Characterisation of Gene-Trap Integrations Expressed During Mouse Heart Development

GURMAN SINGH PALL

Thesis presented for the degree of Doctor of Philosophy
University of Edinburgh

1998

I declare that the work presented in this thesis is my own,
except where otherwise stated.

To my Grandparents
Sant Singh Pall and Dhandey Kaur Pall

Acknowledgments

I would like to thank my supervisor Lesley Forrester for encouragement, support and advice during the course of my PhD in her laboratory.
My thanks also go to John Mullins, my second supervisor, and to John Bishop, Bill Skarnes and Austin Smith, who as members of my PhD committee, discussed and encouraged my research.
I have enjoyed the company of many members of Lesley's laboratory past and present and would especially like to thank Alistair Watt, Peter McClive, Philippe Gabant and Jane Brennan for their discussions and for generating a pleasant environment to work in. I have also had the pleasure of discussing science with many other members of the CGR and take this opportunity to thank them, particularly Matt Sharp. Thanks also go to the animal house staff, particularly Vanessa and Andrew Jeske, for providing excellent facilities.
Finally I thank my family for support, love and encouragement, and especially my children for the happiness they shine into my life.

Abstract

The molecular mechanisms involved in heart development are poorly understood and only a few heart specific molecular markers are known. A gene-trap strategy in embryonic stem (ES) cells has been employed to identify and characterise genes involved in heart development. This work describes the characterisation of two gene-trap integrations (R68 and R124), including identification of endogenous trapped gene sequences and analysis to determine the function of the trapped genes.

Molecular analysis of the R68 and R124 gene-trap integrations has shown the use of cryptic splice sites within the gene-trap vector suggesting the vector has integrated into an exon in both gene-trap cell lines. The unpredicted integration of a gene-trap vector into an exon still results in the expression of the reporter gene. Sequence data indicates a novel gene has been trapped by the R124 gene-trap integration. The structure of the R124 integration in the genome has been predicted based upon sequencing data and restriction fragment length polymorphism analysis. The R124 gene-trap integration has been mapped to chromosome 5 in the mouse genome using fluorescent *in situ* hybridisation labelling.

Beating cardiomyocytes generated from the *in vitro* differentiation of R68 and R124 ES cells express the reporter gene. Embryos heterozygous for the R124 gene-trap integration express reporter gene activity in the developing heart throughout gestation. In the adult the reporter gene is expressed in the heart, kidney, testis, ovary and brain.

Function of the trapped gene was assessed by generating animals homozygous for the R124 integration. 60% of animals homozygous for the integration die shortly after birth. This lethality is associated with a right ventricle heart defect. Surviving homozygote males show enlarged hearts and kidneys. Histology has shown generalised hypertrophy in the enlarged hearts and vacuoles in the tubules of the enlarged kidneys indicating renal dysfunction. The surviving homozygote males are also infertile, histological analysis has shown no mature sperm in the testes of homozygote males.

Characterisation of the R124 gene-trap integration has led to the discovery of a novel gene which is important for heart development. This gene is also important for male fertility as well as heart and kidney function in the adult.

Contents

Declaration
Dedication
Acknowledgments
Abstracts
Contents

Chapter 1

1.0	**General Introduction**	1
1.1	Heart Development	2
1.2	The Looping of the Linear Heart Tube	3
1.3	Chamber Specification	7
1.4	Chamber Maturation and Septation	9
1.5	Neural Crest Cell Contribution to the Heart	10
1.6	Function of the Heart	11
1.7	Changes in the Heart at Birth	13
1.8	Congenital Heart Disease	15
1.9	Effects of Retinoic Acid on Heart Development	16
1.10	*In Vitro* Early Cardiogenesis Model	17
1.11	Identification of Novel Genes	19
1.12	Random Mutagenesis	19
1.13	Identification of Conserved Genes	21
1.14	Embryonic Stem Cells	22
1.15	Directed Mutagenesis (or "Reverse" Genetics)	23
1.16	Mouse Mutations Affecting Heart Development	23
1.17	Gene-Trap Technology	24
1.18	Enhancer-Trap Vectors	25
1.19	Gene-Trap Vectors	26
1.20	Modified Gene-Trap Vectors	27
1.21	Transfection of ES Cells	29
1.22	Modified Gene-Trap Screens	30
1.23	Trapping in Practice	30
1.24	Generation and Isolation of Gene-Trap Cell Lines R68 and R124	31
1.25	Aim of the Project	33

Chapter 2

2.0 Materials and Methods 34

2.1 Molecular Techniques 34

2.1.1 General Cloning Procedures 34

2.1.2 Ligations 35

2.1.3 Transformation of Bacterial Cells 35

2.1.4 Plasmid Preparation - "Miniprep" 36

2.1.5 Sequencing Double Stranded DNA Templates 37

2.1.6 Isolation of Genomic DNA from ES cells 40

2.1.7 Isolation of Genomic DNA from Tissue 41

2.1.8 Southern Blot Analysis 41

2.1.9 Isolation of Total RNA from ES Cells and Tissue 41

2.1.10 Northern Blot Analysis 42

2.1.11 ^{32}P-Labelling Probe Fragments 43

2.1.12 Hybridisation Conditions 43

2.1.13 PCR Amplification of Genomic DNA 44

2.1.14 Rapid Amplification of cDNA Ends (5' RACE) 45

2.1.15 RNase Protection 48

2.1.16 Screening cDNA Libraries 50

2.1.17 Plasmid Constructs used for Molecular Analysis 53

2.2 ES Cell Methodology 54

2.2.1 Reagents for ES Methods 54

2.2.2 ES Cell Growth and Maintenance 55

2.2.3 Freezing ES Cells 55

2.2.4 Thawing ES Cells 55

2.2.5 Exposure of ES Cells to Retinoic Acid 56

2.2.6 *In Vitro* Cardiac Differentiation 57

2.3 Biological Specimens 58

2.3.1 Animal Maintenance and Breeding 58

2.3.2 Breeding Protocol for the R124 Gene-Trap Line 58

2.3.3 Animals Housed in Metabolic Cages 59

2.3.4 Specimens for Histology 59

Chapter 3

3.0 Preliminary Characterisation of Gene-Trap Integrations 64
3.1 Response of Reporter Gene Activity to Retinoic Acid 64
3.2 *In Vitro* Differentiation 65
3.3 Germline Transmission of the Gene-Trap Integrations 68
3.4 *In Vivo* Reporter Gene Activity in R124 Heterozygotes 69
3.5 Chromosomal Mapping 72
3.6 Detection of Fusion Transcripts 73
3.7 Identification of Trapped Sequences using RACE PCR 73
3.8 RNAse Protection Analysis of RACE Clones 75
3.9 Summary 77
3.10 Discussion 78
3.11 Conclusion 86

Chapter 4

4.0 Molecular Analysis of the R124 Gene-Trap Integration 87
4.1 Sequence Analysis of R124 ES Cell RNA derived RACE Clone 11 87
4.2 Vector Integration into an Endogenous Gene Exon 88
4.3 Isolation of cDNA Clones 88
4.4 Sequence Analysis of cDNA Clones 89
4.5 Detection of Endogenous Transcripts by Northern Blot Analysis 90
4.6 Genomic Structure of the R124 Gene-Trap Integration 91
4.7 Summary 94
4.8 Discussion 95
4.9 Conclusion 100

Chapter 5

5.0 Phenotype Analysis 101
5.1 Neonatal Lethality 101
5.2 Adult Phenotype 103
5.3 Kidney Function 104
5.4 Male Infertility 105
5.5 Summary 106
5.6 Discussion 107
5.7 Conclusion 116

Chapter 6
6.0 General Discussion 117

Appendix I 125
Appendix II 130
Abbreviations 136
List of Figures 138
List of Tables 139
References 140

Chapter 1

1.0 General Introduction

Introduction

The heart is one of the first organs to develop and become functional during vertebrate embryogenesis. It plays the essential role of pumping blood, nutrients, cellular products (such as hormones, cytokines and waste products) to and from all the tissues around the body via the vascular system. Congenital heart anomalies are the most common form of human birth defect and affect one in every 125 live births each year in the UK. Cardiovascular disease is the most common cause of premature death in the UK, claiming the lives of approximately 1 in 5 men and 1 in 10 women before the age of 75 i.e. 43% of the population (statistics from the British Heart Foundation).

