.}, 1983). Finally, a chimeric \textit{lacZ}-\textit{P} element gene under the control of the weak \( P \)
element promoter was used in *Drosophila* in the first enhancer trap experiments (O'Kane and Gehring, 1987). The integration of the enhancer trap element near a host sequence upregulated its transcription and led to an increase in *lacZ* expression in a tissue-specific manner.

The zebrafish, *Danio rerio*, has recently become a vertebrate model system because it is susceptible to both classical embryological manipulations and extensive genetic analysis. Because eggs are fertilized externally and the embryo is transparent, virtually every cell can be examined in the live embryo by light microscopy. Zebrafish offer several advantageous features for genetic analysis, compared to other vertebrates, including a short gestation time, and the ability to make genetic mosaics, haploids and triploids (Felsenfeld 1996). These features made it possible and encouraged two independent groups to undertake large scale chemical mutagenesis screens in zebrafish with remarkable results (Driever et al, 1996; Haffter et al, 1996). A few thousand mutations have been identified and several hundred of them characterized. Mutant phenotypes have been identified with maternal effects, in gastrulation and cell movement, body axes, notochord, CNS, neural crest, blood, heart and internal organs, eye and behavior (http://zebrafish.mgh.harvard.edu/database.html). Although positional cloning of these mutations will be labor-intensive, it is achievable as tools for genetic analysis, mapping and cloning become available (Knapik et al, 1996). Insertional mutagenesis and rapid cloning of essential genes using a retroviral vector have also been successfully attempted in zebrafish (Gaiano et al., 1996), which could circumvent positional cloning. Furthermore, gene interactions can be revealed in zebrafish by enhancer/suppressor screens and conditional mutations (Johnson and Weston, 1995). Thus, zebrafish has great potential in enhancing our understanding of molecular and genetic mechanisms underlying vertebrate development.

Avian and amphibian systems will also continue to provide insights into vertebrate development through classical embryological manipulations, while the mouse
will continue to be the model of choice in understanding human genetic disorders and behavior. Creating null as well as subtle mutations in known genes using the mouse embryonic stem cell technology will also continue to provide insights about gene function, regulation and interactions in mammalian system.

The mouse as a model for mammalian development

The mouse has been the subject of embryological and developmental studies for almost a century. Its small size, relative ease of maintenance and generation of inbred strains of genetically homogeneous mice has made this organism the model of choice in many laboratories. Early work by Snell, Little, Castle and others established the basis for future studies in the fields of embryology, genetics, cancer, and transplantation (Russell 1985). Today, complicated genetic and developmental problems, i.e. genetic interactions, modifier genes and complex genetic traits, can be addressed in the mouse due to advances in genetic mapping and analysis such as the use of microsatellite repeats as genetic markers (Dietrich et al., 1992). In addition, the burst in our knowledge of mammalian development is attributable to three inter-dependent technological advances in the mouse: i) generation of chimeric animals, ii) establishment of mouse embryonic stem cells, and iii) production of transgenic mice.

Chimeric mice

A mammalian animal composed of genetically dissimilar cells is called a 'mosaic' if it is derived from a single zygote, and a 'chimera' if it is derived from different zygotes (Slack 1991). In the early 1960s, chimeric mice were generated by aggregation of two eight-cell stage embryos (Mintz 1962; Tarkowski 1961). The chimeras were generated to
investigate cell movement, assortment, and to a limited extent lineage development by using genetic (lethal mutants), radioisotopic and cytological markers (Rossant 1987). Sex-chromosome mosaicism and hermaphroditism as a model of sexual disorders were also investigated in XX/XY chimeras (Tarkowski 1965).

In 1968, Gardner succeeded in creating mouse chimeras by injecting cells into blastocyst stage embryos, thus making it possible to control the extent of chimerism by the number of cells injected (Gardner 1968). Gardner detected the extent of chimerism by coat color. Pluripotent embryonic stem cell lines were established in culture and chimeras were generated by injection of these cells into blastocysts, followed by blastocyst transfer into pseudopregnant recipient mice.

Today we can create mouse chimeras by simple aggregation of pluripotent embryonic stem cells with morulae-stage embryos (Wood et al., 1993). This new technique is less cumbersome, more economical and just as efficient as blastocyst injection. It is also possible now to generate mice and embryos that are derived completely from embryonic stem cells by aggregating stem cells with developmentally compromised tetraploid embryos (Nagy et al., 1993a).

ES-diploid and ES-tetraploid aggregate chimeras are of great value in direct analysis of dominant negative, gain of function and homozygous mutations that may interfere with normal development or germ line transmission. For example, the effect of a mutation may not be as revealing when all cells are mutant as when mutant cells are intermixed with wild type cells in a chimera. Mutant cells in a wild type environment can reveal whether the mutation is lethal for all cells or has only a temporal or tissue-specific effect. Furthermore, chimeric analysis can determine whether the gene product acts cell autonomously or whether neighboring wild type cells can rescue the mutant phenotype (Beddington 1992). Chimeras are also a rich source of differentiated tissues and tissue progenitors carrying genomic alterations (Chen et al., 1993).
**Mouse embryonic stem cells**

Pluripotent cells were known to be present in a mouse embryo until early post-implantation stage; however, early attempts to culture these cells were unsuccessful. In 1981, Evans and Kaufman established embryonic pluripotent cells in culture by direct *in vitro* culture of blastocysts (Evans and Kaufman, 1981). These cells gave rise to embryoid bodies and a complex of differentiated tissues *in vitro*, and to teratocarcinoma *in vivo*, indicating their pluripotentiality. They carried a normal karyotype and generated chimeric mice. In the same year, Martin established a similar cell line from the inner cell mass of early embryos (Martin, 1981). This cell culture was maintained in medium conditioned by teratocarcinoma stem cells.

Today several pluripotent embryonic stem (ES) cell lines capable of germ line transmission have been established. These cells proliferate on feeder layers of irradiated cells, and in the absence of the feeder cells, differentiate into embryoid bodies. Leukemia inhibitory factor can substitute for the feeder layer (Smith *et al.*, 1988). ES cells can be injected into blastocysts or aggregated with morulae stage embryos to create chimeras. When the chimerism extends to the germ cells, the genome of the ES cells, carrying any genetic alteration, can be transmitted to the progeny of the chimeras, thus creating transgenic mice with a desired genetic change (Robertson 1986). ES cells are now routinely used to introduce mutations in the mouse germ line by homologous recombination (Capecchi 1989).

As mentioned above, it is now possible by tetraploid aggregation to produce totally ES cell-derived mice, indicating the remarkable potential of cultured ES cells. However, in these experiments many ES cell-tetraploid aggregates die before reaching term making the technique inefficient and infeasible for germ line transmission of a genetic manipulation (Nagy *et al.*, 1993a).
In vitro differentiation of ES cells depending on culture conditions allows access to different developmental processes such as mesoderm commitment and neuronal and hematopoietic development. In this system, experimental conditions can be controlled and readily altered, and the amount of material available is neither as limited nor as expensive as the in vivo system. The progressive differentiation of ES cells can be easily followed by molecular markers and techniques such as cytohistochemistry or RT-PCR (Wiles 1993).

Transgenic mice

Foreign DNA sequences can be stably inserted into the mouse genome to create a transgenic animal. Transgenic mice can be created by microinjection, viral infection and manipulation of ES cells in culture. In 1976, Jaenisch infected mouse embryos with Molony leukemia virus and showed for the first time the incorporation of exogenous DNA into the mouse germ line (Jaenisch 1976). In the next five years, several groups achieved the integration of foreign DNA into the mouse genome through the microinjection of DNA into the zygote pronuclei (Brinster et al., 1981; Costantini and Lacy, 1981; Gordon and Ruddle, 1981; Harbers et al., 1981; Wagner et al., 1981; Wagner et al., 1981). Such random integrations led to various genomic rearrangements such as translocation, deletion, duplication and inversion. Before the advent of transgenesis, investigators had been primarily restricted to observation and description. Thus, the ability to introduce at will genetic changes or new genes into animals represented a dramatic advance.

To achieve transgene expression in differentiated adult cells, Brinster et al. fused the herpes simplex virus thymidine kinase gene with the metallothionein-I promoter region (Brinster et al., 1981). The fusion gene was injected into fertilized eggs which
were then reimplemented into pseudopregnant mice. The fusion gene was present in the genome of the progeny mice and it was highly expressed in the liver and kidney. In one of the first gene functional studies in transgenic mice, Löhler et al. examined the result of insertional mutagenesis by the Moloney murine leukemia virus in the \( \alpha_1 \) collagen gene in the Mov13 mice (Lohler et al., 1984). The integration at the Mov13 locus resulted in a recessive lethal mutation leading to midgestation embryonic lethality. The cause of death was determined to be rupture of major blood vessels. There were also pathological events in hemopoietic cells of the liver and necrosis of mesenchymal cells, indicating a functional role for \( \alpha_1 \) collagen in the normal development of these tissues.

In another transgenic experiment, the mouse mammary tumor virus long terminal repeat fused to the \( c\text{-}myc \) gene created an insertional mutation in the mouse genome that led to a severe defect in limb patterning (Woychik et al., 1985). The inserted element provided for a molecular link with the genetic control of limb formation and was used to examine the integration site. The integration of the transgene deleted about 1 Kb of the genomic sequences with no other gross rearrangement. The integration site was mapped to the limb deformity (\( ld \)) region and in complementation tests was shown to be indeed an allele of \( ld \). These early transgenic experiments in which the mouse germ line was stably transformed by exogenous DNA are extensively reviewed by Palmiter and Brinster (Palmiter and Brinster, 1986).

Subsequently, retroviral vectors were used to introduce transgenes into cultured pluripotential stem cells for generation of chimeric mice and germ line transmission (Gossler et al., 1986; Robertson et al., 1986). The proviral vector sequences containing \( neo \) gene were stably transmitted through several generations and provided drug resistance as well as new chromosomal molecular markers for linkage studies. They also caused insertional mutations. These and other similar experiments indicated the feasibility of using the easily manipulatable embryonic stem cells to produce transgenic animals, opening up a whole new set of experimental approaches for \textit{in vitro} selection.
and screening for a desired genetic change in cultured stem cells before generating the transgenic animal. Soon after, transgenes carrying the *E. coli* β-galactosidase (*lacZ*) reporter gene were introduced into the mouse genome and were detected with a histochemical stain to follow endogenous gene expression patterns and to study cell lineages *in vivo* (Allen *et al.*, 1988; Kothary *et al.*, 1988; Sanes *et al.*, 1986). Tissue-specific and inducible promoter sequences were also successfully used in transgenic mice to direct spatio-temporal expression patterns of transgenes during mouse embryogenesis (Kothary *et al.*, 1989; Rossant *et al.*, 1991). Today, it is possible to transfer intact yeast artificial chromosome (YAC) DNA into transgenic mice enabling the analysis of large genes and multigenic loci *in vivo* (Peterson *et al.*, 1997).

**Gene targeting by homologous recombination**

In transgenic mice, gene integration into the genome is usually random, based on non-homologous recombination, and expression is often unpredictable. The ability to target genes to specific locations in the yeast genome by homologous recombination (Hinnen *et al.*, 1978) allowed for specific gene inactivation, mutation, and correction, and prompted investigators to seek a similar technique in the mouse. Reconstruction of a mutant herpes simplex virus thymidine kinase (*tk*) gene in mammalian cells was one of the earliest indications that the genome of mammalian cells can undergo homologous recombination with exogenous DNA (Small and Scangos, 1983). Mutants of dominant selectable markers like *neo* gene introduced into mouse or human cells also generated a functional gene by homologous recombination (Kucherlapati *et al.*, 1984). Homologous recombination was also shown to correct a mutation in the selectable hypoxanthine phosphoribosyl transferase (HPRT) gene and to inactivate a wild type HPRT gene in mouse ES cells (Doetschman *et al.*, 1987; Thomas and Capecchi, 1987).
In 1988, Mansour and colleagues developed a positive-negative selection method that allowed for positive selection for the correct homologous recombination events, while at the same time selected against non-homologous events, thereby substantially enriching for the recovery of targeted events (Mansour et al., 1988). Gene targeting is now widely used in the mouse ES cells to understand gene function in the whole animal. When the targeted gene is transmitted through the germ line of a chimeric animal, it can be made homozygous by breeding and its phenotypic effects can be examined. Based on the number of knockout mice published by 1995, it can be estimated that to date more than 800 genes have been mutated by gene targeting in the mouse ES cells (Brandon et al., 1995).

The combination of homologous recombination with site-specific recombination systems such as the bacteriophage Cre-loxP has allowed the more powerful tissue- and time-specific modification of genes in the mouse genome during development (Gu et al., 1993; Orban et al., 1992). The Cre-loxP recombination system has also been used to induce site-directed chromosomal translocations and chromosome loss in the mouse (Lewandoski and Martin, 1997; Smith et al., 1995).

Genetic manipulation of mice by random transgenic integration and by homologous recombination has increasingly allowed genetic dissection of development, studies of gene function, expression and regulation, and the production of mouse models of human disease. More than 100 mouse models of human disease exist where the homologous mouse and human genes have been mutated, and in a majority of cases the mouse phenotype closely resembles the human disease, thus providing valuable models to understand disease and to devise ways to prevent or treat such diseases (Bedell et al., 1997). A transgenic database on the internet provide an update of all mouse transgenics and knockouts to date (http://www.gdb.org/Dan/tbase/tbase.html).
Gene trap insertional mutagenesis

Two main strategies for obtaining and studying mutations are: i) seeking genes and mutations that display specific phenotypes of interest (forward genetics), and ii) starting with a gene of choice and then looking for the phenotype of a mutation in that gene (reverse genetics). In forward genetics, one selects and is sure about the phenotype but may miss redundant genes or modifiers that do not play an important role under the set of conditions used. In reverse genetics, the gene of interest is molecularly defined but may lead to a variety of phenotypes ranging from lethality to no obvious phenotype. Gene trapping provides for a third alternative in obtaining and studying mutations where random gene mutations are created by gene trap vector insertion and then mutations are selected for further study based on, i) regulatory or signaling pathways involved, ii) in vitro or in vivo expression pattern of the trapped gene, or iii) partial sequence information of the gene.

In a trapping experiment in the mouse, a construct consisting of a promoterless lacZ reporter gene and a selection marker, commonly the neo gene, is randomly integrated into the genome of ES cells grown in culture. ES clones that have successfully integrated the trapping vector are selected in the presence of the drug G418. Then, histochemical staining for β-galactosidase activity reveals the expression of lacZ gene under the regulation of the trapped endogenous genomic sequences. Expression of lacZ indicates a successful trapping event.

In order to achieve lacZ expression, the exogenous lacZ construct must insert near a gene enhancer (enhancer trap), in a gene intron (gene trap), or in a gene exon (promoter trap) (Figure 1.1) (Zachgo and Gossler, 1996). In the case of the enhancer trap, lacZ carries a minimal promoter at its 5' end and its activation by genomic enhancer sequences indicates a trapping event (Figure 1.1A). In the case of gene trap, the lacZ construct lacks any promoter sequences, instead carries a 5' splice acceptor site. Thus, its integration into
an intron leads to its expression and splicing to 5' exons of the endogenous active gene (Figure 1.1B). Such a genomic insertion leads to the generation of a fusion transcript between lacZ and the 5' exons of the endogenous gene. This fusion transcript may be used in 5' rapid amplification of cDNA ends-polymerase chain reaction (5' RACE-PCR) to clone cDNA sequences of the trapped endogenous gene. The experiments described in this thesis are based on a similar gene trap strategy. Finally, in the case of the promoter trap, lacZ carries no promoter or splice acceptor sites in its 5' end. The LacZ sequence starts in this case with its open reading frame and it therefore must integrate in-frame into an exon of an active gene to be translated (Figure 1.1C).
**Figure 1.1** Structure and activation of different trapping constructs. In the enhancer trap construct, \( \text{lacZ} \) has an initiation ATG codon. In the gene trap and promoter trap constructs, \( \text{lacZ} \) may or may not have an initiation ATG codon, resulting respectively in a distinct \( \text{LacZ} \) protein or a fusion \( \text{LacZ} \) protein. [ATG]: optional ATG translation initiation codon, eRE: endogenous regulatory element, mP: minimal promoter, \( \text{lacZ} \): bacterial \( \beta \)-galactosidase gene, neo: neomycin resistance gene, P: promoter, pA: polyadenylation signal, SA: splice acceptor site, line: intron or intergenic genomic sequences, solid boxes: endogenous genomic exons, gray boxes: promoter regions on the vector, open boxes: reporter and selector genes on the construct.
A. Enhancer Trap

B. Gene Trap

C. Promoter Trap

Genomic locus after insertion
Following the enhancer trap experiments in *Drosophila* mentioned above, Gossler et al. successfully attempted gene and enhancer trap experiments in mouse ES cells. In these experiments, a *lacZ* reporter gene with a minimal promoter (i.e. an enhancer trap vector) or with a 5' splice acceptor site (i.e. a gene trap vector) was inserted into the genome and activated by flanking mouse sequences (Gossler *et al.*, 1989). Developmental regulation of *lacZ* expression was demonstrated in chimeric mice and indicated the potential to introduce insertional mutations into the mouse germ line. Using the vector sequences as genomic markers also indicated the possibility for molecular cloning of the flanking endogenous genes. Thus, these trapping experiments demonstrated that it was now feasible to: i) insertionally mutagenise random genes; ii) reveal their developmental expression pattern; and iii) clone the gene sequences using the trapping vector as an aid.

In a gene trap screen of ES cells, it was found that a trapping construct bearing the fusion *lacZ* and *neo* gene (*βgeo*) is much more efficient in picking up *lacZ* expressing trapped ES clones as compared to when the *neo* gene is independently driven by its own promoter (95% versus 5%) (Friedrich and Soriano, 1991). In this experiment similar efficiencies were observed when either plasmid-based or retroviral constructs were used. The trapped loci were transmitted through the germ line and the *lacZ* expression patterns were examined in 27 strains of mice. Nine strains showed restricted expression patterns, ten had widespread and eight ubiquitous expression patterns. Heterozygotes were crossed in 24 strains, nine of which led to embryonic lethality in the homozygotes. The rest did not manifest any overt phenotype, suggesting that a substantial portion of genes identified in this approach are not essential for development.

In another gene trap approach, three insertions were examined for sequence, expression and phenotype by transmission through the germ line (Skarnes *et al.*, 1992). The results demonstrated correct use of splice acceptor sites in the fusion transcripts and revealed distinct *lacZ* expression patterns. Abnormalities were detected in two of the
three mutant insertions in the homozygous state. Ribonuclease protection assays detected negligible amounts of normal endogenous transcripts in the homozygotes as a result of splicing around the gene trap insertion.

Using retroviral promoter trap vectors, Melchner et al. characterized nine trapped clones, one of which was the previously characterized gene, REX-1, and two of the four clones passed to the germ line resulted in homozygous embryonic lethality (Melchner et al., 1992). The murine homolog of the yeast RNA1 gene, fugl, was also identified and mutated in a gene trap screen and was shown to be required for gastrulation in postimplantation mouse development (DeGregori et al., 1994). A novel gene, jumonji, expressed at the midbrain-hindbrain boundary was captured in another gene trap experiment, and shown to be required for neural tube formation (Takeuchi et al., 1995). In the gene trap construct used, an internal ribosome entry site (IRES) was inserted between the neo and lacZ genes to allow dicistronic expression of the two genes. The nuclear transporting signal peptide from SV40 large T gene was also placed at the 5' end of lacZ to concentrate β-galactosidase in the nucleus to facilitate detection of lacZ expressing cells. Several other genes, including α-E-catenin, the Eck receptor tyrosine kinase and cordon-bleu were also identified and/or characterized in gene trap approaches (Chen et al., 1996; Gasca et al., 1995; Torres et al., 1997). In addition, in vitro prescreening of gene trapped ES cells has allowed isolation of trapped ES clones expressing lacZ specifically in the embryonic nervous system (Shirai et al., 1996). By including an N-terminal signal sequence in the LacZ protein, Skarnes et al. captured genes targeted to cell membrane and involved in mouse development (Skarnes et al., 1995).