Anatomical studies in man and animal models (pig, dog, rabbit, hamster, rat and mouse) of the normal and diseased heart have revealed heart development is essentially the same in all mammals and involves complex morphogenetic changes (via cell proliferation, migration, differentiation and shape change). Three stages of morphogenetic change are characteristic of mammalian heart development: formation of the linear heart tube; looping of the linear heart tube; septation of common atria and ventricles to give rise to the chambers of the adult heart (DeHaan, 1965; Kaufman *et al.*, 1981; Manasek *et al.*, 1972). Physiological studies have shown that normal heart function can be described via characteristic profiles and changes in these can be used for diagnostic purposes to determine the disease state of the heart (Colucci, 1994; Rockman *et al.*, 1994). Molecular approaches have begun to allow an understanding of the complex genetics regulating heart development and function (Chien, 1993; Doetschman *et al.*, 1993; Field, 1993; Fishman *et al.*, 1997; Fung *et al.*, 1996).

In this study advantage has been taken of the well described developmental stages of mouse embryogenesis and the ability to manipulate the genetics of the mouse, using embryonic stem cells, to isolate and characterise genes involved in heart development and function.

1.1 Heart Development

Cardiogenesis is initiated early during vertebrate embryogenesis, shortly after the onset of gastrulation. Some of the first mesodermal cells emerging from the primitive streak give rise to cardiac mesenchyme. The cardiac mesenchyme is thought to be induced by the underlying pharyngeal endoderm to proliferate and form angiogenic clusters (DeHaan, 1965; Jacobson *et al.*, 1988; Nascone *et al.*, 1995; Schultheiss *et al.*, 1995; Schultheiss *et al.*, 1977). The mesodermal derived clusters migrate laterally and then rostrally to surround the cranial neural plate by the late presomite stage (Gerhart *et al.*, 1986). The clusters extend across the midline and unite to form the cardiogenic region/crescent and the pericardial cavity. With formation of the cranial neural folds the cardiogenic region moves ventrally and rotates so that it comes to lie ventral to the foregut, forming the cavity of the foregut as it moves progressively more caudally. Splanchnopleuric mesoderm gives rise to the right and left primitive heart tubes, which fuse (in a rostral to caudal direction) to form a single endocardial tube, while the adjacent mesoderm thickens to from the myocardial mantle (Kaufman *et al.*, 1981). The anterior/posterior (rostral/caudal) axis of the heart, which is in line with the body axis of the embryo, is clearly defined at this stage (McGinnis *et al.*, 1992). Evidence of positional information along the anterior/posterior axis has come from exposure of embryos to retinoids, these cause anterior truncations of the heart tube in a dose dependent manner (Stainier *et al.*, 1992; Yamamura *et al.*, 1997). As soon as the precardiac mesoderm forms, the cells begin to express muscle specific gene products including myosin heavy chain (MHC), myosin light chain (MLC), actin, tropomyosin and various other structural gene products (Han *et al.*, 1992). Shortly after the single heart tube has formed it starts to beat. Initially the endocardium and myocardium are separated by the cardiac jelly (Manasek, 1975; Manasek, 1977), a extracellular matrix dense material, which plays an important role in the early modelling of the heart and is later invaded by endocardial cells. The endocardial tube forms the endocardium, the

internal lining of the heart, and the envelope of thickened myocardial tissue surrounding it give rise to the myocardium, which forms the ventricular and atrial walls. The endocardium is a single layer of flattened granulated cells and the myocardium is comprised of two to three cell layers of irregularly shaped myoblasts. At this stage the primitive heart is a double walled tube subdivided into three regions by constrictions: the atrium which recovers blood from the primitive veins; the ventricle which pumps blood out and the bulbus which is continuous with the aorta. Internally, a pair of sinus valves (right and left) guard the entrance into the atrium. By 8.5 d.p.c. (8-10 somites) the primitive heart is one of the most prominent structures of the embryo and starts to beat regularly (Kaufman, 1992; Moorman *et al.*, 1997). The primitive heart undergoes complex conformational changes in the following stages of development which are complete by 15 d.p.c., resulting in the recognisable four chambered heart connected to the major inflow and outflow blood vessels (Figure1.1).

1.2 **The Looping of the Linear Heart Tube**

Looping of the linear heart tube (between 8.5 - 9.5 d.p.c.) is an essential process that aligns the heart to the vascular connections in the embryo and arranges the presumptive chambers of the heart into their relative spatial positions observed in the adult. Overall the looping process translates the anterior/posterior positional information of the heart into a right/left orientation (Danos *et al.*, 1995).

Classical embryological studies show that the linear cardiac tube starts growing via cell division to a length longer than the dimensions of the pericardial cavity (Kaufman *et al.*, 1981, Manasek *et al.*, 1972). The linear cardiac tube is observed to physically bend to remain within the constraints of the pericardial cavity into the characteristic S-shaped conformation observed at late 8.5 d.p.c.. With further asymmetric growth through cell division and cell shape change the bulboventricular loop is pushed caudal

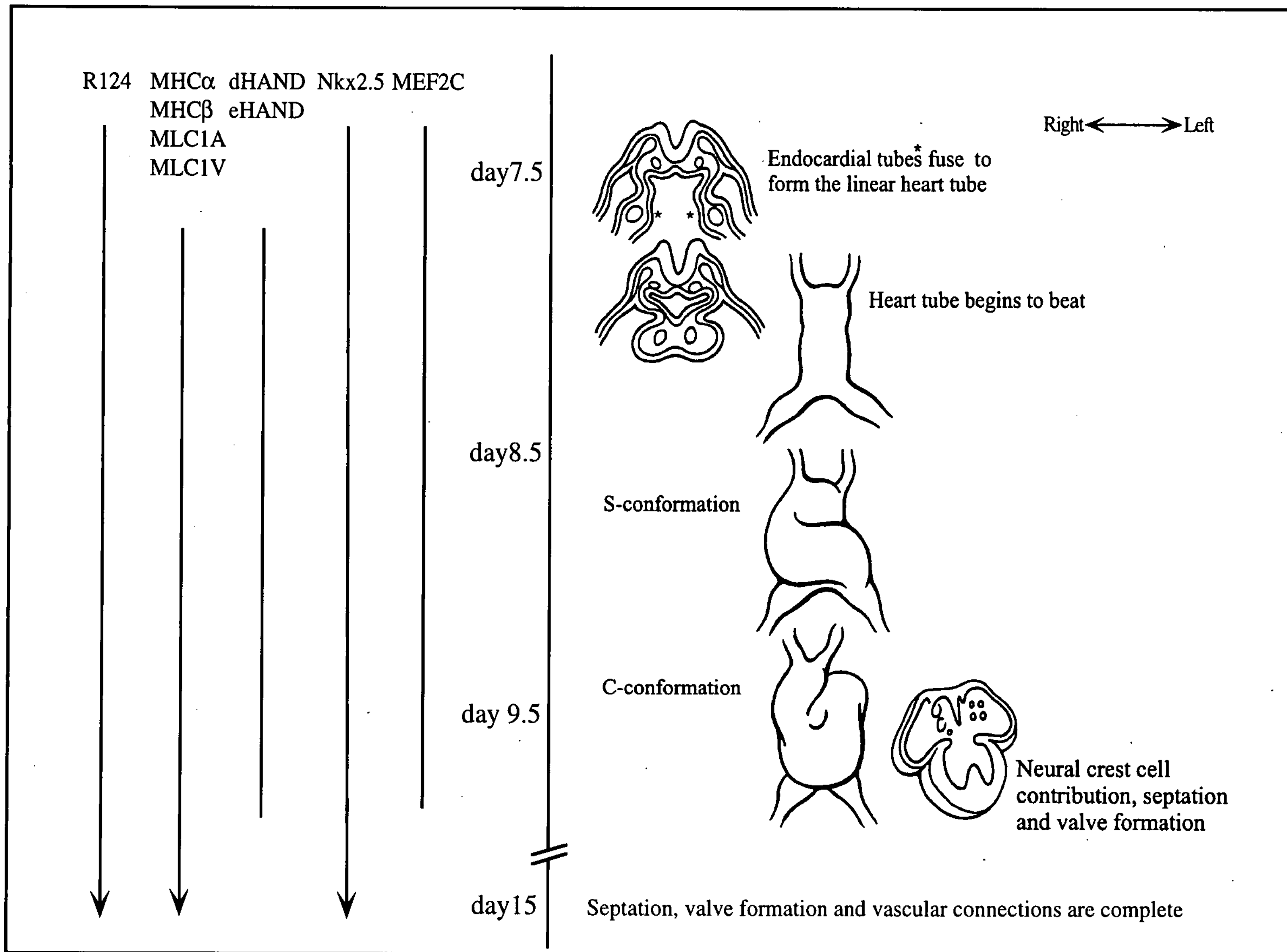


Figure 1.1: Mammalian Heart Development
A simplified schematic to highlight the major morphogenetic changes during heart development and documentation of earliest expressed genetic markers in the cardiac lineage including detection of reporter gene expression in embryos heterozygous for the R124 gene-trap integration.

forming the C-shaped conformation. Therefore by 9 d.p.c. the heart tube is in a conformation to generate the four chambers observed in the adult heart.