Large scale gene trap screens have also been performed to identify and characterize developmentally regulated genes. Wurst et al. isolated 393 ES cell clonal integrations with lacZ reporter gene expression (Wurst et al., 1995). Of 279 clones that generated chimeras, 13% showed restricted expression at E8.5 embryos, 32% showed
widespread expression, and 55% failed to show any expression at E8.5. One-third of these latter clones displayed reporter gene activity at E12.5, indicating spatio-temporal gene regulation during embryogenesis for a large proportion of the genes expressed in ES cells. Baker et al. isolated 86 gene trapped ES cell clones and differentiated them to embryoid bodies *in vitro* to assay for *lacZ* expression in different cell types (a strategy they refer to as 'in vitro preselection') (Baker et al., 1997). Twenty three percent of the clones exhibited uniform expression, 22% exhibited no expression and 55% expressed *lacZ* in limited cell types. Combined with partial sequence analysis from the clones, the *in vitro* preselection is efficient and cost effective in prescreening for trapped clones of interest before generating mice.

To assess the feasibility of large scale molecular cloning, 55 cDNA clones were isolated and characterized from gene trapped ES cell lines using the SAβgeo vector (Chowdhury et al., 1997). SAβgeo vector consists of a 5' splice acceptor (SA) site and the fusion β-galactosidase-neomycin genes. Isolated DNA sequences varied between 52 and 882 bp. Thirty percent encoded known gene products including nuclear, cytoplasmic, cytoskeletal, membrane and extracellular proteins; 21% were present in the databanks as ESTs; and 50% were new. These results suggested stochastic insertion of gene trap vectors in the ES cells. Additionally, the site of the SAβgeo vector insertion within known genes was random. In a similar large scale approach, Hicks et al. devised a gene trap shuttle vector to disrupt genes in ES cells and quickly obtain genomic sequence information from the 5' flanking region (Hicks et al., 1997). Out of 400 inserts cloned and sequenced, 42 (10%) were in known genes and 21 (5%) were present in EST databanks indicating 85% unknown sequences, many of which could be intronic sequences. Gene targets identified were various and included all major classes of cellular proteins. Sixteen mutations were introduced into the germline, six of which resulted in obvious phenotypes; however, in at least one case, insertion in an intron did not ablate endogenous gene expression. Improvements in direct sequencing of RACE-PCR products
make sequence prescreening of gene trap integrations possible before further characterization is undertaken (Townley et al., 1997). Sixty percent of trapped cell lines in the study by Townley et al. were either spliced inefficiently or contained deletions of the vector arguing for caution in such experiments.

It is of great interest to know if all genes can be eventually trapped in the mouse. All genomic loci may not be equally accessible to insertion of exogenous DNA. Nevertheless, at present, it seems possible to trap all genes expressed in ES cells, estimated to be around 10,000 (Evans et al., 1997). The goal would be to create a gene trap library of ES clones and perhaps thousands of mouse strains in an international effort to extract biological information from the genomic sequences. Because gene trap strategies cannot introduce subtle or gain-of-function mutations, gene trapping will not replace gene targeting by homologous recombination; rather it provides a rapid strategy for gene discovery, mapping, mutagenesis, and expression studies.

In a gene trap screen, ES-tetraploid chimeras can be used to visualize gene expression patterns. Generation of such chimeras is less demanding than the blastocyst injection techniques and is more likely to represent accurately the endogenous gene activity as ES cells, but not the tetraploid cells, contribute to embryonic tissues.

One shortcoming of the gene trap strategy is that a single clone of each gene trap integration is available; thus independent confirmation of the resulting phenotype is not possible. In contrast, gene targeting generates several independently derived clones of targeted ES cells allowing the confirmation of the phenotype.

**Retinoic acid response gene trap screen**

By preselecting gene trapped ES cells in vitro, Forrester et al. developed a novel retinoic acid induction approach that selected for integrations into genes that lie downstream of the RA signaling pathways (Forrester et al., 1996). The screening strategy
is described in Figure 1.2. Essentially, it is based on the fact that ES colonies can be replica plated and while the master plates of colonies may be saved for further characterization, replicas of the plates are used in in vitro screens for lacZ response to RA. Once colonies of interest are detected in the screen, master colonies can be used for germ line transmission, expression studies and cloning of the trapped genes. This approach can be adapted to other receptor/ligand-mediated signaling pathways.

In the above-mentioned RA screen, 20 gene trap integrations were identified, 9 of which were induced and 11 were repressed after exposure to exogenous RA. All but one of these integrations showed unique spatially restricted or tissue-specific expression patterns during embryogenesis. Five out of six analyzed cDNA sequences were novel genes and one was the protooncogene c-fyn. Germ-line transmission and breeding of 4 integrations uncovered one homozygous embryonic lethal (R140 described in Chapter 4 of this thesis), and three homozygous viable insertions.

The work described in this thesis is based on characterization of two gene trapped clones, 1193 and R140, isolated in this RA screen. Chapter 2 describes the novel RA-induced Dr repetitive elements cloned from the 1193 gene trapped ES cell line, the Dr nucleotide sequence and its expression pattern. Chapter 3 describes the RA-induced Aquarius gene, its sequence analysis and expression pattern during embryonic development. Chapter 4 describes the identification and characterization of the mym gene, from the R140 gene trapped ES cell line. Finally, I summarize in Chapter 5 the work done for this dissertation and describe some future experiments that might be undertaken to better understand the Aquarius and mym genes.
Figure 1.2 Retinoic acid gene trap screen in ES cells. Replica plating allows for in vitro manipulation and screening of gene trapped ES colonies before further characterization. Gene trapped clones that are induced or repressed in response to RA are selected for cloning cDNA sequences and generating chimeric animals.
Chapter 2: A Novel Family of Repeat Sequences in the Mouse Genome Responsive to Retinoic Acid

Chapter 2 is a version of the following publication:
ABSTRACT

Repetitive DNA sequences form a substantial portion of eukaryotic genomes and exist as members of families which differ in copy number, length and sequence. Various functions, including chromosomal integrity, gene regulation, and gene rearrangement have been ascribed to repetitive DNA. Although there is evidence that some repetitive sequences may participate in gene regulation, little is known about how their own expression may be regulated. During the course of gene trapping experiments with embryonic stem (ES) cells, we identified a novel class of expressed repetitive sequences in the mouse using 5' rapid amplification of cDNA ends-polymerase chain reaction (5' RACE-PCR) to clone fusion transcripts from these lines. The expression of these repeats was induced by retinoic acid (RA) in cultured ES cells examined by Northern blot analyses. In vivo, their expression was spatially restricted in embryos and in the adult brain as determined by RNA in situ hybridization. We designated this family of sequences as Dr (developmentally regulated) repeats. The members of the Dr family, identified by cDNA cloning and through database search, are highly similar in sequence and show peculiar structural features. Our results suggest the expression of Dr-containing transcripts is part of an ES cell differentiation program induced by RA.

INTRODUCTION

A substantial fraction of eukaryotic genomes is made up of repetitive DNA sequences. Different classes of repeat sequences, ranging from a few copies per genome to about one million Alu family members, are present in mammalian genomes (Hellmann-Blumberg et al., 1993). The repeat units range from a few to a few thousand
nucleotides, and their genomic organization varies from tandemly repeated to widely dispersed (Darnell et al., 1990). In mammals, interspersed sequences are classified as SINEs (short interspersed elements) ranging in size from 90 to 400 bp, and LINEs (long interspersed elements) which can be up to 7,000 bp (Deininger et al., 1992). In the mouse, SINE and LINE sequences account for about 15% of the genome. These elements are dispersed throughout the genome, and there is evidence to suggest that the dispersion occurred by retrotransposition, i.e. by means of an RNA intermediate (Schmid and Maraia, 1992). While LINEs may encode their own reverse transcriptase, SINEs are too small to encode any proteins. Two major families of SINE sequences in the mouse, B1 and B2, have relatively short repeat units of ~140 bp and ~190 bp, respectively (Silver 1995). The majority of mouse middle-repetitive DNA sequences consist of numerous families and are only a few hundred base pairs in length (Schmid and Deininger, 1975).

It has been proposed that repetitive sequences may provide a mechanism for widespread and coordinated gene regulation (Britten and Davidson, 1969), although evidence for their involvement in gene regulation remains limited. Alu repeat sequences in humans are involved in transcriptional regulation (Brini et al., 1993; Saffer and Thurston, 1989; Tomilin et al., 1990), and they exhibit tissue specific patterns of expression (Koga et al., 1988; Watson and Sutcliffe, 1987). In the mouse, Alu-like repetitive sequences have been reported to confer growth- and transformation-dependent regulation on transgene expression (Vidal et al., 1993). Repetitive elements upstream of the murine erythropoietin receptor and the immunoglobulin Kappa light-chain genes negatively regulate the activity of these genes (Saksela and Baltimore, 1993; Youssoufian and Lodish, 1993).

Repetitive sequences may also be involved in gene rearrangements (Rudiger et al., 1995; Wallace et al., 1991). Alu repetitive elements have been frequently found at the site of human gene rearrangements that occur by homologous recombination, such as in sex chromosome exchange in XX males (Rouyer et al., 1987), rearrangements in the
LDL-receptor (Hobbs et al., 1986; Rudiger et al., 1991), α-globin (Nicholls et al., 1987), β-globin (Henthorn et al., 1986; Vanin et al., 1983), and apolipoprotein B (Huang et al., 1989) genes, and in the c-sis protooncogene (Smidt et al., 1990). A short, well conserved region of the Alu repeat has been proposed to be involved in gene rearrangements (Rudiger et al., 1995).

Repetitive sequences have also been found at the site of transgene insertions (Mark et al., 1992; Wilkie and Palmiter, 1987), suggesting that these sequences may be preferred sites for the integration of foreign DNA (Kato et al., 1986). Repeat-containing sites may also be hotspots for mitotic recombination (Wilkie et al., 1991). Finally, instability of hypervariable minisatellite repetitive sequences have been associated with several human diseases, including carcinomas, leukemias, and insulin-dependent diabetes (Krontiris 1995). Furthermore, these sequences bind transcription factors and activate transcription (Krontiris 1995).

In this paper, we describe a novel class of expressed repetitive sequences in the mouse genome whose expression is restricted during development and is induced in response to retinoic acid (RA) in vitro. These sequences, which we have termed Dr repeat sequences, were identified during gene trapping experiments while characterizing fusion transcripts between promoter-less lacZ transgenes, and unknown but transcriptionally active endogenous genes (Forrester et al., 1996). The Dr element and its flanking sequences contain palindromes, and direct and inverted repeats potentially capable of forming secondary structures in RNA and foldback structures in double stranded DNA.

MATERIALS AND METHODS
Gene trapped embryonic stem (ES) cell lines and tissue culture. Construction, handling and characterization of the gene trapped mouse ES cell lines from the wild type R1 cells (Nagy et al., 1993b) used in this study have been described (Forrester et al., 1996).

5' RACE, PCR DNA amplification and cloning. 5' RACE-PCR on fusion transcripts from the gene trapped ES cell lines was performed using a 5' RACE kit (Life Technologies), essentially according to the manufacturer's instructions, with the following modifications: Reverse transcription was carried out at 42°C for 30 min in the presence of [α-32P]dCTP to monitor the synthesis of the first strand cDNA. SuperScript II (Life Technologies) and GTlacZ-1 [5'-GCAAGGCGATTAAGTTGGGT-3'] primer were used for reverse transcription. DNA amplification of the RACE products was performed in two rounds in 50-μl volumes containing 1X PCR Buffer II (Perkin Elmer), 1.25 mM MgCl₂, 200 mM deoxynucleoside triphosphates, 200 nM each primer, and 2.0 U of AmpliTaq DNA Polymerase (Perkin Elmer) in a DNA Thermal Cycler (Perkin Elmer) (1 cycle of 94°C for 5 min; 1 cycle of 80°C for 8 min during which Taq polymerase was added to the reaction mix; 35 cycles of 94°C for 1 min, 60°C for 2 min, and 72°C for 3 min; 1 cycle of 72°C for 5 min). The second round of amplification was performed under the same conditions using nested primers (see below) and 5 μl of the reaction product from the first round. The primers used for DNA amplification were as follows: for the first round, GTlacZ-2A [5'-CCGTCGACTCTGGCGCCGCTCTGTCAG-3'] and Anchor [5'-GGCCACGCCTCGACTAGTACGGGGGIGGGGIGGG-3'] (Life Technologies); for the second round, Nested2AU [5'-CAUAUAUAUAATTGTCGACCTGT TGGTCTGAAACTCAGCCT-3'], and Universal Amplification Primer [5'-CUACUACUACUAGGCCACGCCTCGACTAGTAC-3'] (Life Technologies).
The RACE-PCR products were digested with *Sal*I and cloned into pBlueScript II vector (Stratagene). Alternatively, they were cloned into the pAMP1 plasmid using CloneAmp System (Life Technologies). The cDNA inserts were sequenced on both strands using the AutoRead Sequencing kit (Pharmacia) on an A.L.F. DNA Sequencer (Pharmacia).

**cDNA cloning and sequencing.** A mouse embryonic cDNA library was screened by standard procedures (Sambrook *et al.*, 1989) using Dr-1a as a DNA probe. Hybridization was done in 50% formamide-5X SSPE (1X SSPE is 150 mM NaCl, 10 mM NaH2PO4 and 1 mM EDTA)-5X Denhardt's solution-100 mg/ml, boiled and chilled salmon sperm DNA-0.5% SDS at 42°C for 24 hours. High-stringency washes were done twice in 0.2X SSC, 0.2% SDS at 60°C for 15 min each. DNA was prepared from two positive plaques (Dr-2 and Dr-3) and their cDNA inserts were sequenced as described above. GenBank search and sequence analyses were performed using Fasta, Bestfit, Pretty and Pileup softwares in the Wisconsin Package, Version 8.0-Open VMS, September 1994 (Genetics Computer Group, Inc., Madison, Wisconsin).

**Retinoic acid induction, RNA preparation and Northern blot analyses.** In the retinoic acid induction experiments, ES cells were maintained in ES cell medium (without leukemia inhibitory factor - LIF) containing 5% fetal calf serum (FCS) and 10⁻⁶ M all-trans retinoic acid (RA) (Sigma). The medium was changed every 12 hours on all plates so that cells were in the continuous presence of RA for varying time periods (0, 12, 24 or 48 hours). Control samples (0 hours) were in ES cell medium containing 5% FCS, in the absence of LIF and RA throughout the experiment.

Total RNA was prepared from ES cell lines using a modified version of the method described by Chomczynski and Sacchi (Chomczynski and Sacchi, 1987). Confluent plates (60 mm) of cells were lysed with 5 ml guanidinium thiocyanate (4.4 M) solution containing 25 mM sodium citrate (pH 7.0), 0.6% sarcosyl and 100 mM β-
mercaptoethanol. The lysate was mixed with 5 ml TE-saturated phenol, 1 ml chloroform and 0.5 ml 2 M sodium acetate. After centrifugation, RNA was precipitated from the aqueous phase with an equal volume of isopropanol. Northern blot analyses were carried out according to standard procedures (Sambrook et al., 1989). The blots were hybridized overnight with $^{32}$P-labelled probes, and were washed twice under stringent conditions in 0.2X SSC (1X SSC is 150 mM NaCl plus 15 mM sodium citrate)-0.2% sodium dodecyle sulphate (SDS) at 60°C for 15 min each.

RNA in situ hybridization. RNA in situ hybridization was performed as described (Hogan et al., 1994). Briefly, tissues were cryostat-sectioned at 10 μm, mounted on glass slides and refixed in 4% paraformaldehyde. Prehybridization treatments were performed as described (Hogan et al., 1994). $^{35}$S-labeled single stranded RNA probes were prepared using T3 and T7 RNA polymerases (Boehringer Mannheim). Adjacent sections were hybridized with Dr-2 sense and antisense probes. Post hybridization washings included treatment with 50 μg/ml RNase A (Sigma) at 37°C for 30 min. Following dehydration, the slides were dipped into NTB-2 film emulsion (Kodak), exposed at 4°C for 6-10 days, developed, and stained with toluidine blue.

Genomic screening and Southern blot analysis. A 129/Sv mouse genomic library on 20 large LB plates was screened (1X10⁶ plaques, representing four times the haploid genome) using a Dr-la probe. For Southern blot, genomic DNAs from two mouse species, C57BL/6 and Mus spretus (SPRET/Ei, Jackson Laboratory) were prepared and digested overnight with either SacI, TaqI or XbaI restriction enzymes. Digested DNA (10μg per lane) was fractionated on a 0.9% agarose gel and transferred to GeneScreen Plus nylon membrane (Dupont) and probed using standard procedures. Hybridization and high-stringency washing conditions for genomic screening and Southern blots were as described above for cDNA cloning.
**FISH.** Mouse chromosomes were prepared for fluorescence *in situ* hybridization (FISH) according to the previously published procedure (Feng *et al.*, 1994). Briefly, lymphocytes were isolated from mouse spleen, cultured and then treated with 0.18 mg/ml bromodeoxyuridine for 14 hours. The synchronized cells were washed and recultured at 37°C for 4 hours. Chromosome slides were prepared as described previously (Heng *et al.*, 1992). The Dr-3 cDNA fragment in p-Bluescript (Stratagene) was used as probe and was biotinylated with dATP (15°C, 1 hr) using the BioNick labeling kit (Life Technologies). Hybridization was carried out overnight at 37°C. The slides were washed and FISH signals were detected and amplified as previously described (Heng *et al.*, 1994). Mouse chromosomes were stained with diamidino phenylindole (DAPI). A Leitz-Aristoplan epifluorescent microscope with DAPI and fluorescein isothiocyanate (FITC) filters was used for photographing the slides.

**RESULTS**

The experiments described in this report grew out of a study designed to identify genes responsive to exogenous RA, by a novel gene trap approach in totipotent mouse ES cells (Forrester *et al.*, 1996). In this approach, we introduced a promoterless *lacZ* construct randomly into the mouse genome by electroporation in ES cells. Because the construct contains a splice acceptor site 5’ of *lacZ*, a fusion transcript can be generated that contains upstream exons of the trapped gene and the *lacZ* sequence. The *lacZ* can be used as a marker to monitor the expression patterns of the trapped gene in cell culture and in the embryo. We were therefore able to screen the trapped ES cells for clones in which the *lacZ* reporter was up- or down-regulated in response to RA. The *lacZ* transgene was also used to clone the trapped gene. Using this approach, we isolated a number of ES cell
clones which contain the gene trap lacZ construct integrated within an RA-responsive transcriptional unit (Forrester et al., 1996).

5' RACE-PCR cloning of fusion Dr-1 transcripts from three independent gene-trapped ES cell lines induced with retinoic acid. To clone the trapped genes, we employed a 5' RACE-PCR procedure to amplify fusion transcripts from several independent gene-trapped ES cell clones. The cDNAs obtained were subcloned and sequenced. We repeated the RACE-PCR procedure on three clones, and repeated the subcloning and sequencing of the RACE products. The cDNAs obtained were identical to the ones isolated previously for each line. Analysis of cDNA sequences indicated proper splicing between an exon from an endogenous gene and the splice acceptor site of the lacZ reporter gene in all cases. So far we have RACE products from three RA induced and four RA repressed independent clones. The fusion transcripts from the three induced clones contain at their 5' ends a common DNA sequence of 114-117 bp in addition to the lacZ reporter and unique sequences. The RACE products from the four repressed clones on the other hand do not contain this common DNA sequence (data not shown). This finding prompted us to study this common sequence further. We name these sequences the Dr repeat family, as their expression turned out to be developmentally restricted and they are repetitive in the mouse genome. The three family members, Dr-1a (117 bp), Dr-1b (115 bp) and Dr-1c (114 bp) were derived from the RA-responsive gene trapped ES cell clones I193, I163 and I214, respectively. These three sequences are 92-97% identical (Fig. 2.1). We described previously that the LacZ expression pattern of the three ES cell clones, analyzed either in chimeric or normal embryos, are distinct during embryogenesis (Forrester et al., 1996). In I193-derived embryos, LacZ is expressed in limb buds, branchial arches, neural crest cells, and yolk sac endoderm. In I163-derived embryos, lacZ is expressed in the posterior spinal cord, adjacent mesoderm, and yolk sac.
mesoderm. In I214-derived embryos, \textit{lacZ} is expressed in the most rostral somites and two lateral stripes along the spinal cord.

**Molecular cloning and nucleotide sequence analysis of Dr-2 and Dr-3 cDNAs.** To obtain longer cDNAs containing Dr elements, an 8.5 day post coitum (dpc) total embryonic cDNA library was screened using Dr-1a as a probe. Two positive clones of 0.4 and 1.1 Kb were isolated, subcloned and sequenced (Dr-2 and Dr-3, respectively; Fig. 2.1). Dr-2 and Dr-3 are 95% similar to Dr-1a and are themselves 95% similar throughout the entire length of the Dr-2 sequence (359 bp). The common Dr element and its flanking regions in the Dr-2 and Dr-3 cDNAs have interesting sequence features, including long and short palindromes, and direct and inverted repeats. These regions have the potential to form stem-loop and other foldback structures, and the inverted repeat regions are reminiscent of some transcription factor binding sites (Fig. 2.2). The 54-base long palindrome close to the 3' end of the Dr-3 sequence is capable of forming a large stem loop structure. There is also a polyadenylation signal and a poly A track, suggesting a retrotransposition event in the history of these sequences, close to the 3' ends of the Dr-2 and the Dr-3 sequences (underlined in the case of Dr-3 in Fig. 2.2). In Dr-3, the long palindromic sequence is immediately adjacent to the poly A track. All Dr sequences obtained lack any significant open reading frame.
Figure 2.1. Alignment of the Dr cDNA sequences using the GCG Pretty software program. Dr-1a, Dr-1b, and Dr-1c are from the I193, I163, and I214 gene-trapped ES cell lines respectively. Dr-2 and Dr-3 were cloned from 8.5 dpc embryonic cDNA library (see Materials and Methods). Deletions (-) are indicated based on the best alignment. The numbering is based on the Dr-3 sequence.
**Figure 2.2.** Dr-3 nucleotide sequence. Palindromes (thick arrows), direct repeats (thin arrows), and inverted repeats (two-headed arrows) are indicated. Some of the small repeat regions are not indicated for clarity. The presence of long and short palindromes is indicative of the region's potential to form foldback structures. The polyadenylation signal and the poly A track are underlined.
In a GenBank search, three other mouse sequences were found to contain Dr elements. Long mosaic repeated sequence (LMRS, Accession # X55036), KIF2 (Accession # D12644) and Segment 4 of the transgenic mouse Myk-103 (Accession # M16214) contain Dr elements that are respectively, 96%, 89%, and 98% similar to the Dr-1a sequence. No other sequences in the GenBank from other organisms showed such a degree of similarity to the Dr element. Best matches with other sequences in mouse or sequences from other organisms were in the 60-70% range. Such sequences may be more distantly related to the Dr family. The nucleotide sequence data reported here have been submitted to GenBank and have been assigned the accession numbers U51725 (Dr-1), U51726 (Dr-2), and U51727 (Dr-3).

Dr-containing transcripts are induced by retinoic acid in ES cells. Because the expression of the lacZ marker in the ES cell lines that were used to clone the Dr sequences was induced in response to RA, we examined the RA responsiveness of Dr-containing transcripts (Dr transcripts) by exposing wild type R1 and the gene-trapped I193 ES cell lines to RA in the absence of LIF for up to 48 hours (see Methods). Northern blot analyses of cytoplasmic RNA from these lines revealed a size range of Dr transcripts (0.5-18 Kb) whose expression did not alter in the presence of RA after 12 hours, but was induced after 24 hours and was further overexpressed after 48 hours in RA (Figs. 2.3A & 2.3B). Induction of the Dr-lacZ fusion transcript in the I193 line also confirmed its RA responsivity (Fig. 2.3B).

Developmental expression of Dr repeats. We next examined the temporal and tissue-specific expression pattern of the Dr element in embryos and adult tissues. RNA transcripts containing Dr sequences were observed at low levels in 8.5 dpc embryos and their expression levels increased by midgestation (Fig. 2.3C). These transcripts had a size range similar to those observed in ES cells. All adult tissues examined (testes, brain and
spleen) also express Dr transcripts (Fig. 2.3C). To define this expression pattern further, we performed RNA in situ hybridization analysis on day 16 embryonic sections as well as on adult brain (Fig. 2.4). Dr transcripts are widely expressed in day 16 embryos but the highest expression is observed in the developing brain, specifically in the cortical areas. In the adult brain, expression is widespread in the cerebral cortex with the strongest expression in the hippocampus. Because of the high sequence homology between the Dr family members, our probe was not specific to any particular member of this repeat family. Thus, it is likely that we are detecting multiple family members that might be differentially regulated in different cell lineages. Despite this lack of specificity of the probe used for this RNA in situ analysis, it is interesting to note the developmentally restricted pattern of expression.
Figure 2.3. RA induction of Dr expression in ES cells, and Dr expression in embryos and adult tissues. Northern blot analyses were performed on ten micrograms of total cytoplasmic RNA using radiolabelled Dr-1a probe and then a mouse actin probe. (A) Wild type R1 ES cells, or (B) I193 gene-trapped cell line were treated with RA for 0 to 48 hours. In B, the same blot was reprobed with lacZ indicating the induction of its 4.5 Kb fusion transcript. (C) RNA from 8.5 to 15.5 days post coitum embryos, and adult testes, brain and spleen tissues. The migrations of ribosomal RNAs are indicated as size markers.
Figure 2.4. Developmental expression of Dr transcripts. RNA *in situ* hybridization was performed in 16.0 days post coitum mouse embryo, sagittal sections (A-F), and in adult mouse brain, horizontal sections (G-J) using radiolabelled Dr-2 sense (B & E) and antisense (C, F, H & J) RNA probes. Abbreviations: cb, cerebrum; cc, cerebral cortex; cp, cortical plate; fc, frontal cortex; and hp, hippocampus.
Genomic representation and organization of Dr sequences. To gain information about the genomic representation of the Dr elements, we screened a mouse genomic library representing a fourfold equivalent of the haploid genome (10^6 PFU). A significant number (0.12%) of the total clones screened contained Dr sequences, suggesting a rough estimate of about 300 copies of Dr family members in the mouse haploid genome. We then performed Southern blot analysis to examine in two related but distinct mouse species the degree of conservation of the Dr loci (Fig. 2.5). At this level of analysis, it was evident that there were both similarities and multiple differences between laboratory mice and Mus spretus, indicating a high degree of polymorphism in the Dr loci. Similar polymorphisms have previously been reported for several middle-repetitive DNA sequences in two non-interbreeding strains of Drosophila (Young 1979).

The genomic organization of repetitive sequences range from tandemly repeated to widely dispersed across the genome. The Southern blot and genomic library analyses described above both suggested that Dr sequences were widely dispersed throughout the genome. To test this further, we performed FISH analysis. This analysis revealed that Dr sequences are widely dispersed across the whole genome on all the chromosomes (Fig. 2.5). A few of the signals were quite strong (Fig. 2.5, arrows) suggesting that some Dr sequences may be clustered, a conclusion supported by the presence of several strong bands of hybridization in the Southern blot analysis. The number of positive plaques from the genomic screen and the number of signals from the FISH analysis were roughly in correspondence, giving a minimum estimate for the number of Dr repeats in the mouse genome of 300. However, this number may be an underestimate of the Dr family size since the strong signals in the FISH analysis may be the result of multiple copies of Dr elements repeated in tandem. Furthermore, stringent conditions were used in these analyses, suggesting that less highly related members of the Dr family might have been missed in these experiments.
Figure 2.5. (A) Genomic representation of Dr sequences in two mouse species. Southern blot analysis of C57BL/6 (lanes 1) and Mus spretus (lanes 2) genomic DNA (10 μg) digested with three different restriction endonucleases and probed with radiolabelled Dr-la cDNA. One Kb ladder (GibcoBRL) was used as size marker. (B-C) Localization of Dr sequences in two sets of mouse metaphase chromosome spreads by FISH. (B) Metaphase chromosomes stained with DAPI. (C) PI-stained same chromosomes with FITC signals. Strong signals indicative of Dr clustering are marked by arrows.
DISCUSSION

We have identified a novel class of expressed repetitive sequences in the mouse genome, the Dr repeat family. We isolated these sequences from three independent gene trapped ES cell lines (I193, I163, and I214) responsive to RA. These lines were isolated based on the insertion of a lacZ reporter into a gene that is upregulated in response to RA. The three lines have distinct lacZ expression patterns during embryogenesis, indicating the gene trapped in each line, and therefore the insertion locus, is also distinct (Forrester et al., 1996).

The number of Dr elements per haploid mouse genome is of the order of 300 copies. This number is a rough estimate based on hybridization data from library screening and FISH analysis. Dr sequences contain palindromes and direct and inverted repeats. The repeat unit is at least 100 bp, and they are widely dispersed across the genome. The Dr element can thus be considered as a low copy number SINE. The presence of a polyadenylation signal and a poly A track in the Dr-2 and the Dr-3 sequences is indicative of a retrotransposition event in the history of these sequences. The Dr-3 sequence also includes a 54-base long palindromic sequence close to its 3' end which is capable of forming a stem loop structure in both DNA and RNA. Such palindromic sequences have been found in most eukaryotic genomes and shown to be capable of forming snap-back DNA structures (Hardman 1986). It is been suggested that these sequences are mobile genetic elements (Perlman et al., 1976).

The three Dr-containing sequences identified in a GenBank search provide some insights into the possible roles of these elements. The first Dr sequence (KIF2, # D12644) is in the 5' untranslated region of a kinesin-related gene (Aizawa et al., 1992), raising the possibility that Dr elements may be included in RNA transcripts of functional genes where the primary sequence or the secondary structure of the Dr sequence may affect RNA stability or translation. Inclusion of other repetitive sequences in the RNA

46
transcripts of functional genes has already been reported. For example, there is a B2 repeat in the 3' untranslated region of the mouse muscle γ-phosphorylase kinase gene (Maichele et al., 1993), and a short interspersed repetitive element in the 3' untranslated region of four mammalian genes form the polyadenylation signal (Murnane and Morales, 1995).

The second Dr sequence from GenBank (# M16214) is in a DNA fragment displaced into the integration site of the MyK-103 transgenic mouse (Wilkie and Palmiter, 1987). It is noteworthy that the Dr elements described here were also identified on the basis of an insertion event, by a gene trap construct. These results raise the interesting possibility that Dr elements might be preferred sites of transgenic insertions in the mouse genome.

The third Dr sequence (# X55036) is part of a 15 Kb long mosaic repetitive sequence (Decoville et al., 1992), suggesting that Dr elements may function as part of a long and complex repetitive sequence. The length and the structural features of the Dr-3 sequence, such as palindromic and inverted sequences, also support this notion. The high degree of similarity between different Dr sequences in the mouse genome and their absence in other organisms, indicated by GenBank search, suggests a role for these sequences in the mouse.

There are three possible ways in which Dr elements may have been introduced into the 5' end of the lacZ fusion transcripts that we cloned. First, they may have existed downstream of the promoter within the 5' region of the genes where the lacZ transgene inserted. In this case, the presence of the Dr element may render the locus favorable for the integration of foreign DNA. It has been proposed that Alu repeats increase susceptibility to gene rearrangements in humans (Rudiger et al., 1995). Other repeats have also been found at the sites of integration of foreign DNA (Kato et al., 1986). If such preferred loci for integration do exist, gene trap insertions may not be as random as it is generally thought.
Second, at the time of insertion, Dr elements may have cointegrated with the transgene. There are precedents for the integration of repetitive sequences during gene rearrangement. In characterizing the insertional mutation locus in line 4 transgenic mice, Mark et al. found insertion of DNA fragments similar to the rat LINE-1 repeat family at the junction between host and foreign DNA (Mark et al., 1992). As noted above, Wilkie and Palmiter also identified repetitive DNA fragments at the genomic integration site of the MyK-103 transgenic mouse (Wilkie and Palmiter, 1987).

Finally, Dr and lacZ transcripts could have been brought together in our gene trapped lines by trans-splicing. Trans-splicing has been well characterized in trypanosomes (Agabian 1990; Matthews et al., 1994), and also occurs in organisms such as C. elegans (Blumenthal and Thomas, 1988) and A. lumbricoides (Nilsen 1989). Trans-splicing may also occur in mammalian cells (Eul et al., 1995; Vellard et al., 1992).

At present, there is no evidence that would favor any of the above three models in explaining the apparent involvement of Dr family members in the RA responsive gene trapped lines. However, the presence of Dr elements within the lacZ transcripts in three RA induced lines, and the activation of multiple Dr-containing transcripts in response to RA, suggest the involvement of Dr sequences in an ES cell differentiation program that is triggered by RA. Evidence is now accumulating that repeat-containing transcripts may indeed be involved in regulatory pathways. Mouse B2-containing (Alu type 2) transcripts are strongly induced by mitogens and transforming agents (Edwards et al., 1985; Lania et al., 1987; Singh et al., 1985), and their expression is regulated by RA during the differentiation of F9 embryonic carcinoma (Murphy et al., 1983) and PC13 cells (Bennett et al., 1984). Transcripts from gamma satellite DNA repeats are differentially expressed during mouse development and are strongly repressed by RA in P19 cells (Rudert et al., 1995). Furthermore, the consensus sequence of a major Alu subfamily contains a functional RA response element, raising the possibility that thousands of such sites in the
genome may be potential targets for RA regulatory pathways (Vansant and Reynolds, 1995).

The experiments reported here, as well as those described previously (Forrester et al., 1996) do not address whether the RA-responsive genes, and the Dr repeats contained within them, are directly responsive to RA or are further downstream in a RA response pathway. In these experiments, the kinetics of induction by RA is over a 24-48 hour period. Such long kinetics of transcription activation by RA is not unusual (Simeone et al., 1990; Tini et al., 1993). The HOXB1 gene, for example, has an RA response element in its promoter and is activated by RA after 24 hours (Ogura and Evans, 1995).

In summary, this paper describes the isolation of novel fusion transcripts containing highly similar repetitive Dr elements that respond to RA. Although the physiological significance of Dr repeats remains to be elucidated, our results suggest they may be involved in developmental and inducible patterns of gene expression.
Chapter 3: *Aquarius*, a novel gene isolated by gene trapping with a RNA-dependent RNA polymerase motif

This chapter is a version of the following publication:

ABSTRACT

In a retinoic acid gene trap screen of mouse embryonic stem cells, a novel gene, named *Aquarius* (*Aqr*), was identified and characterized. A promoterless *lacZ* marker was used to disrupt and trap the genomic locus and to determine the expression pattern of the gene. *Aqr* transcripts are strongly induced in response to RA *in vitro*. During embryogenesis, *Aqr* is expressed in mesoderm, neural crest cells and their target tissues, as well as in neuroepithelium. Expression was first detected at 8.5 dpc when neural crest cells are visible at the lateral ridges of the neural plate. The gene trapped *Aqr* locus was transmitted through the mouse germ line in three genetic backgrounds. In the F2 generation, the expected mendelian ratio of 1:2:1 was observed in all backgrounds indicating that homozygous mice are viable. Homozygotes are normal in size and weight, and breed normally. The *Aqr* ORF has weak similarity to RNA-dependent RNA polymerases (RRP) of the murine hepatitis viruses and contains a RRP motif. *Aqr* was mapped to mouse chromosome 2, between regions E5 to F2 by FISH analysis.

INTRODUCTION

Over the next decade, the complete DNA sequence of a mammal will become available. This information will serve as a resource with which to decipher the genetic instructions required for embryological development, organogenesis, and the functioning of various cell types in the adult. The challenge for biology will be to characterize functionally all the genes involved in these processes and in disease. Over the past few years, various
strategies have been developed and refined to clone the genes corresponding to existing mutant or disease phenotypes, or to isolate genes that are the mammalian homologs of genes in other organisms that are more amenable to genetic analysis. These approaches can also be coupled with gene targeting in embryo-derived stem (ES) cells to generate mice with mutations in any desired gene. However, these procedures are expensive, slow and labor intensive, and hence are not yet sufficiently robust to apply on a genome-wide level.

Insertional mutagenesis, based on a stochastic insertion of exogenous DNA into the genome of ES cells, provides another approach to generate novel mutations in the mouse germ line. Furthermore, if the insertional construct includes features such as a promoter-less lacZ cassette, then the inserted DNA provides an easy readout of the normal expression pattern of the 'trapped' gene, and partial sequence information of flanking exons can be derived by various PCR-based amplification strategies. This process of gene trapping offers a feasible and potentially robust strategy to mutagenize and characterize the entire mouse genome.

Given the very large size of the mammalian genome and large number of genes, pseudogenes and non-coding regions, many of the ES clones isolated by a totally non-selective gene trapping strategy will harbor insertions in regions of the genome that are never expressed. Furthermore, a large portion of the remaining insertional events are likely to be in genes that may not be of immediate interest. For example, in a recent test for feasibility of large-scale gene trap mutagenesis, Wurst et al. recovered 279 gene-trapped clones in a screen for developmentally regulated genes (Wurst et al., 1995). Thirty-six clones (13%) exhibited restricted patterns of gene expression in embryonic and extraembryonic tissues, 88 (32%) showed widespread expression and 155 (55%) failed to show detectable levels of expression at 8.5 days of gestation.

Two approaches have been described that are designed to enrich for particular subsets of trapping events in vitro, prior to the generation of mice. Skarnes et al (1995)
developed a gene trapping construct that was designed to trap genes that encode secreted and membrane-spanning proteins, such as receptor tyrosine kinases, phosphatases, cadherin and laminin. The screening strategy was based on an inframe insertion of the CD4 transmembrane domain with the βgeo (β-galactosidase-neomycin fusion) reporter gene as the trapping vector, thereby capturing any insertion occurring downstream of a signal sequence (Skarnes et al., 1995).

In another approach, we have previously reported identification and preliminary characterization of 20 gene trap integrations in ES cells that were screened in vitro for response to exogenous retinoic acid (RA) (Forrester et al., 1996). All but one of these integrations subsequently showed unique and tissue-specific patterns of expression during embryogenesis. Sequence analysis from six integrations revealed five novel genes and one previously identified gene, the protooncogene c-fyn. Germline transmission and breeding uncovered one homozygous embryonic lethal and three homozygous viable insertions. A novel family of RA-responsive repetitive sequences was also identified in this RA screen (Sam et al., 1996).

In other gene trap studies, novel genes with developmentally-restricted expression patterns have been identified and it has been shown that their expression is not disturbed by the insertion of the gene trap vector. Loss-of-function phenotype of homozygous mutant embryos have also been characterized. Takeuchi et al. generated a gene trap mouse mutation, jumonji, that leads to defective neural tube closure and embryonic lethality before day 15.5 of gestation (Takeuchi et al., 1995). The gene is predominantly expressed in the cerebellum and at the midbrain-hindbrain boundary and its deduced amino acid sequence shares a portion of significant homology with human retinoblastoma-binding protein RBP-2.

Chen et al. inactivated the murine Eck receptor tyrosine kinase by a retroviral gene trap insertion and showed that the expression of the Eck promoter was not affected by provirus integration (Chen et al., 1996).
The function of \( \alpha-E\text{-catenin} \) in mouse preimplantation development was also defined by a gene trap approach (Torres et al., 1997). It was shown that a loss-of-function mutation in \( \alpha-E\text{-catenin} \) results in disruption of the trophoblast epithelium and subsequent developmental block at the blastocyst stage.

To better understand B-cell development at the molecular level, the genetic response of B-lineage cells to bacterial lipopolysaccharide (LPS, a potent stimulator of B-cells) was also examined in a gene trap approach. Novel LPS-responsive genes were identified and shown to have restricted expression within the B-lymphoid lineage (Kerr et al., 1996).