More recently asymmetric expression of molecular markers such as *nodal* (Collignon *et al.*, 1996), *lefty* (Meno *et al.*, 1996), *HNF3β* (Echelard *et al.*, 1993), *Shh* (Johnson *et al.*, 1994), *cSnR* (Isaac *et al.*, 1997), *e-HAND/Hxt/Th1* (Cross *et al.*, 1995; Cserjesi *et al.*, 1995) and *d-HAND/Hed/Th2* (Hollenberg *et al.*, 1995; Srivastava *et al.*, 1995) have been detected before the linear heart tube loops. In addition the asymmetric distribution of extracellular matrix components locally around the cardiac region and within the linear heart tube, provided by the cardiac jelly (Yost, 1992) e.g. fibronectin (Linask *et al.*, 1986; Icardo *et al.*, 1983), flectin (Tsuda *et al.*, 1996) and actin (Itasaki *et al.*, 1991), before and during looping of the heart have been demonstrated. These examples of asymmetric markers have accumulated from studies on various model organisms (mouse, chick, xenopus, zebrafish) and have been used to propose a genetic pathway that would lead to looping (Figure 1.2), but here I will concentrate mainly on the mouse. The asymmetric expression of markers would suggest that looping of the heart is an active process regulated by expression of specific gene products in contrast to the interpretation of how the heart loops from anatomical studies. The rightward-looping of the linear heart tube is a process conserved in all vertebrates studied to date i.e. man, pig, sheep, cat, rat, mouse, frog, fish and chick. Recently, Biben *et al.*, 1997 have documented that during mouse heart development an initial leftward displacement of the caudal/posterior end of the heart tube is observed before the rightward looping is initiated, this is also conserved in other vertebrates.

nodal and *lefty*, members of the transforming growth factorβ (TGFβ) superfamily, are expressed in the lateral plate mesoderm to the left of the linear heart tube prior to looping suggesting that the heart receives signals from this region which influence looping of the heart. Ectopic expression of *nodal* in chick and frogs leads to abnormal

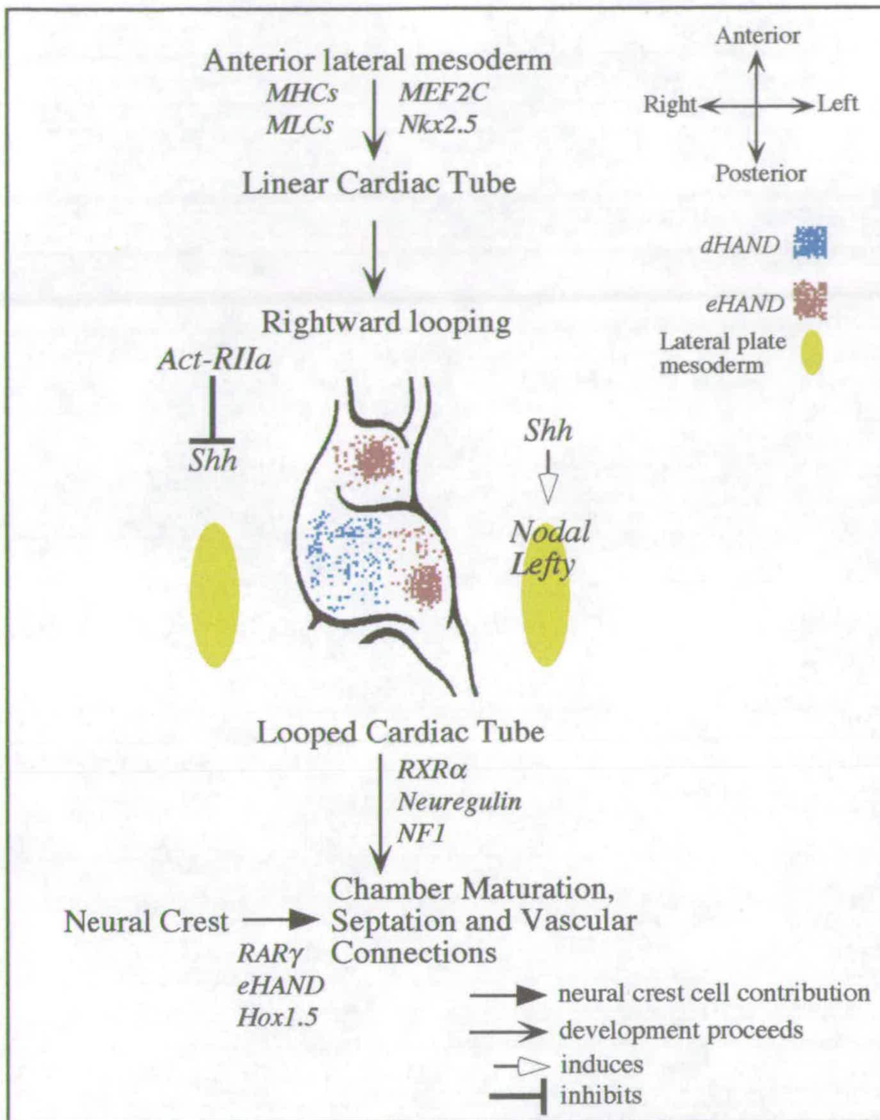


Figure 1.2: A genetic pathway determining asymmetry of the heart Expression of the HAND genes is correlated with the presumptive right and left ventricles. The pathway leading to the asymmetric expression of *Shh* has been determined from experiments in chick. No asymmetry of *Shh* expression has been detected in mouse.

looping of the heart, which implicates *nodal* directly in the right/left axis of the heart (Levin *et al.*, 1997).

d-HAND and *e-HAND* are the first helix-loop-helix transcription factors to be identified that are expressed in the cardiac lineage (Srivastava *et al.*, 1995; Cserjesi *et al.*, 1995). In the mouse, both genes are initially expressed throughout the linear heart tube but become restricted to the future right and left ventricle regions respectively (Cserjesi *et al.*, 1995). *d-HAND* null embryos fail to complete the process of looping and die by the 10th day of gestation from heart failure (Srivastava *et al.*, 1997). Analysis of the null mice shows the region destined to become the right ventricle is missing, suggesting that *d-HAND* regulates the proliferation and specificity of cardiogenic precursors destined to form the right ventricle. Evidence has also been documented that shows *e-HAND* maybe involved in the initial leftward displacement of the cardiac tube and determining the future left ventricle cardiomyocytes (Biben and Harvey, 1997).

In chick asymmetric expression of the *HAND* genes is not observed (Srivastava *et al.*, 1995). However, the involvement of the *HAND* genes in looping of the heart in chick is shown by disruption of looping by antisense oligonucleotides designed to inhibit the expression of the *HAND* genes (Srivastava *et al.*, 1995).

Complete reversal of the right/left axis is observed in 100% of mouse mutants lacking the normal function of *inv (situs inversus)* and in 50% of mutants lacking the normal function of*iv (inversus viscerum)* (Collignon *et al.*, 1996; Layton, 1976; Yokoyama *et al.*, 1993). Expression of *nodal*, *lefty* and the *HAND* genes is reversed and the heart loops to the left in these mutants (Icardo *et al.*, 1991; Lowe *et al.*, 1996). This suggests that the looping of the heart is interlinked with the symmetry of the body axis.

Other genes such as *Mef2C* and *Nkx2.5* also affect looping of the heart but do not show asymmetric patterns of expression during heart development. *Mef2C*

(Edmondson *et al.*, 1994; Molkentin *et al.*, 1995) and *Nkx 2.5/CSX* (DeRuiter *et al.*, 1992; Komuro *et al.*, 1993; Lints *et al.*, 1993) are expressed bilaterally in the early cardiac progenitors before the linear heart tube forms implicating them in commitment to and maintenance of a cardiac specific lineage.

Interestingly both genes have *Drosophila* homologues, *Dmef 2* and *tinman* respectively that display high conservation in sequence and expression patterns. *Dmef 2* loss-of-function mutants block differentiation of all muscle cell types in the embryo, however normal myoblasts are present which give rise to the dorsal vessel (heart equivalent in *Drosophila*) but contractile protein genes are not expressed. *Drosophila tinman* loss-of-function mutants have no hearts or visceral muscles and many of the somatic body wall muscles also appear to develop abnormally (Bodmer, 1993).

MEF 2C, a member of the MADS box containing transcription factors (Pollock *et al.*, 1991) is first expressed at 7.5 d.p.c. in cardiac mesodermal cells that give rise to the primitive heart tube and thereafter is expressed throughout the primitive heart tube. However the phenotype observed in embryos null for *MEF2C* affects the looping morphogenesis of the linear heart tube (Lin *et al.*, 1997). In addition the right ventricle does not seem to form in these mutants and is correlated with down regulation of dHAND and other right ventricle markers. *Nkx 2.5*, also known as *CSX,* is expressed at 8.5 d.p.c. in myocardial progenitor cells and continues to be expressed throughout the embryonic and adult heart at later stages (Lints *et al.*, 1993). Homozygous *Nkx 2.5* loss-of-function mouse mutants form a linear heart tube but looping morphogenesis of the heart is not initiated. In addition, expression of *eHAND* and myosin light chain 2V (*MLC 2V*) is not observed in mutant hearts implicating *Nkx 2.5* in left ventricle specification (Lyons *et al.*, 1995; Biben *et al.*, 1997). This data suggest that *Mef2C* and *Nkx2.5* lie upstream of the *HAND* genes in a genetic pathway determining chamber specificity and looping. The lack of severity of *MEF2C* and *Nkx 2.5* mutants, as would be predicted from their early expression, can be explained by

the overlapping expression and possible functional redundancy between the family members of *MEF2* and *Nkx* genes during development (Lin *et al.*, 1997; Lints *et al.*, 1993).

Thus, determination of regional specificity of the linear heart tube, demarcated by the expression of gene markers, is an essential part of the mechanism involved in the looping of the heart tube. Regional specification of the linear heart tube into presumptive chambers may be sufficient for the basis of the genetic pathways which determine the cell fate of precursor cardiomyocytes. The regulated morphogenetic behaviour of cardiomyocytes i.e. activation of cell growth, shape change and division, in specific regions of the linear heart tube would lead to asymmetry of the heart.

1.3 **Chamber Specification**

As has been stated above, a number of markers have been identified that are implicated in looping of the cardiac tube and are also involved in regional specification of the right/left ventricles. In addition to these, genes that encode structural proteins allow early distinction between different cell types of the presumptive atria and ventricles of the heart before actual physical regionalisation (Lyons *et al.*, 1990). Many of these are also expressed later in skeletal muscle except for cardiac troponin I which is exclusively expressed in the cardiac cells (Schiaffino *et al.*, 1990). Expression of these structural protein encoding genes in skeletal muscle is regulated by the helix-loop-helix transcription factors *Myo D*, *myf* and myogenin, but no homologues of these genes have been found in cardiac muscle. This makes it difficult to analyse the regulation of the myosin proteins in cardiac muscle.

Myosin heavy chain (*MHC*) genes and myosin light chain (*MLC*) genes are coexpressed throughout the myocardium of the primitive linear heart tube, but by 8 d.p.c. expression of *MHC*α becomes increasingly restricted to cells that will contribute to the atria, while *MHC*β is restricted to the ventricle. As development

proceeds, late 17.5 d.p.c., expression of *MLC1a* and *MLC1v* also become restricted to the atria and ventricles respectively. This variation in expression patterns is indicative of chamber specific cardiomyocytes arising as cardiogenesis proceeds (Lyons *et al.*, 1990). The complete partitioning of these chamber specific markers occurs after morphologically distinct atrial and ventricle regions can be identified. In contrast, the regional specific expression of the *MLC2a* and *MLC2v* has been shown to become restricted, to the presumptive atria and ventricles respectively, earlier during cardiogenesis than any other cardiac chamber specific marker. *MLC2v* becomes restricted to the presumptive ventricle at 8 d.p.c. in the linear heart tube. Following this the expression of *MLC2a* is down regulated in the ventricular region by 9 d.p.c. and becomes restricted to the atrial region. The ventricular specificity of *MLC2v* gene has been localised to a 250bp cis regulatory promoter sequence that contains positive (*HF-1a*, *HF-1b*, *HF-2*) and negative (*HF-3* and *E-box*) regulatory sites (O'Brien *et al.*, 1993; Lee *et al.*, 1994; Zhu *et al.*, 1991). The 250bp sequence can be replaced by the *HF1* binding site (*HF-1a* and *HF-1b*/*Mef-2* sites) to confer ventricular specific expression of β-galactosidase reporter transgene during early cardiogenesis. However the expression of the transgene becomes restricted to the right ventricle/conotruncal region as heart development proceeds unlike the expression of the endogenous *MLC2v* which is expressed in both ventricular chambers suggesting that sequences for complete spatial regulation are not present within the 250bp sequence (Ross *et al.*, 1996).

The modular nature of promoters in eukaryotes can potentially be used to study regionalisation of the heart by directing for example reporter gene expression to identify distinct regions within the physically obvious compartments of the heart. Recently, promoter sequences from the MLC-3F gene, which is normally expressed in more abundantly in atria than ventricles, have been used to direct reporter gene expression to distinct regions of the developing heart eg. embryonic right atrium,

atrioventricular canal and left ventricle (Franco *et al.*, 1997). This data adds support to expression within distinct regions of the heart being regulated by specific sets of transcription factors and thus particular promoter sequences direct expression of genes to specific regions of the heart.

1.4 Chamber Maturation and Septation

As the conformational changes are observed externally between 8 and 9.5 d.p.c., important changes are occurring within the heart to compartmentalise the heart into four chambers (Vuillemin *et al.*, 1989a; Vuillemin *et al.*, 1989b). These morphogenetic changes include: the partitioning of the common atria into right and left by the formation of the septum primum and septum secundum; the absorption of the sinus venosus (which becomes the vena cava) into the wall of the right atrium and of the pulmonary veins into the left atrium; the division of the atrioventricular canal into two by the generation of the atrioventricular valves from endocardial cushions; the merging of the bulbus into the prospective right ventricle; the partitioning of the right and left ventricles; the longitudinal division of the conotruncus/bulbus into the cardiac outflow tracts (aorta and pulmonary artery) via a complex spiral interaction between the conus cordis and truncus arteriosus; small regions of the dorsal myocardium breakdown between the inflow and outflow tracts to create the transverse pericardial sinus; and the histogenic differentiation (thickening and trabeculation) of the cardiac wall, including the development of the tricuspid and bicuspid valves (Huang *et al.*, 1995; Meyer *et al.*, 1995). The internal morphogenetic changes of the heart involve extensive remodelling that is achieved through selective growth of the epicardial compact zone and cell death zones of the myocardium (Manasek, 1969; Okamata *et al.*, 1975). The developing heart being sculpted by cellular morphogenesis is also under the influence of the hemodynamic forces generated from the circulatory system (Bishop, 1995).

After the specification of the primitive ventricle, the ventricular wall consists of three zones, the endocardium, the myocardium and the epicardium. The external surface of the myocardial wall (known as the compact zone) undergoes expansion to form the thickened ventricle wall (spongy layer), the trabeculation within the chamber and contributes to the formation of the interventricular septum between the future right and left ventricles. Specialised trabeculae develop into papillary muscles which anchor the atrioventricular valves.