The above experiments demonstrate the power of gene trapping to identify and analyze novel genes and their expression and function in the mouse. In this report, we describe the isolation and characterization of the mammalian \textit{Aquarius} gene, identified by the RA induction gene trap strategy described above, and by using Internet-based database access and resources. We believe the combined gene trap and database approach has considerable potential in increasing the pace of gene discovery and functional analyses, particularly as the sequence of the mouse genome comes online.

**MATERIALS AND METHODS**

\textbf{Gene trapped embryonic stem (ES) cell lines, lacZ staining and germ line transmission.} Construction, handling and characterization of the I193 gene trapped ES cell line from wild type R1 cells (Nagy et al., 1993b), lacZ staining and transmission of the trapped locus through the mouse germ line have been described (Forrester et al., 1996).
Mouse genotyping and F2 breeding in different backgrounds. Genomic DNA was isolated from tails at the time of weaning, digested with EcoR1 restriction enzyme and separated by electrophoresis on a 1% agarose gel. Southern blots were double probed with both engrailed-2 (en-2) and lacZ probes. En-2 was used as an internal control for quantitative Southern and lacZ probe detected the trapping vector inserted in the Aqr locus. Densitometric analyses were used to measure the intensity of the lacZ signal (4.0 Kb band) relative to the en-2 signal (11 Kb band). Three F1 males and six F1 females, heterozygous for the disrupted locus, were bred to obtain F2 progeny in a 50% 129sv-cp/50% C57BL6 outbred background. Because the chimeric animals generated by blastocyst injection died shortly after germ line transmission, outbred F1 males were also bred to CD-1 and 129 females to obtain 50% CD-1/25% 129/25% C57BL6, and 75% 129/25% C57BL6 backgrounds. Three males and six females from these backgrounds were also bred to obtain F2 progeny. Heterozygous animals (+/-) would be expected to have a lacZ/en-2 signal ratio of 1, while homozygotes (-/-) would be expected to yield a lacZ/en-2 signal ratio of 2 on Southern blots. To confirm the genotyping, two heterozygotes and two homozygotes were backcrossed to wild type animals and their progeny were genotyped for the presence or absence of the transgene. In all four crosses, half of the progeny from the heterozygotes and all the progeny from the homozygotes carried the transgene, confirming the original genotypes.

RACE, PCR amplification, cloning of the 5' Aqr exon, and genomic PCR. 5' RACE-PCR on fusion transcripts from the gene trapped ES cell lines were performed using a 5' RACE kit (Life Technologies, Burlington, On), essentially according to the manufacturer's instructions, with the following modifications: Reverse transcription was carried out at 42°C for 30 min in the presence of [α-32P]dCTP to monitor the synthesis of the first strand cDNA. SuperScript II (Life Technologies) and GTlacZ-1 [5'-GCAAGGCGATTAAGTTGGT-3'] primer were used for reverse transcription.
DNA amplification of the RACE products was performed in two rounds in 50-μl volumes containing 1X PCR Buffer II (Perkin Elmer, Foster City, Ca), 1.25 mM MgCl₂, 200 mM deoxynucleoside triphosphates, 200 nM of each primer, and 2.0 U of AmpliTaq DNA Polymerase (Perkin Elmer) in a DNA Thermal Cycler (Perkin Elmer) (1 cycle of 94°C for 5 min; 1 cycle of 80°C for 8 min during which Taq polymerase was added to the reaction mix; 35 cycles of 94°C for 1 min, 60°C for 2 min, and 72°C for 3 min; 1 cycle of 72°C for 5 min). The second round of amplification was performed under the same conditions using nested primers (see below) and 5 μl of the reaction product from the first round. The primers used for DNA amplification were as follows: for the first round, GTlacZ-2A [5'-CCGTCGACTCTGGCGCCGCTGCTCTGTCAG-3'] and Anchor [5'-GGCCACGCCTCGACTAGTACGGGGIIGGGIIGG-3'] (Life Technologies); for the second round, Nested2AU [5'-CAUCAUCAUCAUTTGTCGACCTGTGGTCTGAAACTCAGCCT-3'], and Universal Amplification Primer [5'-CUACUACUACUAGGCCACGCGTCGACTAGTAC-3'] (Life Technologies, Burlington, On).

The RACE-PCR products were digested with SalI and cloned into the vector pBlueScript II (Stratagene, La Jolla, Ca). Alternatively, they were cloned into the pAMP1 plasmid using the CloneAmp System (Life Technologies). The cDNA inserts were sequenced on both strands using the AutoRead Sequencing kit (Pharmacia, Baie d'Urfe, Qc) on an A.L.F. DNA Sequencer (Pharmacia).

PCR amplification on genomic DNA from +/-, +/-, and -/- animals was performed with primers flanking the 5' Aqr exon, 1 [5'-GCATGTAAATACTGGGCT-3'], and 2 [5'-CTAAATTCCAGCAGCATT-3'] and the GTlacZ-1 primer (above). PCR conditions were as described above.

cDNA cloning, sequencing, GenBank and internet utilities. A 12.5 days post coitum mouse embryonic cDNA library was screened by standard procedures (Sambrook
et al., 1989) using a non-repeat region of the RACE product (Aqr exon) as a DNA probe. Hybridization was done in 50% formamide, 5X SSPE (1X SSPE is 150 mM NaCl, 10 mM NaH$_2$PO$_4$ and 1 mM EDTA), 5X Denhardt's solution 100 mg/ml, boiled and chilled salmon sperm DNA, 0.5% SDS at 42°C for 24 hours. High-stringency washes were done twice in 0.2X SSC, 0.2% SDS at 60°C for 15 min each. DNA was prepared from the only positive plaque and its 2.4 Kb cDNA insert was sequenced as described above.

GenBank searches and sequence analyses were performed using Fasta, Translate, Bestfit, Gap and Pileup softwares in the Wisconsin Package, Version 8.0-Open VMS, September 1994 (Genetics Computer Group, Inc., Madison, Wisconsin). For the Aqr ORF sequence alignment with the viral RNA-dependent RNA Polymerases, I used the GCG Pileup program with a gap weight and a gap weight length of 2. Only Aqr residues that were in common with at least two of the three other RRP s were considered significant. To evaluate the alignment significance obtained by the Pileup program, Bestfit and Gap programs with 100 randomizations were used in pairwise alignments of Aqr with the three RRP sequences. Bestfit finds the best region of similarity between two sequences while gap generates the best alignment of the two sequences in their entirety. In randomizations, the second sequence is repeatedly shuffled while its length and composition maintained, and then it is realigned to the first sequence. The average alignment quality, plus or minus the standard deviation, of all randomized alignments is reported in the output file (Genhelp Bestfit, Genetics Computer Group, Inc. 1992). z score is a measure of the alignment significance and is calculated as: (actual alignment quality - average random quality)/(standard deviation). z values above 3 are considered possibly significant and above 10 as significant (Lipman and Pearson, 1985).

Aqr open reading frame was submitted to the National Center for Biotechnology Information XREF internet service (http://www.ncbi.nlm.nih.gov/XREFdb/) where it was searched against the most recent update of the EST database. Positive matches were examined for homology with Aqr. Online sequence databases for yeast (100% of the
genome complete, http://speedy.mips.biochem.mpg.de/mips/yeast/), and C. elegans (60% of the genome complete, http://www.sanger.ac.uk/-sjj/C.elegans_Home.html) were also searched for homology with Aqr.

**Retinoic acid induction, RNA preparation, Northern blot analyses, and RT-PCR.** In the retinoic acid induction experiments, ES cells were maintained in ES cell medium (without leukemia inhibitory factor - LIF) containing 5% fetal calf serum (FCS) and 10⁻⁶ M all-trans retinoic acid (RA) (Sigma, St. Louis, Mo). The medium was changed every 12 hours on all plates so that cells were in the continuous presence of RA for varying time periods (0, 12, 24 or 48 hours). Control samples (0 hours) were in ES cell medium containing 5% FCS, in the absence of LIF and RA throughout the experiment. In *in vivo* RA experiments, pregnant mice at E8.5-E9.5 were fed 2 mg of RA per 100 gram of body weight and then embryos were surgically removed 24 or 48 hrs after RA treatment and stained for *lacZ* histochemical staining.

Total RNA was prepared from ES cell lines using a modified version of the method described by Chomczynski and Sacchi (Chomczynski and Sacchi, 1987). Confluent plates (60 mm) of cells were lysed with 5 ml guanidinium thiocyanate (4.4 M) solution containing 25 mM sodium citrate (pH 7.0), 0.6% sarcosyl and 100 mM β-mercaptoethanol. The lysate was mixed with 5 ml TE-saturated phenol, 1 ml chloroform and 0.5 ml 2 M sodium acetate. After centrifugation, RNA was precipitated from the aqueous phase with an equal volume of isopropanol. Northern blot analyses were carried out according to standard procedures (Sambrook *et al.*, 1989). The blots were hybridized overnight with ³²P-labelled probes, and were washed twice under stringent conditions in 0.2X SSC (1X SSC is 150 mM NaCl plus 15 mM sodium citrate)-0.2% sodium dodecyle sulphate (SDS) at 60°C for 15 min each.

RT-PCR was done with the GeneAmp RNA PCR kit (Perkin Elmer, Foster City, Ca) according to the manufacturer's instructions. The primers used for PCR amplification
were the Aqr exon flanking primers A & B, and the Dr-1a repeat flanking primers C & D; A [5'-GAGCTCTTGCTATTGCTC-3'], B [5'-TAGAGTCGGCAGCCCAAT-3'], C [5'-GCATGTAATCTGGCT-3'], and D [5'-CTAAATTCCAGCAGCATT-3']. The PCR products were run on a 1% gel, blotted and hybridized with $^{32}$P-labelled Dr1a-Aqr probe.

**RNA and whole mount in situ hybridization.** RNA in situ hybridization was performed as described (Hogan et al., 1994). Briefly, tissues were cryostat-sectioned at 10 μm, mounted on glass slides and refixed in 4% paraformaldehyde. Prehybridization treatments were performed as described (Hogan et al., 1994). $^{35}$S-labeled single stranded RNA probes were prepared using T3 and T7 RNA polymerases (Boehringer Mannheim, Laval, Qc). Adjacent sections were hybridized with Aqr sense and antisense probes. Post hybridization washes included treatment with 50 μg/ml RNase A (Sigma, St. Louis, Mo) at 37°C for 30 min. Following dehydration, the slides were dipped into NTB-2 film emulsion (Kodak, Rochester, NY), exposed at 4°C for 6-10 days, developed, and stained with toluidine blue. Whole mount in situ hybridization was performed as described previously (Conlon and Herrmann, 1993). Single-stranded Aqr RNA probes were labeled with digoxygenin labeling kit according to the manufacturer (Boehringer Mannheim Biochemicals). After RNA in situ hybridization, embryos were postfixed in 4% paraformaldehyde at 4°C overnight. Sections were cut at 5-6 μm and some sections were counterstained lightly with eosin, and photographed using a Leitz Orthoplan compound microscope and Nomarski optics.

**FISH.** A genomic clone for Aqr was obtained and used as a probe for fluorescence in situ hybridization (FISH) analysis. FISH was performed according to published procedures (Heng et al., 1992; Heng and Tsui, 1993).
RESULTS

The *Aquarius* gene described in this report was identified in a gene trap screen of totipotent mouse ES cells, designed to capture genes responsive to exogenous RA (Forrester *et al.*, 1996). In this approach, the gene trap construct carrying a splice acceptor site 5' of the promoterless *lacZ* reporter gene was randomly integrated into the mouse genome by electroporation. Twenty gene trapped ES clones were isolated that were either induced or repressed after 48 hours of exposure to RA. This paper describes the cloning of the *Aqr* gene from one such clone, I193. This ES clone was of particular interest to us since its *lacZ* was strongly induced in response to RA. Retinoic acid and its derivatives regulate important biological processes, including cell proliferation, differentiation and morphogenesis in a variety of developmental systems (Mangelsdorf *et al.*, 1995).

Cloning of the RA-responsive *Aquarius* gene

To examine RA responsivity, I193 cells were grown in the presence or absence of leukemia inhibitory factor (LIF) which maintains ES cells in an undifferentiated state, and in the absence of LIF and in the presence of RA for 48 hrs, and then stained for *lacZ* activity. As shown in Figure 3.1A-C, the *lacZ* reporter gene was strongly induced in the I193 cell line in response to RA, indicating the RA responsivity of the trapped locus. To determine RA response *in vivo*, we fed pregnant mothers with RA and stained embryos for *lacZ* activity (see Materials and Methods). We did not detect any reproducibly significant change in the extent of staining in the embryos from RA-treated mothers versus those from untreated mothers.

To obtain sequence information about the trapped gene, 5' rapid amplification of cDNA ends-polymerase chain reaction (RACE-PCR) was carried out twice in two
independent reactions on total RNA isolated from RA-treated I193 cells. Both experiments yielded an identical 0.4 Kb cDNA, which was cloned and sequenced. One hundred and fifty bases of the cDNA were derived from the gene trap vector. The 5' half of the RACE product was the Dr-la repeat element that we have previously characterized (Sam et al., 1996). Dr sequences have been submitted to GenBank (Accession numbers: U51725-U51727). The 3' half of the RACE product was novel, with an open reading frame from a 5' exon of a novel gene which we have designated Aquarius (Aqr). The structure of the RACE product is shown in Figure 3.4A.

We used the 5' exon as a probe on Northern blot analysis of RNA isolated at different time points from wild type R1 ES cells treated with RA (Figure 3.1D). Three mRNA isoforms of 2.5, 6.0 and 8.5 Kb were induced after 24 hours of exposure to RA. The 24-hour delay in induction may suggest that Aqr is further downstream of an RA signaling pathway, however, such long kinetics of transcription activation by RA is not unusual (Simeone et al., 1990; Tini et al., 1993).

To confirm the inclusion of the Dr repeat element within Aqr transcripts, we designed two pairs of PCR primers, one from the Dr element and one from the Aqr ORF (Figure 3.1E) and used them in RT-PCR reactions on total RNA from the parental R1 and gene trapped I193 lines (Figure 3.1F). Amplification of the expected 250 bp fragment with A & D primers confirmed the presence of a Dr element on the Aqr transcript in both cell lines. We have previously reported the inclusion of Dr repetitive elements on many RA responsive transcripts (Sam et al., 1996).
Figure 3.1 *LacZ* and *Aqr* expression are induced *in vitro* in response to retinoic acid (RA).

II93 gene trapped ES cells stained for *LacZ* activity, (A) in the presence of LIF and in the absence of RA, (B) in the absence of LIF and the absence of RA, (C) in the absence of LIF and the presence of RA for 48 hrs. (D) RA induction of *Aqr* expression. Wild-type R1 ES cells were treated with RA for 0 to 48 hours. Northern blot analyses were performed on ten micrograms of total cytoplasmic RNA using a radiolabelled *Aqr* probe and then a mouse actin probe. The migrations of ribosomal RNA (1.9 & 4.9 Kb) are indicated as size markers. (E) Diagrammatic representation of the *Aqr* transcript and Dr-la repeat sequence deduced from the RACE-PCR cDNA product. Primers A-D were used in RT-PCR experiment to confirm the presence of Dr-la sequences on the *Aqr* transcript (F).
Aquarius has weak similarity to RNA-dependent RNA polymerases (RRPs)

To determine the sequence of the coding region corresponding to the Aqr ORF, we screened a 12.5 days post coitum embryonic cDNA library using the Aqr 5' exon as a probe. A single 2.4 Kb cDNA was obtained and sequenced. This cDNA contained a novel 552 amino acid open reading frame (ORF) with weak similarity to RNA-dependent RNA polymerases (RRP) of the murine hepatitis and avian bronchitis viruses (GenBank accession: VF1HJH, S15760 and VF1HB2). There is 25% identity between Aqr and the viral RRP genes over the entire deduced amino acid sequence of Aqr (Figure 3.2A). The 2.4 Kb cDNA that we have cloned is similar in size to the 2.5 Kb isoform of the Aqr transcript on Northern blots. However our cDNA is not complete as it does not contain an initiation codon or 5' UTR.

To evaluate the significance of Aqr and RRP sequence alignment, Gap and Bestfit sequence alignments were performed with 100 randomizations to calculate z scores as described in the Materials and Methods (Table 3.1). Since Pileup, Gap and Bestfit programs use different algorithms for aligning sequences, the actual alignments obtained in each case are different even if in some cases they were in the same region (asterisks in Table 3.1). The sequence identities in these alignments were between 22 to 31 percent. In all cases reported in Table 3.1, the sequence similarity is along the whole length of the Aqr ORF although Bestfit finds only the best region of similarity. z scores above 3 are considered possibly significant (Lipman and Pearson, 1985) and is obtained by Bestfit in three alignments (Table 3.1). However, since the z score obtained in the other cases is below 3, the sequence alignment between Aqr and the RRP must be taken with caution and as tentative.

Two sets of values for gapweight and gaplength weight parameters were used for comparison in pairwise alignments. As indicated in Table 3.1, the gapweight and gaplength weight values have a notable effect on z scores, again suggesting caution in evaluating sequence alignments. In such cases of weak sequence similarity with
questionable significance, other biological and biochemical data such as sequence motifs and biochemical assays, or three dimensional structural information about the protein product are highly desirable.

The 25% sequence identity with RRPs may nevertheless be significant because the \( Aqr \) ORF also contains the four segment RRP motif (A-D) corresponding to the polymerase active site (Figure 3.2B) (Poch \textit{et al.}, 1989; Sousa 1996). The RRP motifs within the viral RRPs are in a different region of their sequence from the ones that are shown in Figure 3.2A. Three dimensional structure of RRPs or their active site is not yet determined and therefore the significance of the RRP motif alignment in the figure 3.2B is not clear. Having the active site motif, \( Aqr \) may possess RNA polymerase activity. Alternatively, it may have descended from a common ancestor shared with RRP genes but which later lost its original function.

The \( Aqr \) ORF also contains two putative nuclear localization signals (several lysine and arginine residues) similar to those present in known nuclear proteins such as p53 and DIM (Figure 3.2C), suggesting nuclear localization of the \( Aqr \) gene product (Robbins \textit{et al.}, 1991). Several lysine and arginine residues in a small stretch of amino acids constitute a nuclear localization signal (Robbins \textit{et al.}, 1991). The nucleotide sequence (GenBank: U90333) includes a 5' palindromic region within the ORF capable of forming a 27 bp stem and a 5 base-long loop structure. There is also an RNA instability motif (AUUUA) in the 3' untranslated region (Ross 1996), suggesting a short half-life for the \( Aqr \) transcript.

We searched several online sequence databases to identify other genes with significant homology to \( Aqr \). No significant homology was detected in the yeast or \( C.\ elegans \) databases. However, there were four EST sequences, two from human and two from the mouse EST databases, that were highly similar to the \( Aqr \) ORF. The four ESTs show 88-100% amino acid sequence identity with \( Aqr \) in overlapping regions. Human EST 564231 (GenBank: AA121582) was cloned from NT2 neuronal precursor cells and
is 94% identical to \textit{Aqr} over 167 amino acid residues; human EST 626472 (GenBank: AA188145) was cloned from HeLa cells and is 88% identical to \textit{Aqr} over 92 amino acid residues; mouse EST 765493 (GenBank: AA274735) was cloned from mouse lymph node and is 100% identical to \textit{Aqr} sequence over its full length of 157 amino acids, and mouse EST 535171 (GenBank: AA073476) was cloned from retinoic acid induced P19 embryonic carcinoma cells and is 89% identical to \textit{Aqr} over 123 amino acid residues. This last EST was cloned from an RA induced cell line, suggesting it may also be responsive to RA.