1.5 **Neural Crest Cell Contribution to the Heart**

Neural crest cells arise from the neuroepithelium. They are characterised by a epithelial to mesenchymal transformation as they emigrate from the junction between the neuroepithelium and epidermal ectoderm (Nichols, 1987). The first migratory wave of neural crest cells emerge at the neural plate stage during embryogenesis this is followed by several waves of emigration in a cranio-caudal direction (Morriss-Kay *et al*, 1993). Each wave of neural crest is referred to as the anatomical region from which it emerges eg. cranial neural crest are derived from the midbrain/rostral hindbrain region. Neural crest cells migrate to specific sites in the embryo and differentiate into a variety of cell types which contribute to the development of many organs and tissues eg. heart, craniofacial bone and cartilage, ganglia, melanocytes, neurons etc., [this pluripotentiality has earned them the title of the "fourth germ layer" in some schools of biology] (Fukiishi *et al*, 1992; Levi *et al*, 1987; Morriss-Kay *et al*, 1991; Noden, 1983). Neural crest cells derived from the caudal hindbrain neuroepithelium (parallel to the occipital somites 1-7) contribute to the cardiac ganglia, the cardiac mesenchyme and septation of the outflow tract (Fukiishi *et al.*, 1992; Kirby *et al.*, 1983a; Kirby *et al.* 1983b). Deletion of neural crest that lie adjacent to occipital somites 1-3 produce malformations of the heart which include abnormalities of the aorticopulmonary septum, transposition of the major heart vessels, high ventricular septal defects, single

outflow vessels emerging from the right ventricle or over the ventricular septum leading to abnormal outflow of blood from the heart (Kirby *et al.*, 1983; Stewart *et al.*, 1986). Deletion of more caudal neural crest from the level of somites 10-20 results in the absence of sympathetic cardiac nerve innervation (Kirby *et al.*, 1984). These heart defects are associated with under development of the thymus and parathyroid glands, both of which also receive contribution from neural crest cells. The specificity of defects induced by deletion of different neural crest populations highlights the importance of the spatially and temporally regulated contribution of neural crest cells in development (Beeson *et al.*, 1986). Many of these cardiac defects are also observed in vitamin A deficiency syndrome defects and in excess RA embryopathies (Dickman *et al.*, 1997; Wilson *et al.*, 1949; Wilson *et al.*, 1953; Lammer *et al.*; 1985, Morriss-Kay, 1993). Human syndromes such as DiGeorge Syndrome also display a range of similar cardiac defects some of which can be attributed to abnormal neural crest migration and differentiation.

1.6 **Function of the Heart**

The heart is one of the few organs that becomes functional as it forms emphasising the critical role it performs to pump blood, nutrients and waste through the early embryo and placenta. On formation of the single linear heart tube, the heart starts to beat initially in a peristaltic motion which is initiated in the pacemaker region of the primitive ventricle, and then a regular beating pattern is achieved (DeHaan, 1963; Kamino *et al.*, 1981, Moorman *et al.*, 1997). This beating is an intrinsic function of the cardiomyocytes and can be modulated by parasympathetic and sympathetic neural inputs (not discussed here). The pacemaker activity is taken over by a cluster of cells known as the sinoatrial node which is situated in the right atrium in the right venous valve (Viragh *et al.*, 1977a; Viragh *et al.*, 1977b; Viragh *et al.*, 1980; Viragh *et al.*, 1982). Soon after the establishment of the sinoatrial node pacemaker the formation of

the atrioventricular valves is initiated, this coincides with the development of a second pacemaker centre the atrioventricular node. The development of the atrioventricular node is accompanied by the appearance of the "bundle of His" a bundle of specialised conducting cells extended into the right and left ventricles. Signals from the sinoatrial node are conveyed, at a pace slower than conductance through the atria, via nodal fibres to the atrioventricular node which passes the signal rapidly to the ventricles via the "bundle of His". The differential pace of the signal from the sinoatrial node to the atrioventricular node allows for the atria to contract earlier than the ventricles and thus blood is pumped into the ventricles before the ventricles contract. Impulses from the pacemakers are relayed through the atria and ventricles via specialised conducting purkinje fibres and via the numerous intercellular connections between cardiomyocytes including gap junctions (Flucher *et al.*, 1996). Normal heart function can be characterised by measurements of pressure, volume, electroactivity and heart sound during the cardiac cycle which consists of a period diastole (relaxation) when the chambers of the heart fill with blood followed by systole (contraction) (Figure1.3) (Colucci, 1994; Dyson *et al.*, 1995; Rockman *et al.*, 1991, 1993, 1994; Tsutsui *et al.*, 1993; Jones *et al.*, 1996). The myofibril contracting apparatus of the heart myocardial cells consists of force generating proteins myosin (thick filament), actin and a complex of regulatory proteins namely tropomyosin, troponin C, tropininI and troponin T (thin filaments). The structure of the myofilaments is highly organised in a regular array of thick and thin filaments that is observed as the characteristic striated structure of cardiac muscle (Ishikawa *et al.*, 1975). The ultrastructure of heart muscle i.e. composition and structure of the myofilaments; the activity of the contractile proteins and the structure of intercellular connections have been documented widely in the literature (Colucci, 1994) but are beyond the scope of this introduction.

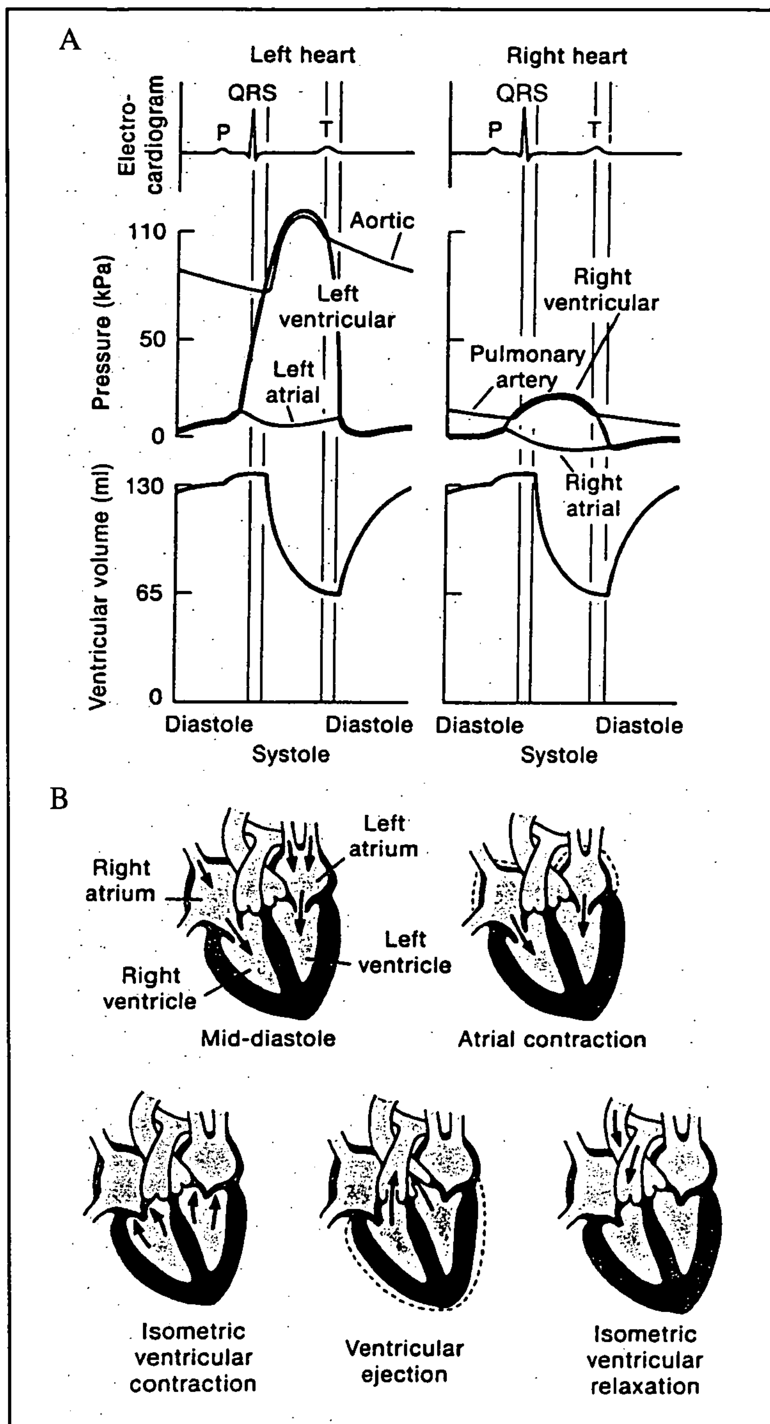


Figure 1.3: Characteristic Normal Cardiac Cycle

A. Changes in pressure, volume and electrical activity in the left and right ventricles during systole (contracting) and diastole (relaxing) phases of a single cardiac cycle. Electrocardiogram shows the normal profile of electrical activity from the heart ventricles consisting of: P wave, spread of repolarization; QRS wave, result of depolarization; T wave, repolarization. B. The flow of blood (arrows) during these phases.

1.7 Changes in the Heart at Birth

The external appearance of the beating heart at late stages of gestation is the same as the adult heart. Distinct atrial and ventricular chambers connected to the great blood vessels can be identified. Internally the development of the tricuspid and bicuspid valves separating the atrial and ventricle chambers is complete, however flow of blood through the right side of the heart is not completely partitioned from the left (DeHaan, 1965).

Before birth oxygen and nutrients are supplied via the placenta. Thus oxygen and nutrient rich blood reaches the fetus via venous circulation into the right atrium and right ventricle. The right atrium is connected to the left ventricle via the oval foramen, this allows the blood to be shunted to the left and be pumped around the body. Blood from the right ventricle is pumped into the aorta via the ductus arteriosus. The ductus arteriosus is an arterial connection between the pulmonary artery and the aorta that diverts blood away from pulmonary circulation around the body (Coceani *et al*, 1988). After birth it is essential to close the oval foramen and the ductus arteriosus to redirect the flow of blood via the pulmonary circuit and prevent shunting of blood from the left to the right side of the heart. Abnormalities arise if these structures do not close (discussed in Section 1.8).

Expansion of the lungs after birth allows blood to flow, with little resistance, from the right ventricle into the lungs via the pulmonary artery to be oxygenated and passed into the left atria via the pulmonary vein and finally into the left ventricle which pumps the blood throughout the body. Change in the direction of blood circulation after birth leads to reduced pressure in the right ventricle and atrium while the reverse is true for the left side of the heart. This causes the septum primum, a thin membranous structure, to be pressed against the septum secundum resulting in the closure of the oval foramen (Smolich, 1995; Teitel, 1988). In man these septa fuse permanently in the year following birth. The ductus arteriosus is a large muscular artery with layers of

immature smooth muscle separated by layers of elastin. The permanent closure of the ductus arteriosus is initiated by constriction of the vessel followed by fragmentation of the elastin layers resulting in release of endothelial cells into the lumen of the vessel. Subsequently, extracellular matrix components accumulate in the lumen followed by invasion of the lumen by smooth muscle cells to form thickened intimal cushions this results in the obstruction of the ductus arteriosus. The final stage of closure involves the degradation of the ductus arteriosus into a fibrous cord referred to as the ligamenum arteriosus (Coceani *et al.*, 1988). The closure of the ductus arteriosus is initiated within 3 hours and is complete after about 12 hours after birth in the mouse. The morphological changes involved in the closure of the ductus arteriosus in mouse closely resemble those observed in larger model organisms (sheep, dog, rat) and in man (permanent closure of the ductus arteriosus takes 1 to 3 months). The direction of closure, from the pulmonary to the aortic end, is also conserved.

Recently prostaglandin synthesis has been implicated in the closure of the ductus arteriosus (Nguyen *et al.*, 1997). Inhibition of prostanoid synthesis using pharmacological inhibitors can induce premature closure of the ductus arteriosus. Mice lacking the prostaglandin E2EP4 receptor die after birth from a patent ductus arteriosus. This suggests that a reduction in the level of prostaglandins circulating after birth is detected by the E2EP4 receptor which initiates the closure of the ductus arteriosus.

After birth the hemodynamic forces of the circulation system change such that the left of the heart dominates the right, a reversal of forces experienced by the heart before birth. This results in the cardiomyocytes of the left ventricle becoming larger than the right ventricle cardiomyocytes. The sum effect is that the left ventricle becomes much thicker and muscular than the right via hypertrophy. This adaptive change is not immediate but becomes apparent after 1 to 2 weeks after birth in the mouse (Bishop, 1995).

1.8 Congenital Heart Disease

Congenital cardiovascular malformations account for about twenty percent of all congenital defects observed in live born infants. Examples of cardiovascular defects affecting every step of cardiogenesis have been documented but defects affecting septation are most common. Incomplete partitioning of the right and left atria leads to shunting of blood from the left to the right causing enlargement of the right ventricle and pulmonary trunk. Ventricular septal defects arise due to incomplete partitioning of the right left ventricles leading to serious shunting of blood from left to right causing pulmonary hypertension. Other defects include transposition of the great vessels, tricuspid stenosis, persistent truncus arteriosus and the double outlet left ventricle malformation all of which involve abnormal blood flow. Hemodynamic forces caused by the flow of blood are critical and aberration of blood flow due to defects mentioned above lead to a syndrome of defects which often result in the hypoplasia of the heart walls. Many of the defects mentioned arise due to aberrant neural crest contribution and an abnormal level of programmed cell death. Programmed cell death is normally involved in fusion of the primitive heart primordia to give rise to the primitive heart tube and modelling of the ventricular septum, atrioventricular valve formation and outflow tract. A majority of the defects described above can be detected either *in utero* or at the neonatal stage using echocardiography and can be corrected by modern surgical techniques. Other defects which are observed to a lesser extent in clinical cases effect the looping of the heart and are often embryonic lethal.

This full spectrum of clinical heart defects and additional defects affecting commitment to the cardiac lineage, chamber specification, looping of the primitive heart tube have been observed in various mouse models of congenital disease (Table1.1). These mouse models will prove very important in dissecting the genetics of how these defects arise .

1.9 Effects of Retinoic Acid on Heart Development

All-trans retinoic acid (RA), a biologically active derivative of vitamin A, has been shown to play an important role in vertebrate pattern formation and cell differentiation (Conlon, 1995; Morriss-Kay, 1991; Papalopulu *et al.*, 1991, Tabin, 1991). During embryogenesis maternal RA excess and Vitamin A deficiency (VAD) are teratogenic and cause a syndrome of congenital malformations. Tissues affected include neural crest derivatives, the axial skeleton, the limbs, the neural tube, the brain and the heart. The defects induced by RA are dose and stage specific (Lammer *et al.*, 1985; Morriss-Kay, 1993; Webster *et al.*, 1986; Wilson *et al.*, 1949; Wilson *et al.*, 1953). *In vitro* RA can induce the differentiation of embryonic carcinoma (EC) cells and embryonic stem (ES) cells through activation of specific genes (Simeone *et al.*, 1991). RA transcriptional regulation is mediated through complexes of RA with nuclear retinoic acid (RARs) and retinoid X receptor (RXRs) heterodimers. These complexes bind to specific DNA regulatory sequences known as retinoic acid response elements (RAREs) in a ligand dependent manner to regulate gene expression. Diversity of RAR/RXR complexes is generated through the existence of three different subtypes (α, β and γ) of each receptor which are extensively spliced into many different isoforms. Each isoform is expressed in a spatially restricted pattern during embryogenesis. No *RAR* isoform or subtype is exclusively expressed in any tissue (Chambon, 1994; Manglesdorf *et al.*, 1994).