\textbf{Germ line transmission of the gene trapped \textit{Aquarius} locus}

The trapped \textit{Aqr} allele was transmitted through the germ line of mouse chimeras to produce heterozygous I193 mice. Heterozygotes were identified and bred to obtain F2 progeny in three different outbred backgrounds (see Materials and Methods). All genotyping was performed using a quantitative Southern blot analysis (Figure 3.3A & 3.3B). The F2 progeny results generally indicate a 1:2:1 Mendelian distribution of +/-, +/- and +/- animals (Figure 3.3C). The distribution in the 75\% 129 background was slightly skewed but not statistically significant (Chi-Square Goodness of Fit Test, P = 0.09), suggesting that a few homozygous mutant embryos may not have survived to birth. However, when embryos were examined from mid- to late-gestation, no gross abnormalities were observed in +/- embryos. At 10 months of age, +/- and +/- animals were the same weight and size as their wild type (+/+) littermates, bred normally and did not manifest any discernible abnormalities. Newborn skeletons were also prepared and stained with alizarin red and alcian blue stains for skeletal and cartilage structures. No abnormalities were detected in craniofacial bone or cartilage structures where \textit{Aqr} is expressed.
To confirm genomic disruption of the Aqr locus by the gene trap vector, genomic DNA from +/+ , +/- and -/- animals was amplified by PCR using primers from the 5' Aqr exon and the lacZ sequences (Figure 3.3D). We obtained amplified recombinant DNA products from +/- and -/- , but not from +/+, animals, confirming disruption of the Aqr gene by lacZ sequences (Figure 3.3E).

In order to determine if Aqr transcripts are present in the homozygous animals, we performed Northern blot analysis on total RNA from genotyped midgestation embryos (Figure 3.4B). Aqr transcripts were present in the homozygous, as well as heterozygous and wild type embryos, indicating that the disruption of the Aqr locus by the gene trap vector does not lead to a null mutation. The gene trap vector employed in these experiments were designed such that transcription is terminated after the lacZ polyadenylation site, thus leading to a transcription disruption of the endogenous trapped gene. It is possible however that RNA polymerase proceeds with transcription downstream of the polyadenylation site. Polyadenylation site selection by RNA polymerase II and RNA transcript cleavage are not well understood (Alberts et al., 1994; Darnell et al., 1990), and remains a concern in gene trapping experiments. In the case of a run-through transcription, splicing across the gene trap vector sequences could produce a wild type mRNA. The presence of a wild type Aqr transcript may indicate why we did not detect any abnormalities in the homozygous mice.
Figure 3.2. *Aqr* is similar to viral RNA-dependent RNA polymerases (RRPs).

(A) Alignment of the *Aqr* deduced amino acid sequence with the murine hepatitis virus RRP (Vfihjh & S15760), and the avian infectious bronchitis virus RRP (Vfihb2). Gaps (.) are introduced for the best alignment but are not counted in numbering of residues. Numbering is based on the *Aqr* amino acid residues. Amino acid identities are shaded.

(B) Alignment of the RRP conserved motifs for the *Aqr* and the three viral RRP sequences in (A). The strictly conserved residues are in bold and the conserved hydrophobic residues are indicated by a star. Numbers indicate the amino acid residues between the fragments or from the beginning of the sequence. (C) *Aqr* contains two Nuclear Localization Signals (NLSs) aligned here with known NLSs from four other genes. Numbers indicate the residues from the beginning of the sequence.
### A

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aqr</td>
<td>PTVQAL...</td>
<td>DITD...</td>
<td>DITD...</td>
<td>DITD...</td>
<td>DITD...</td>
</tr>
<tr>
<td>Vfihj</td>
<td>FVQD...</td>
<td>DLFH...</td>
<td>DLFH...</td>
<td>DLFH...</td>
<td>DLFH...</td>
</tr>
<tr>
<td>S15760</td>
<td>PTVQAL...</td>
<td>DLFH...</td>
<td>DLFH...</td>
<td>DLFH...</td>
<td>DLFH...</td>
</tr>
<tr>
<td>Vfihb2</td>
<td>PTVQAL...</td>
<td>DLFH...</td>
<td>DLFH...</td>
<td>DLFH...</td>
<td>DLFH...</td>
</tr>
</tbody>
</table>

### B

#### A

<table>
<thead>
<tr>
<th>41</th>
<th>80</th>
<th>116</th>
<th>162</th>
<th>203</th>
<th>240</th>
</tr>
</thead>
<tbody>
<tr>
<td>HIKKSPFDI...</td>
<td>TVKPGN...</td>
<td>TTVKPGN...</td>
<td>TTVKPGN...</td>
<td>TTVKPGN...</td>
<td>TTVKPGN...</td>
</tr>
</tbody>
</table>

#### B

<table>
<thead>
<tr>
<th>43</th>
<th>195</th>
<th>316</th>
<th>519</th>
<th>74</th>
<th>185</th>
</tr>
</thead>
<tbody>
<tr>
<td>KKKspfdiKeleisieKevKsR</td>
<td>KKKspfdiKeleisieKevKsR</td>
<td>KKKspfdiKeleisieKevKsR</td>
<td>KKKspfdiKeleisieKevKsR</td>
<td>KKKspfdiKeleisieKevKsR</td>
<td>KKKspfdiKeleisieKevKsR</td>
</tr>
</tbody>
</table>

### C

<table>
<thead>
<tr>
<th>44</th>
<th>162</th>
<th>203</th>
<th>240</th>
</tr>
</thead>
<tbody>
<tr>
<td>TVKPGN...</td>
<td>TTVKPGN...</td>
<td>TTVKPGN...</td>
<td>TTVKPGN...</td>
</tr>
</tbody>
</table>

### D

<table>
<thead>
<tr>
<th>45</th>
<th>162</th>
<th>203</th>
<th>240</th>
</tr>
</thead>
<tbody>
<tr>
<td>TVKPGN...</td>
<td>TTVKPGN...</td>
<td>TTVKPGN...</td>
<td>TTVKPGN...</td>
</tr>
</tbody>
</table>

---

41 717 895 444 515760 717 895 444
215 729 252 795 444 515760 717 895 444
177 252 729 252 795 444 515760 717 895 444

---

45 162 203 240
215 729 252 795 444 515760 717 895 444
177 252 729 252 795 444 515760 717 895 444

---

43 KKKspfdiKeleisieKevKsR
Aqr 195 KKKspfdiKeleisieKevKsR
p53 316 KKKspfdiKeleisieKevKsR
DIM 519 KKKspfdiKeleisieKevKsR
TGA1A 74 KKKspfdiKeleisieKevKsR
TGA1B 185 KKKspfdiKeleisieKevKsR
Table 3.1. Measurements of Aqr alignment significance by randomization with Gap and Bestfit.

GCG programs, Gap and Bestfit were used with 100 randomizations on pairwise sequence alignments between Aqr and RRP sequences. z scores above 3 are considered as possibly significant and were calculated as described in the Materials and Methods. Two sets of values for gapweight and gaplength weight parameters were used for comparison. Asterisks indicate sequence similarity in the same region as in Figure 3.2.
<table>
<thead>
<tr>
<th>Aqr aligned with</th>
<th>gapweight, gaplength</th>
<th>Gap</th>
<th></th>
<th></th>
<th>Bestfit</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>% identity</td>
<td>z score</td>
<td>% identity</td>
<td>z score</td>
</tr>
<tr>
<td>Vfihjh</td>
<td>1, 4</td>
<td>25.1</td>
<td>0.38*</td>
<td>25.6</td>
<td>4.06</td>
</tr>
<tr>
<td></td>
<td>2, 2</td>
<td>29.8</td>
<td>0.15*</td>
<td>30.4</td>
<td>0.36*</td>
</tr>
<tr>
<td>S15760</td>
<td>1, 4</td>
<td>21.8</td>
<td>1.94</td>
<td>24.9</td>
<td>3.50</td>
</tr>
<tr>
<td></td>
<td>2, 2</td>
<td>29.0</td>
<td>0.95</td>
<td>30.8</td>
<td>0.37</td>
</tr>
<tr>
<td>Vfihb2</td>
<td>1, 4</td>
<td>22.5</td>
<td>2.20*</td>
<td>26.0</td>
<td>3.92</td>
</tr>
<tr>
<td></td>
<td>2, 2</td>
<td>27.8</td>
<td>0.88*</td>
<td>29.7</td>
<td>1.06</td>
</tr>
</tbody>
</table>
Figure 3.3. Genotyping of F2 progeny and the genomic disruption of the Aqr locus.

(A) Southern blot. The upper band is the endogenous en-2 gene used as internal control for DNA loading. The lower two bands contain the lacZ transgene. The ratio of the intensity of the faster lacZ band (peak 3 in B) to the en-2 band (peak 1 in B) indicates the genotype of the animal: +/+ (ratio = 0, lane 5), +/- (ratio = 1, lanes 2 & 4), and -/- (ratio = 2, lanes 1 & 3). (B) The densitometric measurements of the band intensities. (C) Genotypes of F2 progeny in three mixed genetic backgrounds. i) 50% 129sv-cp/50% C57-B16, ii) 25% 129sv-cp/25%C57-B16/50% CD-1, iii) 75% 129sv-cp/25% C57-B16. (D) LacZ disrupts Aqr genomic locus. Aqr exon was cloned by 5' RACE-PCR. The primers 1-3 were used for PCR on genomic DNA to confirm disruption of the Aqr locus by the gene trap vector. (E) PCR on genomic DNA from +/+, +/- and -/- animals and Southern blot probed with Aqr exon.
Figure 3.4 Aqr RACE product and Northern analysis of Aqr embryonic expression.

(A) The structure of the RACE product. Dr-1a: 119 bp Dr repeat element, Aqr: 133 bp Aqr exon, and the LacZ reporter sequences are indicated. (B) Northern blot analysis on total midgestation embryonic RNA from wild type (+/+), heterozygous (+/-) and homozygous (-/-) embryos. The blot was hybridized with the 2.4 Kb Aqr cDNA probe. Three Aqr transcript isoforms are present in the embryos of the three genotypes. Ribosomal RNA are indicated as size markers.
**Embryonic expression in neural crest and mesodermal tissues**

The developmental pattern of *Aqr* gene expression was examined during embryogenesis by staining for β-galactosidase activity (Figure 3.5), and by RNA and whole mount *in situ* hybridization techniques using the *Aqr* cDNA as a probe (Figures 3.6 & 3.7). At day 8.5 of gestation, expression was detected in the cephalic neural folds and mesoderm, paraxial mesoderm, somites, and throughout the lateral ridges of neural folds including the fused regions in between the caudal and rostral extremities of neural tube closures (Figure 3.5A). The lateral ridges of neural folds are where premigratory neural crest cells emerge and subsequently migrate in a rostral-to-caudal sequence at the cranial, trunk and tail levels (Serbedzija *et al.*, 1992; Serbedzija *et al.*, 1990). Trunk neural crest cells migrate through the somites along well-characterized pathways in a segmental fashion and give rise to numerous derivatives including the sensory and sympathetic ganglia (Krull *et al.*, 1995).

At day 10.5, expression is detected in forebrain, posterior midbrain, hindbrain, craniofacial structures, limb buds and internally along the length of the body (Figure 3.5C), indicating that *Aqr* expression is maintained in neural crest target tissues after their migration is terminated. To examine the expression pattern in more detail, histological sections of the *lacZ*-stained embryos at day 11 were prepared (Figures 3.5E-3.5I). Ventricular zone of neuroepithelium throughout the nervous system was positive for *lacZ*. Ventricular expression in neuroepithelium was observed in all staining embryos and was also confirmed by whole mount *in situ* (Figure 3.7B). Figures 3.5E to 3.5G present the expression in the neural tube, telencephalon and diencephalon. Widespread expression was also detected in limb mesoderm (Figure 3.5G), in the mesenchyme within the branchial arches (Figure 3.5H), and in paraxial mesoderm at the lower lumbar region (Figure 3.5I).
To determine whether the gene trap vector insertion deregulated *Aqr* expression, wild type E9.5 embryos were stained by whole mount RNA *in situ* hybridization with *Aqr* cDNA and compared with *lacZ* staining of embryos (Figure 3.6, A & B, D & E). Both *lacZ* and *in situ* hybridization patterns recapitulate the expression patterns for *Aqr* in forebrain, trigeminal and facioacoustic neural crest tissues, olfactory placode, nasal process, branchial arches 1 to 3, limb buds and the lateral ridges of neural folds. These results indicate that the pattern of *lacZ* expression faithfully reproduces the endogenous *Aqr* gene expression pattern. The expression surrounding the otic vesicle and in olfactory placode was more pronounced in the whole mount *in situ* than in the *lacZ*-stained embryos, perhaps due to an unspecific staining rather than a deregulation in gene expression.
Figure 3.5. *LacZ* expression in whole mount embryos and sections.

(A) E8.5: *lacZ* staining in the cephalic neural folds (cnf), cephalic mesoderm (cm), at the lateral ridges of neural folds including the fused regions in between the caudal (cc) and rostral (rc) extremities of neural tube closures, somites (s), and paraxial mesoderm (pm).

E10.5: non-staining control (B) and *lacZ*-expressing (C) littermate embryos. Expression in forebrain, posterior midbrain, hindbrain, craniofacial structures, limb buds and internally along the length of the body. Note the expression in both anterior and posterior regions of the limbs (arrowheads). (D) E13.0: forefoot plate, expression in the circumference of the cartilaginous anlagen of digits. (E-I) Expression in E11 embryonic sections. (E) Ventricular zone in the dorsal region of neuroepithelium (ne) in the neural tube. (F) Ventricular cells of neuroepithelium in telencephalon (tl) and diencephalon (dn).

(G) Widespread expression in limb mesoderm (lm), and in ventricular zone (vz) of neuroepithelium. (H) Widespread expression in mesenchyme within the first branchial arch. (I) Expression in paraxial mesoderm at lower lumbar region.
Figure 3.6. RNA in situ hybridization recapitulates lacZ expression.

RNA in situ hybridization of whole mount embryos (B & E) recapitulates lacZ expression in E9.5 embryos (A & D). (A & B) Expression in forebrain (fb), trigeminal neural crest tissue (tg), region surrounding the otic vesicle (ot), facioacoustic neural crest tissue (fa), olfactory placode (op), nasal process (np), branchial arches 1 to 3 (a1-a3), and limb bud (lb). (D & E) Expression in the lateral ridges of neural folds where premigratory neural crest cells originate (arrows). RNA in situ of embryonic sections, (C) E12.5, sagittal section, expression in neopallial cortex (neo), midbrain (mb), infundibulum of pituitary (ip), Rathke's pouch (rp), cartilage primordium of lumbar vertebral body (cv), and neural-crest derived structures: intrinsic muscle of tongue (imt) and Meckel's cartilage (mc). Expression also in liver (li) and lung (lg). (F) E14.5, sagittal section, expression in neopallial cortex (neo), infundibulum of pituitary (ip), lung (lg), liver (li), cartilage primordium of ribs (cr), thyroid and cricoid cartilages (tc & cc), dorsal root ganglion (drg), metanephros (met), and skin (sk).
Aqr expression was further delineated by examining day 9 and 9.5 embryonic sections from whole mount in situ (Figure 3.7). Aqr RNA was expressed in trigeminal neural crest, branchial arch mesenchyme (Figure 3.7A & 3.7D), laterally migrating neural crest cells and the ventricular region of the neuroepithelium (Figure 3.7B). Furthermore, Aqr was expressed in a dynamic and spatially regulated manner in the limbs, in the Zone of Polarizing Activity (ZPA) at the posterior end of limb bud at day 9.5 (Figure 3.7C). The ZPA constitutes a group of mesenchyme cells that establish the anteroposterior patterning of the limb and express Sonic hedgehog, a key signal involved in limb patterning (Cohn and Tickle, 1996). At day 10.5, Aqr was expressed in both anterior and posterior ends of the limb bud (Figure 3.5C), and at day 13 expression of Aqr in the foot plate was restricted to the circumference of the cartilaginous anlagen of digits (Figure 3.5D). This distinct limb expression suggests a role for Aqr in the determination or patterning of limb structures.

To observe Aqr expression at later stages of embryonic development, we performed RNA in situ hybridization on embryonic sections at days 12.5 (Figure 3.6C) and 14.5 (Figure 3.6F). Aqr was expressed in the neopallial cortex, midbrain, the infundibulum of the pituitary, Rathke's pouch, intrinsic muscle of the tongue, Meckel's cartilage, thyroid and cricoid cartilages, cartilage primordium of ribs and lumbar vertebral body, liver, lung, metanephros, dorsal root ganglion and skin. Some of these structures, such as Meckel's cartilage and ganglions are derived from neural crest cells indicating the persistence of Aqr expression during the differentiation of these cells. Aqr may potentially be a useful marker in following the development of these lineages.

Chromosome localization by FISH analysis

To localize Aqr in the mouse genome, we isolated a phage clone from a genomic mouse library and used it as a probe in FISH analysis (Figure 3.8). The assignment between
signal from the probe and mouse chromosome 2 was obtained by DAPI banding and superimposing FISH signals with DAPI banded chromosomes. The detailed position on chromosome 2, region E5 to F2, was based on the analysis of 10 mitotic chromosome spreads. There are two semidominant mutations, strong's luxoid and tight-skin, in the E5 to F2 region of mouse chromosome 2 that are potentially interesting since they exhibit phenotypes in Aqr-expressing tissues. Strong's luxoid (lst) mice present limb phenotypes including polydactyly, and reductions or duplications of the radius (Lyon et al., 1996). Tight-skin (Tsk) mice have increased growth of cartilage and bone, tight skin and hyperplasia of the subcutaneous connective tissue (Lyon et al., 1996).
Figure 3.7. Expression in sections from whole mount *in situ*.

*Aqr* expression in E9-9.5 embryonic sections. (A) Expression in trigeminal neural crest tissue (arrowheads) just above the rostral extension of dorsal aorta, and in the mesenchyme within the first branchial arch (arrow) around the first branchial arch artery. (B) Expression in the neural crest cells migrating laterally next to surface ectoderm (arrowheads), and in the ventricular zone of neuroepithelium (arrows). (C) Expression in the posterior end of limb bud (arrow). (D) Expression in the mesenchyme within the first branchial arch, stronger expression in the rostral regions.
Figure 3.8. FISH localization of the *Aqr* gene.

Localization of the *Aqr* gene was determined in a set of mouse metaphase chromosome spread by FISH. (a) PI-stained chromosomes with FITC signals. (b) Chromosomes stained with DAPI to assign signal from the probe with the mouse chromosome 2 bands. Under the conditions used, the hybridization efficiency was 93% for the probe (93 out of 100 mitotic figures showed signals on one pair of the chromosomes).
DISCUSSION

We have identified a novel mammalian gene, named Aquarius (Aqr), which may be related to RNA-dependent RNA polymerases (RRPs). Aqr shares 25% amino acid identity over the length of its ORF with murine hepatitis virus RRP s and it has the conserved active site RRP motif. Our sequence analysis and measurement of alignment significance indicates that the RRP sequence similarity by itself may or may not be significant. In addition, very little is known about the structural and functional domains of RRP s and thus the significance of amino acid identities and the presence of the RRP motif is not clear. Following entry into the host cell and uncoating in the cytoplasm, positive-stranded RNA viruses replicate their genetic material by using their (+) RNA genome as a template to synthesize a complementary (-) RNA molecule which in turn serves as a template for the synthesis of progeny genomic (+)-strand RNA. There have been a few reports suggesting the presence of RNA-dependent RNA synthesis activity in eukaryotic cells (Schiebel et al., 1993a; Schiebel et al., 1993b; Volloch et al., 1991). In addition, although hepatitis delta virus does not encode any polymerase, it can still replicate autonomously via an RNA-dependent RNA synthesis mechanism in animal cells, raising the possibility that eukaryotic cells have the enzymatic machinery necessary to replicate RNA molecules (Lai 1995).

The polymerase subunit of the Tetrahymena telomerase, p95, also yields an alignment with the polymerase active site of a family of viral RNA-dependent RNA polymerases (Collins et al., 1995). Telomerase is a specialized polymerase that synthesizes long repetitive telomeric DNA (Shay 1996). The mouse homolog of the p80 subunit of Tetrahymena telomerase has been identified but the mouse p95 homolog, which is thought to carry the catalytic polymerase domain, has not been identified (Harrington et al., 1997). Clearly, it will be of interest to determine whether the Aqr gene
product possesses RNA-dependent RNA polymerase activity. The Aqr ORF also carries two putative nuclear localization signals, suggesting that the Aqr protein is localized to the cell nucleus.