Defects arising from an excess or deficiency of RA during embryogenesis in the heart include hypoplasia of the ventricle walls, transposition of the great arteries and septal defects. The septal defects arise due to spatial and temporal disruption of cardiac neural crest cell migration. Cardiac neural crest cells normally contribute to the septation of the heart and to the outflow tract. Direct evidence implicating RA in cardiogenesis comes from genetic studies in which RA receptors have been disrupted using homologous recombination in ES cells. Animals homozygous for the majority of

single RAR knockouts are normal or display only subtle abnormalities supporting the hypothesis that there is functional redundancy between the different RARs (Mendelsohn *et al.*, 1994). However, double knock-out mutants e.g. generated from intercrossing single *RAR* knock-outs mutants display abnormalities which affect the heart, including myocardial deficiency, persistent atrioventricular septal defects, outflow tract defects, persistant atrioventricular canal and abnormal aortic arch patterning (Kastner *et al.*, 1994; Lohnes *et al.*, 1994; Mendelsohn *et al.*, 1994). A majority of these abnormalities are observed in the fetal VAD syndrome (Wilson *et al.*, 1953). Further evidence for the role of RA in cardiogenesis comes from a study in which the *RXRα* was targeted to generate a loss-of-function mutation. Homozygous *RXRα* null animals die between 13.5 and 16.5 d.p.c. and the main defect observed in these fetuses is hypoplastic development of the ventricular chambers of the heart with concurrent defects in the ventricular septation (Dyson *et al.*, 1995; Kastner *et al.*, 1994; Sucov *et al.*, 1994). Again these defects are identical to a subset of defects observed due to VAD and do not affect cardiac neural crest cell derived structures.

1.10 *In Vitro* Early Cardiogenesis Model

Embryonic stem (ES) cells are pluripotent stem cells derived from the inner cell mass of blastocyst stage mouse embryos (for further details see 1.14). ES cells can be differentiated *in vitro* to give rise to cell types derived from all three germ layers: endodermal, mesodermal and ectodermal (Keller *et al.*, 1995). ES cells allowed to aggregate in suspension or in hanging drops form small embryo like structures called embryoid bodies. These embryoid bodies spontaneously differentiate, in the absence of DIA (Differentiation Inhibiting Activity)/LIF (Leukaemia Inhibiting Factor), and recapitulate several aspects of early embryogenesis which include hematopoiesis (Lindenaum *et al.*, 1990), cardiogenesis (Doetschman *et al.*, 1985; Robbins *et al.*, 1990; Sanchez *et al.*, 1991), neurogenesis (Strübing et al, 1995) and angiogenesis

(Risau *et al.*, 1988). Some differentiated cell lineages can be distinguished by their morphology - one of the most striking and obvious is the presence of beating cells. A number of studies using physiological and molecular markers have shown that these beating cells are cardiomyocytes (Wobus *et al.*, 1991). The temporal expression of cardiac markers e.g. MHCs and MLCs has been shown by RT-PCR to be consistent with the *in vivo* expression pattern observed during early cardiogenesis (Robbins *et al.*, 1990; Maltsev *et al.*, 1994). Further evidence comes from physiological patch clamp experiments which show the expression of cardiac specific ion channels, adrenoceptors and cholinoceptors in beating cells (Wobus *et al.*, 1991). In addition beating cells have been shown to differentiate into regional specific cardiac cell types i.e. ventricular or atrial, by the detection of chamber specific gene expression e.g. *MHC*s, *MLC1v* and *MLC2v* (Miller-Hance *et al.*, 1993; Kubalak *et al.*, 1994).

This evidence suggests that cardiac differentiation of ES cells in embryoid bodies recapitulates early cardiogenesis and therefore could serve as an *in vitro* model to study cardiogenesis at the molecular level. Genetic and pharmacological manipulations of ES cells before aggregating cells to form embryoid bodies would allow dissection of the molecular pathways determining cardiac fate. This approach is exemplified by the *in vitro* differentiation of β1 integrin null-ES cells (Fassler *et al.*, 1996). β1-null ES cell derived cardiomyocytes exhibit delayed expression of cardiac specific genes and incomplete sacromeric architecture resulting in generalised impairment of cardiogenesis. *In vivo* as *in vitro*, β1-null ES cells undergo limited cardiac differentiation (Fassler *et al.*, 1995). Chimaeric mice generated from β1-null ES cells have revealed that mutant cell contribution in hearts is not maintained indicating the importance of β1-integrin in differentiation and maintenance of cardiomyocytes.

1.11 **Identification of Novel Genes**

The generation of a mutation is very valuable in determining the function of a gene and is the principle on which "classical" genetics is based on to identify genes involved in development and disease. Mutants in plants, bacteria, worms, flies, zebrafish, mouse and humans can be generated by spontaneous, mutagen induced or insertional mutagenesis.

[Many spontaneous mutations linked to human disease states have led to the discovery of genes through linkage analysis, which have been later verified by directed mutagenesis of homologous genes in model organisms e.g. muscular dystrophy due to mutations in the human dystrophin gene and *mdx* gene in mouse (Chamberlain *et al.*, 1987; OMIM no.310200).]

1.12 **Random Mutagenesis**

Classical mutants were isolated by screening populations for spontaneous mutations, which include point mutations, insertions, deletions, translocations.

To increase the number of mutants available for study large scale saturation mutagenesis screens have been performed in *D.melanogaster* (Nusslein-Volhard *et al.*, 1980 and 1984); *C.elegans* (Horvitz *et al.*, 1991); *D.rerio* (Driever *et al.*, 1996; Haffter *et al.*, 1996; Streisinger *et al.*, 1981) using X-rays, which cause large chromosomal deletions, chemical mutagens such as ethylnitrosourea (ENU) which induces point mutations in the genome, or chlorambucil (CHL) which induces short chromosomal deletions. The mutants selected from these screens had obvious external morphological phenotypes. For example homeotic transformations in *Drosophila* or disruption of heart morphogenesis in zebrafish were some of the easily identifiable criteria by which mutants were isolated. Random mutagenesis requires breeding programs over multiple generations to segregate mutations, define allelism and identify recessive mutations. Thus, small model organisms with short life spans but large

offspring production are most practical for this type of approach. Both spontaneous and mutagen induced mutations have to be mapped using linkage analysis and then isolated by positional cloning techniques or by candidate gene approaches. Positional cloning and candidate approaches are suited for small well characterised genomes e.g. *Drosophila* (Collins, 1995). This approach is not practical for a model organism like mouse which has a relatively long life span, generates a small number of offspring per generation and has a large uncharacterised genome relative to *Drosophila*. Although many classical mouse mutants have been mapped by linkage analysis (Haldane *et al.*, 1915; Dietrich *et al.*, 1995). If saturation mutagenesis was carried out on a large scale it would require vast resources to manage and maintain mutant mouse breeding programs. This is exemplified by reports from Shedlovsky *et al.*, 1988 and Rinchik, 1991.

To increase the efficiency of cloning genes mutated by random mutagenesis, insertional mutagenesis has been developed in many organisms [plant (Koncz *et al.*, 1989); bacteria (Bellofatto *et al.*, 1984; Casabadan *et al.*, 1980); flies (Spralding *et al.*, 1995); zebrafish (Schier *et al.*, 1996); mouse, (Jaenisch, 1981; Gossler *et al.*, 1993)]. This technique relies on the introduction of a known piece of DNA, which integrates randomly, into the genome and thus 'tags' the site of integration at the molecular level. Some of the first examples of gene identification using insertional mutagenesis in eukaryotes comes from *Drosophila* P-element studies (Bier *et al.*, 1989; Engels, 1989; Spralding *et al.*, 1989; Wilson *et al.*, 1989). The P-element not only integrates into genes but also into regulatory regions this has led to the discovery of enhancer and promoter sequences (Bellen *et al.*, 1989). Insertional mutagenesis has evolved to what is known as entrapment technology in mouse ES cells, this is discussed in the context of gene-trapping below (Section 1.17) (Skarnes, 1990).

1.13 Identification of Conserved Genes

Isolation of gene sequences from model organisms e.g. *D.melangastor* and *C.elegans* has led to the identification of conserved gene family members between species. Many techniques involve performing homology screens at the genomic level using 'Zooblots' to identify evolutionarily conserved genes. A variation of this approach is the identification of gene family members between and within different species by screening with redundant probes at varying stringencies e.g. steroid/thyroid hormone (Evans, 1988; Green *et al.*, 1988). This approach has been superceded by the use of degenerate polymerase chain reaction (PCR) primers to amplify homologous sequences (Gould *et al.*, 1989; Wilks *et al.*, 1989). PCR based approaches can be employed at the genomic and mRNA level.