During embryogenesis, Aqr is predominantly expressed in neural crest (NC) tissues from early on in their ontogeny to their differentiated derivatives in cartilaginous head structures such as the hyoid bone and Meckel's cartilage. Expression was first detected in the dorsal edges along the whole length of the neural groove and in its adjacent mesenchyme in the headfold which contains the earliest migrating NC cells. The NC is a transient embryonic cell population which arises from the dorsal region of the neural tube, extensively migrates and gives rise to a number of tissues and structures including ectomesenchyme which forms most of the skeleton, dermis and connective tissue of the vertebrate head (Couly et al., 1993; Schilling 1997). Cranial mesoderm and NC cells are codistributed in the craniofacial mesenchyme but are distinctly segregated in the branchial arches, although the interactions of these cell populations are not yet fully understood (Trainor and Tam, 1995). Aqr expression in NC and craniofacial and branchial mesoderm suggests a possible role for its gene product in craniofacial development. Aqr is also expressed in paraxial and branchial mesoderm and other mesodermal tissues such as the zone of polarizing activity in the limb buds where Sonic hedgehog signaling is thought to establish the anteroposterior limb pattern (Perrimon 1995). There is also evidence that Sonic hedgehog cooperates with a RA inducible cofactor to establish ZPA-like activity (Ogura et al., 1996). The dynamic expression of Aquarius in the limbs and its in vitro RA-inducibility suggests a possible role for Aqr in the development of limb structures or their patterning.

Through germ line transmission of the trapped locus, we created animals homozygous for the disrupted Aqr gene. Homozygous mutant mice did not present any obvious abnormalities. At 10 months of age, they were of the same size and weight as their heterozygous and wild type siblings, and they bred normally. Newborn skeletons
from homozygotes also did not present any abnormalities in craniofacial structures. The absence of any discernible phenotype in mice homozygous for the gene trap allele does not rule out the possibility that Aqr plays an essential role in the cells that express it. The gene trap allele described here may not be a null mutation as Aqr RNA is present in homozygous animals. The lack of a discernible phenotype is likely due to read-through transcription and splicing across the trapping vector.

Aqr was identified in a retinoic acid gene trap screen of mouse ES cells. In vitro, Aqr expression is induced nine-fold in response to RA (Forrester et al., 1996) and its transcript includes an RA-induced Dr repeat element in its 5' end (Sam et al., 1996). We did not detect RA responsivity of the Aqr locus when we examined embryonic lacZ expression in vivo (E9.5-E11). This lack of in vivo RA-responsivity might be due to the inability to expose embryos to a concentration of RA in vivo that corresponds to the optimal dose in vitro (Morriss-Kay 1993).

RA is known to induce both alterations in neural crest cell migration as well as craniofacial defects (Gale et al., 1996; Lee et al., 1995). Retinoic acid receptor α (RARα) is expressed in migrating crest cells and in facial mesenchyme (Ruberte et al., 1991). Also, mutant mice with targeted mutations in both RARα and other RARs have severe defects in the craniofacial complex (Lohnes et al., 1994). The transcription factor AP-2 is induced in response to RA, is expressed in neural crest cell lineages and its disruption in mice also leads to anencephaly and craniofacial defects (Zhang et al., 1996). Aqr induction in response to RA and its expression in ectomesenchyme presents Aqr as a candidate gene in an RA signaling pathway that may also play a role in craniofacial development.

In a search of EST databases, we identified Aqr itself and three other genes similar to Aqr, two from human and one from the mouse. These genes are highly similar in sequence to Aqr (88-94% aa identity). The high degree of similarity suggests that they may be homologous genes in the two species. We did not detect any other similar
sequences in the GenBank, yeast or *C. elegans* databases, suggesting that *Aqr* may be the founding member of a mammalian-specific gene family.

The availability of a significant number of sequence databases for different organisms, the utility of analytical software products and the ready access to these tools through the internet has led to an increase in gene identification by computer (Fickett 1996). Approximately 50% of all human genes are already represented in the Expressed Sequence Tag (EST) databases (http://www.ncbi.nlm.nih.gov/dbEST) (Gerhold and Caskey, 1996). Up to half of ESTs are sequences from the 5' end of cDNAs, and gene trapping experiments can also be designed to trap genes in their 5' exons, as was the case in the present study. Thus, one can envision rapidly obtaining sequence information about the trapped gene by RACE-PCR cloning of 5' exons, searching the EST databases for the 5' exon sequences and obtaining the commercially available EST clones. This combination of gene trap experimentation and EST databases provides ready access to the sequence of trapped genes and has the potential to increase several fold the pace of gene discovery and functional characterization. At the same time, the gene trapped mice allow for expression and functional analyses of the trapped genes.

In summary, in an RA gene trap screen of mouse ES cells, we have identified and characterized the RA-responsive gene *Aquarius*. *Aqr* is developmentally expressed and is a novel mammalian gene with some similarity to RNA-dependent RNA polymerases. Our results demonstrate the feasibility of *in vitro* manipulation and screening of ES cells to rapidly identify and characterize genes that lie along a cellular response pathway. Modification of this gene trap strategy in ES cells should make it possible to trap and study genes that are either induced by other biological, chemical or physical agents or that are subject to developmental regulation.
Chapter 4: The Novel Gene Trapped *mym* Locus Is Required for Midgestation Heart Development

Chapter 4 is in preparation for publication as: Sam M and Bernstein A. The Novel Gene Trapped *mym* Locus Is Required for Midgestation Heart Development.
ABSTRACT

In a gene trap screen of mouse embryonic stem (ES) cells, a novel locus named myocyte maintenance (mym) was identified and characterized. The promoterless lacZ gene was used to disrupt and trap the genomic locus and to determine the expression pattern of the gene. During development, mym is expressed in heart, limbs, gut, meninges and the choroid plexus. Expression was first detected at day 9.5 of gestation in the heart, branchial arches and craniofacial mesoderm. The disrupted mym locus was transmitted through the mouse germ line in three genetic backgrounds. No homozygous animals were born in the F2 generation indicating embryonic lethality of the mutant embryos. Heterozygous males had reduced fertility in a pure genetic background. Homozygous mutant embryos died during midgestation due to heart failure. Myocardial wall and trabeculae were diminished in thickness and size, and were hypoplastic due to apoptotic cell death of myocytes. In mutant myocytes, the organization of actin-myosin filaments and cell junctions were disrupted and the expression of myosin light chain isoforms was deregulated. Our results indicate that mym is required for the maintenance and organization of heart myocytes during midgestation development.

INTRODUCTION

In vertebrates, the first organ to develop and function in the fetus is the heart. Cardiac development involves the coordinated interaction and development of three lineages, myocytes, endothelial and neural crest cells, to establish the main structures of the heart by midgestation. Heart morphogenesis has been studied in greatest detail in chick, where
at stage 4, the anterior lateral plate mesoderm cells are committed to the cardiogenic lineage. The bilateral cardiac primordia then fuse and give rise to the primitive heart tube at stage 10 when the rhythmic heart beating begins. The heart tube then undergoes looping where atria and ventricles subsequently appear (Litvin et al., 1992). A variety of genes have been recently implicated in vertebrate heart development (reviewed in Olson and Srivastava, 1996; and Rossant 1996). A number of these genes are required for the development of myocardium and heart trabeculae, as determined by gene targeting in mouse embryonic stem cells, and by antisense or dominant negative mutant studies. Many of the mutant embryos in these studies die during midgestation due to heart failure at the time yolk sac-based circulation is replaced by a cardiovascular circulation. Here we provide a summary of gene mutations that lead to embryonic lethality as a result of a defect in the myocardium.

N-Myc is a member of the well-characterized myc nuclear protooncogene family, and has transcriptional regulatory function. N-myc is expressed at its highest levels in the expanding primitive streak and embryonic mesoderm, and subsequently in the myocardium of the cardiac ventricles (Downs et al., 1989; Kato et al., 1991). Targeted disruption of N-myc results in embryonic lethality between embryonic day E10.5 and E12.5 with a primary underdeveloped defective heart retaining the S-shape more typical of E9 embryos (Charron et al., 1992; Sawai et al., 1993; Stanton et al., 1992). Mutant hearts contain four chambers but vulvular, septal and trabecular tissues are underdeveloped, suggesting a defect in endothelium-myocardium interaction which normally lead to epithelium-mesenchymal transformation. Compound heterozygous fetuses with a leaky and a null mutation in the N-myc locus also die around E13.5 due to similar heart defects (Moens et al., 1993).

Transcriptional enhancer factor 1 (TEF-1) regulates the cardiac-specific genes, including troponin C, T, I, and the myosin heavy chain genes (Kariya et al., 1993). TEF-1 disruption by gene trapping leads to embryonic lethality between E11 and E12 with an
abnormally thin ventricular wall and a reduced number of trabeculae in the heart (Chen et al., 1994). An enlarged pericardial cavity and a pale yolk sac, indicative of poor blood circulation, are also detected. Transcription of TEF-1-regulated genes in the heart, however, appears normal and heart development is not otherwise compromised in the TEF-1 mutant fetus.

The homeobox gene Nkx2-5, the murine homologue of the Drosophila gene tinman, is the earliest known myogenic transcription factor (Lints TJ et al., 1993). It is expressed in myocardogenic progenitor cells and continues to be expressed in embryonic, fetal and adult cardiomyocytes. It is also expressed in the future pharyngeal endoderm, the tissue with presumptive heart inducing activity. Murine embryos lacking Nkx2-5 exhibit abnormal heart morphogenesis, growth retardation and die during midgestation apparently from hemodynamic insufficiency (Lyons et al., 1995). The process of looping morphogenesis is not initiated in these embryos and trabeculation and endocardial cushion formation are also blocked. The myosin light-chain 2V gene, the earliest known molecular marker of ventricular differentiation, is not expressed in Nkx2-5 mutant hearts, while myosin heavy-chain B and cyclin D2, two other ventricle-specific genes, were expressed normally. Nkx2-5 is thus a component of a genetic pathway required for myogenic specialization of the ventricles.

The Wilms' tumor-associated gene, WT-1, is a zinc finger transcription factor expressed as early as E9 in intermediate mesoderm and subsequently in differentiating mesothelium, spinal cord, brain and urogenital ridge derivatives. Targeted mutation of WT-1 causes thin ventricular walls, frequent pericardial bleeding and systemic edema (Kreidberg et al., 1993). Mutant fetuses die at day 14.5 of gestation likely due to cardiac dysfunction. Similarly, hypoplastic development of the ventricular chambers was observed in homozygous mutant fetuses lacking the ligand-dependent nuclear receptor transcription factor, retinoid X receptor (RXRα), and dying between E13.5 and E16.5
(Sucov et al., 1994). RXRα mutant fetuses also have a poorly developed ventricular septum.

Fibroblast growth factor (FGF) and the FGF receptor (FGFR) are expressed at high levels early in cardiogenesis but their expression later declines as the mitotic activity of myocytes decreases (Consigli and Joseph-Silverstein, 1991; Engelmann et al., 1993; Parlow et al., 1991; Peters et al., 1992). Antisense RNA and dominant negative mutants of FGFR type 1 (FGFR1) inhibit myocyte proliferation or survival during the first week of chicken embryonic development, but had much less effect after the second week, suggesting that FGF signaling may regulate myocyte growth during tubular stages of cardiogenesis (Mima et al., 1995). Neuregulin growth factors and their receptors ErbB2 and ErbB4 are involved in the paracrine mechanism for intercellular communication in embryonic cardiac and neural development (Carraway 1996). Homozygous mutations in neuregulin, erbB2, and erbB4 genes lead to embryonic lethality before E11 due to heart malformations identical to each other (Gassmann et al., 1995; Lee et al., 1995; Meyer and Birchmeier, 1995). In these three gene mutant mice, ventricle morphology is abnormal due to a complete absence of myocardial trabeculae. Trabeculae are the first anatomical feature of ventricle differentiation and are responsible for the maintenance of blood circulation during early stages of heart development before expansion of the compact zone. The endocardial cushion is also hypoplastic in all three homozygous mutant embryos.

The neurofibromatosis type 1 gene, NF1, is a member of the GAP family of Ras regulatory proteins involved in a Ras-dependent signal transduction pathway. NF1 is highly expressed in the fetal heart (Gutmann et al., 1993). A targeted null mutation of mouse NF1 leads to fetal lethality at E13.5 in homozygotes, with generalized edema and abnormal cardiac development where both great vessels, aorta and pulmonary artery, emerge from the right ventricle instead of the aorta exiting normally from the left ventricle (Brannan et al., 1994; Jacks et al., 1994). The myocardium, particularly in the
ventricles, is lacy in appearance and hypoplastic, and there is a significant ventricular septal defect.

Tyrosine hydroxylase is an enzyme in the catecholamine (neurotransmitters and hormones like dopamine and noradrenaline) biosynthetic pathway. The inactivation of both its alleles in the mouse results in lethality between E11.5 and E15.5 due to cardiovascular failure (Zhou et al., 1995). The atria of mutant fetuses at E12.5 are markedly dilated, and their walls are reduced to one or two cell layers. Ventricular cardiomyocytes are less organized, heterogeneous in size and more vacuolated. Targeted mutation in noradrenaline also leads to fetal lethality at E13.5 with a heart phenotype very similar to tyrosine hydroxylase mutant fetuses (Thomas et al., 1995), indicating a requirement for catecholamines for early fetal development and that the heart may be an important target for their action. The absence of striking morphological defects in these mutant fetuses suggest that the primary effect may be physiological such as an inability to sustain sufficient heart rate or contractility.

Plakoglobin is a component of desmosomal plaques and cadherin-catenin cell adhesion complexes in cell junctions. Its inactivation by homologous recombination in mice leads to embryonic death after E10.5 due to severe heart defects including hypoplastic ventricular wall, trabeculae, and endocardial cushion, and bleeding into the pericardial cavity (Bierkamp et al., 1996). Ultrastructurally, the number of desmosomes is greatly reduced and their intercellular cement (desmogleas) is atypical in appearance.

Vascular cell adhesion molecule 1 (VCAM-1) is a type 1 transmembrane protein belonging to the immunoglobulin superfamily. Half of the VCAM-1-deficient embryos die before E11.5 from a severe placental defect, while the other half survive to E12.5 and die from heart failure due to a reduction of the compact layer of the ventricular myocardium and intraventricular septum (Kwee et al., 1995). The mutant hearts also lacked an epicardium. Mutant homozygous embryos lack the cell surface adhesion protein N-Cadherin, and die at E10 with a dramatic cell adhesion defect in myocardial
tissue. Mutant myocardium initially forms but subsequently dissociate and as a result the heart tube fails to develop normally (Radice et al., 1997). Targeted mutation in the zinc finger EviI proto-oncogene also leads to embryonic lethality around E10.5 with abnormal heart development (Hoyt et al., 1997). The mutant hearts are poorly developed with a definite looping defect and a retarded constriction between atria and ventricle.

As summarized above, functional mutations in several transcription factors, growth factors and their receptors, signaling regulatory proteins, metabolic enzymes and structural proteins lead to abnormal midgestation myocardium development and fetal lethality. The complexity of cell lineage interactions and tissue morphogenesis required for cardiogenesis, and the incompatibility of the absence of cardiovascular function with life in midgestation development explains the myocardial defects and the lethality of mutations discovered in such a wide variety of genes.

Using a gene trap strategy in mouse embryonic stem (ES) cells, we have disrupted and identified the myocyte maintenance (mym) locus that is essential for the formation of the myocardial wall and the heart trabeculae. mym mutant embryos have disorganized hearts with hypoplastic myocardium which is diminished in thickness and size. In mutant myocytes, actin-myosin filaments and cell junctions are disrupted and the cells undergo apoptotic cell death.
MATERIALS AND METHODS

Gene trapped embryonic stem (ES) cell lines, lacZ staining and germ line transmission. Construction, handling and characterization of the R140 gene trapped ES cell line, lacZ staining and the transmission of the trapped locus through the mouse germ line have been described (Forrester et al., 1996).

Mouse and embryo genotyping and F2 breeding in different backgrounds. Genomic DNA was isolated from tails or yolk sac, digested with the EcoR1 restriction enzyme and separated by electrophoresis on a 1% agarose gel. Southern blots were double probed with en-2 and lacZ probes. En-2 was used as an internal control for quantitation and the lacZ probe detected the transgene in the mym locus. Densitometric analyses were used to measure the intensity of the lacZ signal relative to the en-2 signal. Three F1 males and six F1 females, heterozygous for the disrupted locus, were bred to obtain F2 progeny in 50% 129sv-cp/50% C57BL/6 outbred background. The germ line chimeric males were also bred to CD-1 and 129 wild type females to obtain 50% CD-1/50% 129, and 100% 129 backgrounds. Three males and six females from these backgrounds were also bred to obtain F2 progeny. Heterozygotes (+/-) had a lacZ/en-2 signal ratio of 1, and homozygotes (-/-) had a lacZ/en-2 signal ratio of 2 on Southern blots. To confirm the genotyping, four heterozygotes were backcrossed to wild type animals and their progeny were genotyped for the presence or absence of the transgene. In all four cases, half of the progeny carried the transgene confirming heterozygosity of the mym locus in all tested animals carrying the transgene.

RACE, PCR amplification, cloning of the 5' mym exon. 5' RACE-PCR on fusion transcripts from the gene trapped ES cell line R140 was performed using a 5' RACE kit (Life Technologies, Burlington, On), according to the manufacturer's
instructions, with the following modifications: Reverse transcription was carried out at 42°C for 30 min in the presence of [α-32P]dTTP to monitor the synthesis of first strand cDNA. SuperScript II (Life Technologies) and GGTαCZ-1 [5'-GCAAGGCGATTAAGTTGGT-3'] primers were used for reverse transcription.

DNA amplification of the RACE products was performed in two rounds in 50-μl volumes containing 1X PCR Buffer II (Perkin Elmer, Foster City, Ca), 1.25 mM MgCl2, 200 mM deoxynucleoside triphosphates, 200 nM of each primer, and 2.0 U of AmpliTaq DNA Polymerase (Perkin Elmer) in a DNA Thermal Cycler (Perkin Elmer). Cycle rotations were as follows, 1 cycle of 94°C for 5 min; 1 cycle of 80°C for 8 min during which Taq polymerase was added to the reaction mix; 35 cycles of 94°C for 1 min, 60°C for 2 min, and 72°C for 3 min; 1 cycle of 72°C for 5 min). The second round of amplification was performed under the same conditions using nested primers (see below) and 5 μl of the reaction product from the first round. The primers used for DNA amplification were as follows: for the first round, GGTαCZ-2A [5'-CCGTCCACTCTGGCAGCCTGCTCTGTCAG-3'] and Anchor [5'-GGCCACGCGTCACTAGTACGGGGGIGGGIGGGIG-3'] (Life Technologies); for the second round, Nested2AU [5'-CAUCAUCAUCAUTTGTCGACCTGTGGTCTGAAACTCAGCCT-3'], and Universal Amplification Primer [5'-CUACUACUACUAGGCCACGCGTCACTAGTAC-3'] (Life Technologies, Burlington, On). The RACE-PCR products were digested with SalI and cloned into pBlueScript II vector (Stratagene, La Jolla, Ca). Alternatively, they were cloned into the pAMPl plasmid using CloneAmp System (Life Technologies). The cDNA inserts were sequenced on both strands using the AutoRead Sequencing kit (Pharmacia, Baie d'Urfe, Qc) on an A.L.F. DNA Sequencer (Pharmacia).

**Genomic PCR, and FISH.** PCR amplification on genomic DNA from +/+ , +/- , and -/- embryos was performed with primers flanking the 5' mym exon, a [5'-
GCTCTTGTCCTAGAAGCC-3', and b [5'-CAAAGGGCAAACCCAAAA-3'] and the GTlacZ-1 primer (above). PCR conditions were as described above. A genomic clone for mym was obtained and used as a probe for fluorescence in situ hybridization (FISH) analysis. FISH was performed according to published procedures (Heng et al., 1992; Heng and Tsui, 1993).

**RNA in situ hybridization.** RNA in situ hybridization was performed as described (Hogan et al., 1994). Briefly, fixed tissues were cryostat-sectioned at 10 μm thickness, mounted on glass slides and refixed in 4% paraformaldehyde. Prehybridization treatments were performed as described (Hogan et al., 1994). 35S-labeled single stranded RNA probes were prepared using T3 and T7 RNA polymerases (Boehringer Mannheim, Laval, Qc). Adjacent sections from wild-type and homozygous mutant hearts were hybridized with different myocyte and endothelial marker probes (obtained from J. Cross and J. Rossant). The probes were α-cardiac actin (Kumar et al., 1997), atrial myosin light chain 1 & 2 (MLC1A & -2A, Kubalak et al., 1994; Lyons et al., 1990), ventricle myosin light chain 2 (MLC2V, Lee et al., 1992), muscle-specific isoform of adenylosuccinate synthetase (Adss1, Lewis et al., 1996), and Flk-1 (Yamaguchi et al., 1993). Post hybridization washings included treatment with 50 μg/ml RNase A (Sigma, St. Louis, Mo) at 37°C for 30 min. Following dehydration, the slides were dipped into NTB-2 film emulsion (Kodak, Rochester, NY), exposed at 4°C for 6 days, developed, and stained with toluidine blue.

**Scanning electron microscopy.** Whole embryos were fixed in 2% glutaraldehyde in 0.1 M sodium cacodylate buffer, post-fixed in 1% osmium tetroxide in 0.1 M sodium cacodylate buffer, dehydrated in a graded ethanol series followed by propylene oxide, and embedded in Spurr epoxy resin. Sections 100 nm thick were cut on an RMC MT6000
ultramicrotome, stained with lead citrate and uranyl acetate, and viewed on a Philips EM430 transmission electron microscope.

RESULTS

The *mym* gene was originally identified in a gene trap screen of totipotent mouse ES cells that was designed to capture genes responsive to exogenous retinoic acid (RA) (Forrester *et al*., 1996). The gene trap construct used in this study carried a splice acceptor site 5' of a promoterless *lacZ* reporter gene randomly integrated into the mouse genome by electroporation. The R140 insertion characterized here is one of 20 gene trapped ES clones that were isolated and characterized further for their patterns of *lacZ* expression during embryonic development.

**R140 gene trap ES cells and RACE-PCR products.** The R140 ES clone was originally thought to be repressed in response to RA (Forrester *et al*., 1996). We examined *lacZ* staining of cells in the presence and absence of RA and leukemia inhibitory factor (LIF), and observed that repression of *lacZ* activity was minimal and most likely due to the absence of LIF in the media, and not the presence of RA (data not shown). Nevertheless, due to the restricted expression pattern of the trapped locus in the embryonic heart, we were interested in characterizing the R140 clone further.

To obtain sequence information about the trapped gene, the 5' rapid amplification of cDNA ends-polymerase chain reaction (RACE-PCR) technique was employed on total cellular RNA isolated from the R140 clone. A 0.4 Kb cDNA was cloned and sequenced. The 3' 200 nucleotides of the cDNA came from the gene trap vector itself. The 5' half was novel, had the beginning of an open reading frame and was used as a probe for cDNA and genomic library screening and used to design primers for further 5' and 3'
RACE-PCR. Further 5' RACE on total cDNA from El2 embryos provided us with an extra 200 bases of sequence for a total of 400 bp (Figure 4.1, GenBank accession: AF019615). The partial cDNA contains initiation codon and the beginning of an open reading frame (ORF), 14 amino acid residues in length. The amino acid stretch is novel and does not match any sequences in the database. To date, we have not obtained the full length mmc cDNA.

**Genomic disruption and chromosome localization by FISH analysis.** To confirm genomic disruption of the mmc locus by the gene trap vector, genomic DNA from +/-, +/- and --/- embryos was amplified using primers from the mmc exon 5' of the lacZ insertion and the lacZ sequences (Figure 4.2C). We obtained amplified recombinant DNA from +/- and --/-, but not from +/- embryos, indicating the disruption of the mmc gene by the lacZ sequences (Figure 4.2D).

To localize the trapped locus in the mouse genome, we isolated a genomic clone and used it as a probe for FISH analysis (Figure 4.2E-2G). The gene was localized on chromosome 5, region B3. A spontaneous recessive lethal mutation (15H1) has also been mapped to this region close to the W locus (Lyon and Glenister, 1982). W19H, a radiation induced allele of W also leads to preimplantation embryonic lethality in homozygotes (Lyon et al., 1984). W19H is a large deletion covering Kit, the Ph locus, and a recessive lethal located 2 cM distal to W. W19H mice may be useful in crosses with mmc mice for complementation and mapping experiments.

**LacZ expression pattern of the mmc trapped locus.** The pattern of mmc expression was examined during embryogenesis by staining for lacZ expression (Figure 4.3). At days 9.5 to 11.5 of gestation (E9.5-11.5), expression was observed in the heart, gut, craniofacial structures, otic vesicle and the apical ectodermal region of the limb buds (Figures 4.3A-E & 4.4). At E12.5-14.5, expression was detected in the leptomeninges (Figure 4.3D), atrium and ventricle of the heart (Figure 4.3F), choroid plexus (Figures
4.3G & 4.4C) and in the undifferentiated mesoderm of the limbs (Figure 4.3H). *mym* was also expressed in the adult hippocampus and the Purkinje cell layer of the cerebellar cortex (Figures 4.3I-J). At E11.5, *lacZ* expression in the heart was localized in the ventricle and atrial myocardium as well as in the trabeculae (Figures 4.4A-B), in the ependymal cells of the choroid invagination (Figure 4.4C), and in the midgut, hindgut and urogenital ridge (Figure 4.4D).
Figure 4.1. Nucleotide sequence of the partial mym cDNA. Partial cDNA sequence was obtained in two rounds of 3' RACE-PCR. Partial open reading frame starting with a methionine codon is translated to amino acid residues towards the 3' end of the sequence. The splice junction where vector sequences began is indicated by an arrow.
**Figure 4.2. Southern genotyping, gene trap insertion in the *mym* locus and FISH analysis.** (A) Southern blot. The upper band, derived from the *en-2* gene, is used as internal control for DNA loading. The lower band is derived from the *lacZ* transgene. The ratio of intensity of the *lacZ* band (peak 2 in B) to the *en-2* band (peak 1 in B) indicates the genotype of the animal: +/+ (ratio = 0, lanes 1, 6 & 9), +/- (ratio = 1, lanes 3-5 & 7), and -/- (ratio = 2, lanes 2, 8 & 10). (B) An example of densitometric measurements of the band intensities for +/+ , +/- and -/- genotypes. Peak 1: *en-2* band, peak 2: *lacZ* band. (C) *LacZ* disrupts the *mym* genomic locus. *mym* 5' exon was cloned by 5' RACE-PCR. The primers a, b, and c were used for PCR on genomic DNA to confirm disruption of the *mym* locus by the gene trap vector. (D) PCR on genomic DNA from +/-, +/- and -/- embryos and Southern blot probed with the *mym* exon. (E-G) Localization of the *mym* gene in a set of mouse metaphase chromosome spread was performed by FISH. (E) PI-stained chromosomes with FITC signals, (F) same chromosomes stained with DAPI, (G) diagrammatic view of the banding pattern of the mouse chromosome 5 with the corresponding *mym* signals in the B3 region.
Figure 4.3. Whole mount embryonic and adult *lacZ* expression. (A) Expression at E9.5 & (B) E10.5 in the heart, gut, craniofacial structures and the otic vesicle, (C) E11.0, expression in the heart and gut, (D) E13.5 and E14.5, expression in the leptomeninges covering the whole central nervous system, (E) E10.5, expression in the apical ectodermal region of limb buds (arrowhead), (F) E11.5, expression in the heart, (G) E11.5, expression in the choroid plexus (arrowhead), (H) E13.5, in the undifferentiated mesoderm of the limbs (arrow), (I) expression in the adult hippocampus (hp), and (J) in the Purkinje cell layer (pcl) of the cerebellar cortex.
Germ line transmission and analysis of F2 progeny. The mym trapped locus was transmitted through the germ line of mouse chimeras to produce heterozygous animals. Heterozygotes were identified by Southern analysis and bred to obtain F2 progeny in three different genetic backgrounds (see Materials and Methods). Genotyping was performed using a quantitative Southern blot analysis (Figure 4.2A & 4.2B). Out of 263 animals in the F2 progeny of all three backgrounds, no homozygotes were detected (Table 4.1). Therefore, the mym gene is required to complete embryogenesis.

129 males show reduced fertility. Heterozygous males in the pure 129 genetic background were sterile or had reduced fertility resulting in few progeny. The penetrance of sterility/reduced fertility was 82% (ntotal=17). Mating behavior of the 129 heterozygous males was however normal, as they were able to plug females. To understand the cause of reduced fertility of heterozygous males in the 129 background, we examined histological sections of their sexual organs and found that spermatogenesis was severely affected (Figure 4.5). In the testis, the seminiferous tubules structure was loose and disorganized, and spermatid flagella were ruptured. In the epididymis where sperms mature, there were few or no spermatozoa present.
Figure 4.4. *LacZ* expression in embryonic sections at E11.5. (A & B) *LacZ* staining in the ventricle (v) and atrial (a) myocardium, in the myocardial wall (mw) and in the trabeculae (tb); note the organization of endothelial cells (en) surrounding the trabeculae. (C) Expression in the ependymal cells (ec) of the choroid invagination and (D) in midgut (mg), hindgut (hg) and urogenital ridge (ur).
Figure 4.5. Histological sections of heterozygous male sexual organs. Spermatogenesis is affected in the 129 heterozygous males. (A & C) +/+ males, and (B & D) +/- males. (A & B) Testis. Seminiferous epithelium (se), composed of Sertoli and spermatogenic cells, and flagella (f). In the +/- males, the seminiferous tubules structure are loose and disorganized, and spermatid flagella are ruptured. (C & D) Epididymis. Epithelium (e), spermatozoa (sp), connective tissue (ct), and stereocilia (sc). No spermatozoa is present in the epididymis of +/- males.
**Homozygous embryos die during midgestation.** To understand the time and the cause of embryonic lethality more precisely, we undertook systematic dissection of embryos at various gestation time points (Table 4.1). Due to reduced fertility in the 129 background, we performed all embryonic phenotypic analyses in the two mixed genetic backgrounds. Upon dissection of midgestation uteruses from the heterozygote breedings, we noticed some of the conceptuses were about half the size of the others (Figure 4.6A). We dissected out the embryos at E10 and used their yolk sac for isolating genomic DNA for genotyping. The retarded conceptuses were homozygous for the gene trapped locus and were of two distinct phenotypes. About one third of the homozygous embryos were very small, grossly abnormal in appearance and stalled at the headfold stage (Figures 4.6C-D). Histological sections revealed that these embryos had rather well-developed organs and structures such as yolk sac, head folds, neural tube and heart tube (Figures 4.6E-H). Such grossly abnormal homozygous embryos were detected as early as E9.0, but never after E10.5, most likely due to resorption.

The remaining two thirds of the homozygous embryos were about half the size of their wild type littermates, and presented a characteristic heart phenotype with differing degrees of severity (Figures 4.6B-C & 4.7). The less severely affected embryos had a beating heart at the time of dissection. We recovered three embryos which had the heart phenotype and were scored as heterozygotes, but these were likely due to misgenotyping by quantitative Southern. We did not detect any live homozygous embryos beyond E10.5, suggesting embryonic lethality during this stage of development (Table 4.1).
Table 4.1. Genetic analysis of the *mym* gene trap insertion. Genotyping of embryos was done between days 9 and 11.5 of gestation (E9-E11.5). Genotyping of animals was done at 4 weeks of age (P28).
<table>
<thead>
<tr>
<th>Stage</th>
<th>Total</th>
<th>Genotype</th>
<th>-/- Embryos</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>+/+</td>
<td>+/-</td>
</tr>
<tr>
<td>E9-9.5</td>
<td>50</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>E10.0</td>
<td>43</td>
<td>12</td>
<td>21</td>
</tr>
<tr>
<td>E10.5</td>
<td>36</td>
<td>7</td>
<td>21</td>
</tr>
<tr>
<td>E11.5</td>
<td>11</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>P28</td>
<td>263</td>
<td>102</td>
<td>161</td>
</tr>
</tbody>
</table>

<sup>a</sup>The embryos were highly deformed and developmentally delayed.

<sup>b</sup>The embryos were reduced in size and presented the typical heart phenotype.

<sup>c</sup>Includes deteriorated embryos.
Figure 4.6. Phenotypes of *myo* homozygous embryos. (A) Midgestation uterus dissected out from heterozygous breeding showing the smaller mutant conceptuses (arrowheads). (B) A wild type embryo at E10, and its mutant littermate (on the right). (C) One delayed embryo with heart defect on the left and a grossly abnormal mutant embryo on the right. These two mutant embryos were also littermates, E10. (D) Two other grossly mutant embryos, headfolds are indicated by arrowheads. (E-H) Histological sections of the grossly abnormal embryos, indicated are: headfold (arrowheads), neural tube (nt), yolk sac (ys), heart tube (ht), foregut (fg) and dorsal aorta (da).
**mym is required for heart development.** To examine the general morphology of mutant hearts, wild type and mutant hearts at E10.0 were prepared for and micrographed by scanning electron microscopy. The mutant hearts were of a variably irregular shape (Figures 4.7A-B). Histological analyses were performed on haematoxylin and eosin-stained paraffin sections of wild type and mutant hearts and micrographed under UV for higher resolution (Figures 4.7C-E). Atrium and ventricle, as well as the three cell lineages, endocardium, myocardium and pericardium were present in the mutant hearts; however, their organization was affected. The myocardial wall was only one or two cell layers thick, and the trabeculae, a spongiform network of myocytes infiltrated by the endocardium, were also thin and not well-organized throughout the heart including the bulbus cordis. There was also no clearly distinguishable endocardium surrounding the trabeculae. The endocardial cavity was almost absent and there was blood trapped in both atrium and ventricle indicating a failure in cardiovascular circulation (Figure 4.7D).

To understand the progression of the phenotype in the heart, systematic heterozygote test breedings were set up and embryos were dissected out at embryonic days 9-9.5, 10.0 and 10.5. Wild type and homozygous mutant embryos were sectioned and their hearts examined and compared under light microscopy (Figure 4.8). At E9.0, wild type hearts are highly looped and their walls consist of myocardial and endocardial layers separated by cardiac jelly (Viragh and Challice, 1977). Mutant hearts were as well-formed and normal in appearance as their wild type littermates at E9.0, except for some disorganization of the trabeculae and a larger space between the interacting endocardium and myocardium (Figure 4.8A, 4.8D & 4.8E). At E9.5, however, a distinct phenotype was clear in the myocardium. The trabeculae and myocardial wall were disorganized and the pericardial layer was loosely attached to the underlying myocardium (Figure 4.8H). At E10, the disarray in myocardial wall and trabeculae was more severe and the heart retained its less developed tube-like appearance. The endocardial cells also did not form the orderly structures around the trabeculae and endocardial cushion formation was
underdeveloped (Figure 4.8F). In some instances, part of the heart tube had completely collapsed (Figure 4.8I). At E10.5, the severity of the myocardial defect affected the integrity of the myocardial wall, leading to bleeding in the pericardial cavity and disruption of blood circulation (Figure 4.8G). Some mutant embryos had already started deteriorating at this stage (Figure 4.8J).
Figure 4.7. Embryonic heart phenotype. General morphology of +/+ and +/- hearts at E10. Scanning electron microscopy of: (A) a +/+ wild type heart, myocardial wall (mw), trabeculae (tb), endothelium (en), and endocardial cavity (ec); and (B) a +/- mutant heart in which tb, en and ec are not easily recognizable. UV transillumination of histological sections of: (C) a wild type heart; and (D & E) mutant hearts, (v) ventricle. Note the blood trapped in the atrium (a), and the disorganization of the trabeculae (tb) in D & E.
Figure 4.8. Time series analysis of homozygous embryonic hearts. Progression of the phenotype in mutant hearts between E9.0 to E10.5. (A-C) +/- control hearts, (D-J) -/- mutant hearts. Atrium (a), ventricle (v), pericardium (p), myocardium (m), endocardium (e), pericardial cavity (pc), truncus arteriosus (ta), trabeculation (tb), and endocardial cushion (ec) are indicated. Note the underdeveloped endocardial cushion and the tube-like structure in the E10 heart in F. See the text for further description.
Expression of myocyte-specific genes in mym mutant hearts. To understand the defect in the myocardium and the stage of myocyte differentiation in the mutant hearts, we examined the expression of myocyte-specific genes at E10 by RNA in situ hybridization (Figure 4.9). Mutant myocytes, like their wild type counterparts, expressed differentiated myocyte markers such as α-cardiac actin, myosin light chain 1A and the muscle-specific isoform of adenylosuccinate synthetase (Figures 4.9E-I), but they failed to express myofilament genes, myosin light chains 2A and 2V (MLC2A & -2V) (Figures 4.9A-D). MLC2A was expressed in both ventricles and atrium, while MLC2V expression was restricted to ventricles in wild type hearts. A few cells in the mutant ventricles may still be expressing MLC2V (Figure 4.9D), while MLC2A expression was totally absent (Figure 4.9B). Thus, in the absence of mym function, the MLC2A & MLC2V genes were not activated. Expression of the endothelial-specific receptor tyrosine kinase, Flk-1, was also examined to visualize the organization of the endothelium in the mutant hearts (Figures 4.9K-L). The endothelial cells in the mutant ventricles were dispersed and less orderly than their counterpart in the wild type heart, likely due to the disorganization of the myocardium and trabeculae.

Failure in myocyte maintenance. We next examined the structural organization of myocytes at E10 by transmission electron microscopy (Figure 4.10). Cardiac myocyte ultrastructural features include abundant mitochondria, extensive deposition of myofilament arrays and specialized intercellular junctions, desmosomes and facia adherens with anchoring myofibrils (Hirakow and Sugi, 1990). Mutant myocytes failed to organize actin-myosin and thin filaments in their cytoplasm and there was no evidence of cell junctions between mutant myocytes. The less severely affected embryos with beating hearts at the time of dissection had evidence of cell junctions and some actin-myosin contractile elements in their cardiac myocytes (not shown). It seems that myocyte ultrastructural defects are the underlying cause of cardiac failure in mutant embryos,
however, its severity is likely influenced by modifier genes in the mixed genetic backgrounds used in these phenotypic analyses. We also detected the presence of condensed cytoplasm, apoptotic bodies and electron dense nuclei in the mutant myocardium, indicative of apoptotic cell death (Figures 4.10C-D).
Figure 4.10. Structural organization of myocytes analyzed by transmission electron microscopy. (A & B) +/- control myocytes and (C & D) +/- mutant myocytes. Indicated are: myofibrillae (m), Z lines of actin-myosin filaments (arrowheads), nucleus (n), cytoplasm (c), sarcoplasmic reticulum (sr), thin filament (tf), desmosome (d), apoptotic nucleus (an), apoptotic body (ab), and condensed cytoplasm (cc). Note the absence of cell junctions between myocytes in C and in between the two arrows in D where membranes of neighboring cells are in close proximity. Magnifications: (A) X6,500; (B) X50,000; (C) X6,000; (D) X15,000.
DISCUSSION

In a gene trap screen of mouse embryonic stem cells, we have identified a novel gene, *mym*, that is required for myocyte maintenance during embryonic heart development. *mym* is expressed in the heart, gut, craniofacial structures and limb buds. Expression in the heart is first detected at day 9.0 of gestation in the myocardium. At day 10, *mym* is expressed in the myocardial wall and the trabeculae where the phenotype of the homozygous mutant embryos can first be detected during development. The gene trapped *mym* locus is a recessive lethal mutation. Two thirds of the homozygous mutant embryos die during midgestation due to cardiovascular failure. In the mutant hearts the myocardium is hypoplastic resulting in myocardial wall and trabeculae that are thin, reduced in size and disorganized. Myocardial defects often lead to trapping of the blood in the endocardial cavity and bleeding into the pericardial cavity. Both of these defects were observed in the *mym* mutant hearts. At E10.5, necrosis is evident in some mutant embryos indicating a progression toward resorption. No live mutant embryos were observed after E11.0, indicating day 11.0 of gestation as the time of embryonic arrest.

All cardiac morphogenetic processes such as looping, formation of the chambers, valve, and endocardial cushion, and trabeculation are initiated in *mym* mutant hearts. However, their further development is generally arrested due to a defect in myocyte maintenance. The earliest signs of myocardial defects were seen at day 9.0 when some disorganization in the trabecule structure first became apparent. The impairment progressively worsened until day 10.5 of gestation when some mutant embryos were already dead and had started deteriorating. At this stage of development, when the yolk sac-based circulation is being replaced by cardiovascular circulation, the heart is under considerable mechanical load to supply the embryo with blood (Kaufman 1992). Any structural defects that may compromise the integrity of cardiovascular circulation at this stage are going to be detrimental to embryonic life. This explains why mutations in many
genes involved in cardiovascular development or function lead to embryonic lethality during the midgestation (see Introduction).

At day 10.25 to 10.5 of gestation, two obvious morphogenic changes take place within the heart; i) left and right components of the common atrial chamber begin to separate as the first clear indication of the septum primum is seen, and ii) there is marked trabeculation in the bulbus cordis (Kaufman 1992). Both of these changes are directly related to myocardial proliferation which is severely compromised in the mym mutant embryos. In the mym mutant hearts, we observed that myocytes are disorganized, their cell numbers are reduced and trabeculation is severed, precluding the landmark midgestation developmental changes in the heart to take place.

mym mutant myocytes express differentiated lineage gene markers, myosin light chain 1A (MLC1A), α-cardiac actin and the muscle-specific isoform of adenylsuccinate synthetase, indicating that they have initiated and progressed along their differentiation pathway. In contrast, mym myocytes failed to express myosin light chains 2A and 2V (MLC2A & -2 V), demonstrating that mym is required for a component of the cardiac myogenic pathway that is not essential for the expression of the other myosin and actin genes. The expression of different myosin isoforms during cardiac development is suspected to be under different regulatory mechanisms as their differential and dynamic patterns of expression change at different rates during development (Lyons et al., 1990).

MLC1A and -2A expression are normally restricted to the atria, while MLC2V expression is restricted to the ventricular chambers (Kubalak et al., 1994; Lee et al., 1992; Lyons et al., 1990). The fact that MLC2A and -2V, but not MLC1A expression is deregulated indicates that mym is not affecting myogenesis in a chamber-specific manner. Different myosin genes perform different physiological roles in cardiac muscle cells (Malhotra et al., 1979). Lack of MLC2A and -2V expression in the mym hearts has likely compromised myocardial function, and thereby blood circulation, leading to the death of mutant embryos. Interestingly, the anthracycline antibiotic doxorubicin, used in cancer
therapy, decreases in a dose-dependent manner the levels of MLC2 gene expression in cardiac muscle and as well leads to a myofibrillar loss in a characteristic cardiac myopathy (Ito et al., 1990). The deregulation of MLC2 gene expression and its concomitant myocardial defect in doxorubicin treatment is similar to the absence of mym gene function presented here.

MLC2V expression is also inactivated in Nkx2-5 mutant hearts, while the expression of other myofilament genes remain intact (Lyons et al., 1995). Nkx2-5 is a homeobox-containing transcriptional regulator and a marker of the early stages of cardiogenesis (Harvey 1996). It is homologous to the Drosophila tinman gene and its targeted interruption leads to midgestation embryonic lethality due to abnormal heart morphogenesis. The myocardial and trabecule defects in the Nkx2-5 mutant hearts are very similar to the defects in mym mutant hearts presented here. Myofibrillogenesis and the ultrastructural organization of the myocytes are not, however, compromised in the Nkx2-5 mutant hearts. The embryonic lethality in Nkx2-5 mutant embryos is at 9-10 days of gestation which is at least a day earlier than mym mutants. On the other hand, Nkx2-5 mutant hearts present an early looping defect which we did not observe in the mym hearts. Similar myocardial defects and deregulation of myosin genes in both mym and Nkx2-5 mutants, and earlier more extensive phenotype in the Nkx2-5 suggest that mym may be downstream of Nkx2-5 in the myocardial regulatory pathway.

A gene targeted mutation in Plakoglobin, a cell adhesion molecule involved in cell junctions, also leads to embryonic death after E10.5 due to heart defects very similar to those in Nkx2-5 and mym mutant hearts. The Plakoglobin heart phenotype includes hypoplastic ventricular wall, trabeculae, and endocardial cushion, and defects in myofibrillar organization (Bierkamp et al., 1996), suggesting that Plakoglobin may also be in the Nkx2-5 regulatory pathway.

The mym gene is located on mouse chromosome 5, region B3, close to the W locus where a recessive lethal mutation (15H1) has also been mapped (Lyon and
Glenister, 1982). $W^{19H}$, a deletion in the $W$ locus also leads to embryonic lethality in homozygotes (Lyon et al., 1984). $W^{19H}$ and $15H1$ mice may be used in crosses with $mym$ mice for complementation and mapping experiments to understand the genetic relations between these three lethal mutations. $Nkx2-5$ and $plakoglobin$ mutations present very similar phenotypes to $mym$ gene trap mutation in embryonic heart but map to different chromosomes (17 and 11, respectively, Himmelbauer et al. 1994; Guenet et al. 1995). At present we have only a partial $mym$ cDNA with a limited ORF which is novel and is not similar to any sequences in the database. Further cloning is required to obtain complete sequence information.

In the $mym$ mutant myocytes, the degree of myofibrillogenesis seen was severely reduced. We examined mutant myocytes from embryos with severe heart phenotypes at day 10 of gestation. The actin-myosin and thin filaments in the cytoplasm were absent and the desmosomal cell junctions were not present between cells. Some of the less severely affected embryos, however, had beating hearts at the time of dissection, and had evidence of myofibril contractile elements. It is possible that such embryos lose their myofibrils some time later than their more severe mutant littermates and eventually succumb by day 11 of gestation. We have not yet confirmed the deletion of myocytes by apoptotic cell death in a more direct investigation of the phenomenon; however, myocyte cell death was detected by electron microscopy and likely contributed to the myocardial disorganization and trabecule disruption. In mutant embryos, such defects incite heart and circulation failure when rapid proliferation and rearrangement of cardiac myocytes are the primary contributing factor to heart development (Fishman and Chien, 1997).

It is quite likely that the absence of MLC2A and -2V in mutant myocytes is a major reason for the absence of contractile elements in these cells. Interactions between myosin light chains and actin are known to regulate cardiac contractility (Morano et al., 1995). In the less severe mutant embryos, it is possible that other myosin isoforms can
initially replace the function of missing myosin light chains. Later, when the mechanical load on cardiac tissue becomes more demanding, functional replacement by other isoforms may no longer be sufficient for the maintenance of myocytes.

Because the heterozygous males in the pure 129 background were mostly sterile, we performed all the phenotypic analyses in mixed genetic backgrounds. We therefore cannot exclude the action of modifier genes affecting the severity or presentation of the phenotype in the myocardium. An earlier embryonic lethality would be particularly likely since as early as day 9.0 of gestation cardiovascular circulation is required to supplement the yolk sac circulation, and therefore a mutation affecting myocyte structural integrity and their cell junctions would be expected to be detrimental to the embryo shortly after day 9 of gestation. However, if no modifier genes turn out to be involved in the observed phenotype, the severity or time of presentation, then the late effect of the mym mutation may reflect functional redundancy of the myosin isoforms.

In summary, our results with the gene trap mutation in the mym locus demonstrate an essential role for this gene in the maintenance of myocytes during midgestation embryonic heart development. The disruption of structural organization of the myocytes and deregulation of myosin light chain gene expression suggests a key role for mym in proper execution of cardiac myogenic processes.
Chapter 5: Concluding Remarks
Summary and General Discussion

Over the last few years, gene trap strategies using mouse embryonic stem (ES) cells have been developed to perform genetic screens in mammals for novel genes and mutations. The appeal of gene trapping is in its power to characterize simultaneously three different aspects of developmentally regulated genes: i) expression patterns by the use of a reporter gene; ii) cloning by RACE-PCR and sequence analysis; and iii) transmission through the germ line and phenotypic analysis of the heterozygous and homozygous disruptions. Furthermore, as ES cells can be grown and manipulated in vitro, it is convenient to prescreen for features of interest in culture before further characterization in vivo.

The gene trap strategy, combined with other transgenic and gene targeting technology now available in mice, provides a powerful array of experimental approaches to dissect genetic pathways involved in mammalian development. The RA induction gene trap screen in which I193 and R140 clones were isolated indicate the feasibility of this approach to identify genes that lie downstream of ligand/receptor-mediated signaling pathways. The RA induction screen significantly enriched for integration events into genes that respond to an inducer in vitro and that present restricted expression patterns during development. For instance, 19 of 20 (95%) RA-responsive ES cell lines displayed a restricted expression pattern between E8.5 and E11.5 (Forrester et al., 1996). In contrast, gene trap screens in which cell lines were selected solely on the basis of lacZ expression in undifferentiated ES cells, only 30% displayed specific expression patterns at these stages (Wurst et al., 1995). Potentially, by scaling up an induction gene trap screen, it may be possible to identify most genes involved in any given genetic pathway that are active in ES cells.

As described in Chapter 1, other genes involved in developmental processes have been identified in gene trap approaches. In addition, other gene trap experiments have
identified genes that are specific to certain developmental lineages such as hematopoietic lineages, as well as genes that are responsive to other physical, chemical or molecular agents (W. Stanford, G. Caruana, and W. Wurst, unpublished data).

Dr repeats

In Chapter 2. Dr repeats, a novel class of expressed repetitive DNA elements in the mouse genome, were described. Dr expression is spatially restricted in embryos and in the adult brain, and in vitro is induced by RA. The members of the Dr family are highly similar in sequence and show peculiar structural features. Repetitive DNA sequences form a substantial portion of eukaryotic genomes. Although there is evidence that some repetitive sequences may participate in gene regulation and other cellular functions, due to their repetitive nature and global presence in the genome and transcriptome (the sum total of all transcripts in a cell), functional studies on these nucleic acid sequences have been difficult. A few experimental approaches may however be suggested to further elucidate the cellular role of these sequences.

Our results raise the possibility that the expression of Dr-containing transcripts is part of an ES cell differentiation program induced by RA. Employing a transgenic approach, the precise length of Dr repeats required for RA response can be determined and their functional domains can be dissected. In particular, since Dr sequences may be part of a much larger repetitive sequence, it is conceivable that other sequence domains exist within these elements that are essential for other possible structural or regulatory functions such as interaction with regulatory proteins. Using transgenic experiments, the spatial and temporal domains of Dr expression and function can also be addressed. Transgenic expression of Dr repeats in heterologous tissues may also shed some light on their function. We know that Dr sequences are highly dispersed in the genome. As the
mouse genomic sequences become available, the exact length, location and organization of these elements in the genome may further indicate some functional or structural roles for them. Finally, it may be possible to inactivate these sequences by overexpressing their antisense RNA.

Dr sequences include a polyadenylation signal, a poly A track and long 3' palindromic sequences indicating a retrotransposition event in the history of these sequences. Dr repeats may be part of a mobile genetic element and as such may have played a role in the construction of the genome organization.

Aquarius

Aquarius is a novel developmentally regulated gene described in Chapter 3. During embryogenesis, it is expressed in mesoderm, neural crest cells (NC) and their target tissues, and in neuroepithelium. There are many genes with restricted expression in subsets of NC cells. For instance, the transcription factors AP-2, AP-2.2, Dlx-2, Hoxa-2, Pax3, Pax7 and twist are all expressed in cranial NC cells (Chazaud et al., 1996; Chen and Behringer, 1995; Gendron-Maguire et al., 1993; Mansouri et al., 1996; Qiu et al., 1995; Tremblay et al., 1995; Zhang et al., 1996). In a cell lineage marker analysis, it would be of interest to compare Aqr expression pattern with other NC-expressed genes. Such analysis would clarify the subset of NC cells in which Aqr may play a role.

In vitro, Aquarius is responsive to RA and its transcript carries the RA-inducible Dr repeat element. By isolating genomic sequences and the Aqr promoter region, it will be possible to see whether Aquarius carries RA responsive elements or whether its activation by RA is indirect. Alternatively, there may be a novel RA induction pathway through the Dr elements as was suggested in Chapter 2. The purpose of the original RA screen was to identify genes downstream of RA signaling. Future experiments will be
needed to indicate whether \textit{Aqr} is responsive to RA \textit{in vivo}, and if so, where in the RA signaling pathway, \textit{Aquarius} may play a role.

\textit{Aquarius} cDNA sequence shows weak similarity to RNA-dependent RNA polymerases. To understand whether the sequence similarity to RNA-dependent RNA polymerases has a functional relevance, the protein product of the gene produced in an \textit{in vitro} expression system could be examined for RNA binding and polymerase activity. At present, little is known about the structure and function of RNA-dependent RNA polymerases.

The \textit{Aquarius} protein product may also be used to raise antibody for immunohistochemical studies for cellular localization of the gene product and to precipitate from cellular extracts other proteins that may interact with the Aquarius protein. Cloning and sequence analysis of the other \textit{Aquarius} isoforms identified by Northern analysis may also provide more information about other possible functional domains in this gene. The presence of nuclear localization signal in the \textit{Aquarius} sequence suggests the presence of its gene product in the cell nucleus.

The confirmation of an RNA-dependent RNA polymerase in the mouse, would not be that surprising, as several reports have already indicated the presence of such activity in eukaryotic organisms (Schiebel \textit{et al.}, 1993a; Schiebel \textit{et al.}, 1993b; Volloch \textit{et al.}, 1991; Lai 1995). The Tetrahymena telomerase is also similar to viral RNA-dependent RNA polymerases (Collins \textit{et al.}, 1995). It will then be of interest to know, through a biochemical assay, if the Aquarius protein possesses telomerase activity.

Mice homozygous for the \textit{Aqr} gene trap insertion are viable and normal. However, the gene trap allele described is not a null mutation as \textit{Aqr} RNA is present in homozygous animals. The lack of a discernible phenotype may be due to read-through transcription and splicing across the trapping vector. To understand the function of \textit{Aquarius} during embryogenesis, it is essential to mutate it by standard gene targeting techniques.
Using the Cre-\textit{loxP} recombination system, it is possible to mutate a gene in a temporal or tissue-specific manner to more accurately dissect gene function during specific times or in specific tissues. This technique requires the presence of \textit{loxP} sites in the genomic locus, and transient or tissue specific expression of Cre recombinase, both of which are now possible in the mouse using embryonic stem cells to generate the lines (Smith \textit{et al.}, 1995). To study different domains of the Aquarius protein, subtle mutations can now be introduced through the 'knock in' version of gene targeting, where a replacement vector introduces the mutated version of the gene into its genomic locus (Castilla \textit{et al.}, 1996).

\textit{mym}

The gene trapped locus in the R140 ES line was named \textit{myocyte maintenance} (\textit{mym}) as it is required for the development of cardiac myocytes as described in Chapter 4. During midgestation embryonic development, \textit{mym} is expressed in cardiac myocytes, limbs, gut, meninges and the choroid plexus. Homozygous mutant embryos died during midgestation due to heart failure. Mutant hearts presented thin myocardial walls and defective trabeculae. The myocardium, where \textit{mym} is expressed, was hypoplastic due to myocyte cell death. Apoptotic cell bodies and electron dense nuclei indicative of apoptotic cell death were detected by electron microscopy.

To understand the myocyte defects better, myoblasts from early homozygous mutant embryos should be cultured and further investigated \textit{in vitro}. For instance, as myosin light chain genes are deregulated in these myocytes, it would be informative to document \textit{in vitro} the deregulation of these and other myocyte markers. The apoptotic response needs to be confirmed \textit{in vitro} and \textit{in vivo}, and investigated further at the molecular level. Cell division as well as cell differentiation with its ensuing cellular
changes, such as myofilament expression and organization, can also be examined in mutant myocytes *in vitro*.

To bypass midgestation embryonic lethality and to examine other possible functions of *mym* later in development, chimeric analysis can be undertaken in which homozygous mutant ES cells are selected in high G418 concentration and used to generate aggregation chimeras. Homozygous mutant myocytes will then be amongst wild type cells which, depending on the degree of chimerism, might compensate for myocardial defects. This analysis should determine whether the *mym* defect is cell autonomous (the gene function is required in the cells that express the gene), or cell non-autonomous (the gene function is required in neighboring cells).

About one third of homozygous mutant embryos were blocked at the headfold stage (E8.5) with no further development up to E10.5. These early defective embryos suggest a more general cell division or biological defect due to *mym* disruption. Lineage marker analysis and *in vitro* culture of embryonic cells may shed some light on the cause of this early developmental defect. Ideally this defect should be examined in a pure genetic background as the two phenotypes (E10.5 myocardial defect & E8.5 developmental block) and their penetrance (66% & 33%) are likely due to modifier gene effects.

Heterozygous *mym* males had reduced fertility in a pure 129 genetic background due to a defect in spermatogenesis. This defect, although a disadvantage for studying *mym* in a pure genetic background, indicates the likelihood of phenotypic variation in different genetic backgrounds. There are indeed other cases of phenotypic variations in mice in different genetic backgrounds (Ramirez-Solis *et al.*, 1993; Threadgill *et al.*, 1995). The *mym* defect in spermatogenesis also needs further investigation perhaps by lineage marker analysis and by *in vitro* culture of spermatocytes.

Similar myocardial defects and deregulation of myosin genes have been reported in *NkrZ-5* mutant hearts (Lyons *et al.*, 1995). Embryonic death after E10.5, hypoplastic
ventricular walls, trabeculae and endocardial cushion, and defects in myofibrillar organization have also been detected in mice homozygous for a null mutation in the *Plakoglobin* gene (Bierkamp *et al.*, 1996). The remarkable similarities in time and tissue-specificity of these mutant phenotypes suggest functional roles in the same pathway. Since genetic interactions are of great value in unraveling the molecular basis of developmental processes, gene interactions among these three loci can be investigated by breeding mice carrying these mutations, generating double mutant animals and examining their phenotype. This is possible because each mutation manifests phenotypic features that are not shared by the other two. Therefore, by scoring for the presence or absence of the exclusive phenotypic features and the severity of the phenotypes in double mutants, it will be possible to gain some insights into whether the products of these genes functionally interact during heart development.

**Conclusions**

Developmental genetics seeks a molecular understanding of the developmental processes that perpetually produce progeny in each successive generation. Such an understanding at the level of gene function is currently limited in mammalian embryonic development. The development of transgenic technology, encompassing gene targeting and trapping methodologies, has initiated a flourishing era of molecular investigations into mammalian developmental genetics. This approach is promising to narrow the 'phenotype gap,' the gap in the available mutant resources (2% of the full range of phenotypes possible), that now exist in the mouse as a model organism (Brown and Peters, 1996).

The work described here demonstrates the feasibility of *in vitro* manipulation and screening of ES cells to identify and rapidly characterize genes and genetic elements that
are developmentally regulated or lie along a cellular response pathway. The generation of gene trapped mice allows for expression and functional analysis of the trapped genes.

Sequence databases, analytical softwares and ready access to these tools through the internet have led to an increase in gene identification by computer (Fickett 1996). Approximately 50% of all human genes are already represented in the Expressed Sequence Tag (EST) databases (Gerhold and Caskey, 1996). A similar EST project is now underway at the Washington University, St Louis USA, to develop 400,000 mouse ESTs. The combination of gene trap experimentation and EST databases provides ready access to the sequence of trapped genes and has the potential to increase several fold the pace of gene discovery and functional characterization.
Literature Cited


152


Kucherlapati RS, Eves EM, Song KY, Morse BS, and Smithies O. (1984). Homologous recombination between plasmids in mammalian cells can be enhanced by treatment of input DNA. *Proc Natl Acad Sci USA* 81, 3153-3157.


and second branchial arch derivatives and abnormal differentiation in the forebrain. *Genes Dev* 9, 2523-2538.


IMAGE EVALUATION
TEST TARGET (QA-3)

150mm

6"

APPLIED IMAGE, Inc
1683 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/286-5969

© 1993, Applied Image, Inc., All Rights Reserved