Analysis of mRNA allows the identification of transcriptionally active sequences that can be tissue or molecular pathway specific. This adds a dimension of gene regulation to identification of genes. Techniques such as 'differential display' (Liang *et al*, 1993) and 'subtractive hybridisation' (Harrison *et al.*, 1995) analyse mRNA populations isolated from different sources. The sources of mRNA can be different types of tissue or tissue treated for example with some agent to alter differentiation status. Both techniques involve elimination of common cDNAs between the different sources to reveal cDNAs that are unique to one source. In 'differential display' elimination is performed by running the different cDNA pools on polyacrylamide gels and looking for differentially expressed cDNAs. This procedure has to be repeated several times to confirm reproducibility before the differential cDNA is eluted and cloned. 'Subtractive hybridisation' as it suggests involves hybridisation of single stranded cDNA with mRNA or cDNA from a control reference source followed by active removal of the hybridised duplex species thus enriching the single stranded differentially expressed cDNAs. In differential display the elimination step has to be repeated several times to enrich the differential cDNA before cloning procedures.

More recently gene identification has reached a computer database dimension. Progress in genome sequencing projects has led to establishment of databases predicting potential gene encoding cistrons (Olson *et al.*, 1989; Pruitt, 1997). At the mRNA level sequencing of random cDNA synthesis from various tissue sources have generated expressed sequence tag (EST) databases (Adams *et al.*, 1995; Boguski *et al.*, 1993). The authenticity of sequences in these databases have to be confirmed by Southern and northern blot analysis. However these databases provide a useful resource for comparing and extending novel sequences without screening libraries.

1.14 **Embryonic Stem Cells**

The isolation of embryonic stem (ES) cells has revolutionised the ability to genetically manipulate the mouse genome and hence mouse mutational analysis. ES cells are derived from the inner cell mass of blastocyst stage mouse embryos and can be maintained in an undifferentiated state by culturing on fibroblast feeder cell layers or in the presence of DIA (differentiating inhibitory activity), also known as LIF (leukaemia inhibiting factor) (Brook *et al.*, 1997; Smith *et al.*, 1988; Williams *et al.*, 1988). *In vitro* ES cells can be induced under appropriate culturing conditions to give rise to cell types derived from all three germ layers (Section 1.10) (Evans *et al.*, 1981; Martin, 1981). *In vivo*, when ES cells are introduced into preimplantion embryos by morula aggregation or blastocyst injection they are incorporated into the normal development of the embryo and contribute to all tissues, including the germline, giving rise to chimaeric mice (Bradley *et al.*, 1984; Nagy *et al.*, 1993). By incorporating selectable markers into the targeting constructs and gene-trap vectors, ES cells can be screened *in vitro* for the desired genetic alteration before reintroduction into an embryo. Genetically altered ES cells are generally stable in culture and maintain their normal morphology and their pluripotency *in vivo* and *in vitro* (Gossler *et al.*, 1989; Robertson *et al.*, 1986; Thomson *et al.*, 1989).

1.15 **Directed Mutagenesis (or "Reverse" Genetics)**

Identification and isolation of genes from classical mutants and saturation screens has been most successful in *Drosophila* and *C.elegans* consequently many evolutionarily conserved genes have been identified and isolated by homology in other species eg. Hox genes (McGinnis *et al.*, 1992; Miklos *et al.*, 1996; Nusslein-Volhard, 1994). The development of homologous recombination in mouse ES cells has allowed vertebrate homologues to be 'targeted' to introduce (point and deletions) mutations to assess the functional homology of conserved genes (Capecchi, 1989).

This approach is very powerful in dissecting the function of known genes in the mouse and a number of mutant phenotypes have accumulated from targeted knock-out experiments which are proving very valuable in understanding mammalian developmental mechanisms and clinical disease conditions.

1.16 **Mouse Mutations Affecting Heart Development**

The development of the heart has been well described morphologically and more recently the molecular mechanisms are becoming apparent, through the identification of genes, which determine cardiac fate and regulate cardiac differentiation (Chien, 1993). Many genes that have been identified could not be predicted to be involved in heart development from their expression patterns. The generation of targeted mutations by homologous recombination in ES cells has uncovered specific functions of genes in the cardiac lineage. Conversely in some mutants the spectrum of defects is more restricted or less severe than would be predicted from the expression pattern this partially reflects the functional redundancy present between family members and may confuse mutational analysis in some cases.

A list of genes is provided which have been implicated in cardiogenesis (Table 1.1), and some selected genes are discussed below highlighting the advantages of mutational analysis to study development. The heart ventricle matures at 9.5 d.p.c. (Section 1.4)

Table1.1: Genes Implicated in Mouse Heart Development

Gene	Type of Mutation	Time of Death	Cardiac Phenotype	Other Abnormalities
Fibronectin (George *et al.*, 1993)	T, null	day 10-11p.c.	bilateral heart primordia do not fuse	shortened A-P axis, deformed neural tube, lack of mesoderm derivatives
α_4-integrin (Yang *et al.*, 1995)	T, null	Variable	epicardial dissolution	failure of chorioallantoic fusion
β_1-integrin (Fassler *et al.*, 1996)	T, null	day 8.5p.c.	impaired differentiation of cardiomyocytes	none reported
VCAM-1 (Kwee *et al.*, 1995)	T, null	Variable	epicardial dissolution, reduced myocardium	failure of chorioallantoic fusion, placental anomalies
Vinculin (Xu *et al.*, 1998)	T, null	day 10p.c.	dilated pericardiac cavity, heart walls hypoplastic leading to hypodynamic heart	failure of neural tube to close, forelimb bud development retarded
MHCα (Jones *et al.*, 1996)	T, null	day 11-12p.c.	defective myofibrillar and sacromeric organisation,	none reported
Neurcgulin typeI (Meyer *et al.*, 1997)	T, null	day 10.5p.c.	ventricular trabeculation lost	no sensory neurons in cranial ganglia
PDGF-B (Leveen, *et al.*, 1994)	T, null	Term	heart and vessel dilation	kidney, hematological defects
NF1 (Brannan *et al.*, 1994)	T, null	day 14p.c.	hypoplastic heart, septal defects, persistent truncus arteriosus	multiple organ defects, hyperplastic ganglia
Connexin43 (Reaume *et al.*, 1995)	T, null	Term	enlarged ventricles, outflow tract aberrations	none reported
gp130 (Yoshida *et al.*, 1996)	T, null	day 12.5-Term	hypoplastic ventricular myocardium	impaired hematopoiesis
IGF-IIr (Lau *et al.*; 1994)	T, null, imprinted	Term	hypertrophic myocardium	fetal organ overgrowth
RARs (Mendelsohn *et al.*, 1994)	T, null compound	variable	hypoplastic ventricular myocardium, outflow tract and great vessel aberrations	neural crest derivatives, genito-urinary and lower digestive tract abnormal,
RXRα (Sucov *et al.*, 1994)	T, null	day 13.5-16.5p.c.	hypoplastic ventricular myocardium, septal defects	ocular defects, chorioallantoic placenta defects
GATA4 Kuo *et al.*, 1997)	T, null	day 8.5-10.5	bilateral primorda do not fuse	folding of the embryo generally impaired
Hox1.5 (a3) (Chisaka *et al.*, 1991)	T, null	Term	hypertrophic atria, dysmorphogenic ventricles, enlarged blood vessels, stenosis	athymic, aparathyroid, throat abnormalities
MEF2C (Lin *et al.*, 1997)	T, null	day 10.5p.c.	right ventricle absent, abnormal heart looping	none reported
Nkx2.5 (Lyons *et al.*, 1995)	T, null	day 9-10p.c.	abnormal heart looping	none reported
d-HAND (Srivastava *et al.*, 1997)	T, null	day 10.5	right ventricle and trabeculation absent	aortic arch vessels absent
N-myc (Sawai *et al.*, 1993)	T, null	day 11p.c.	hypoplastic myocardium, septal defects	general hypoplasia
N-myc (Moens *et al.*, 1993)	T, hypo/null	day 14p.c.	hypoplastic ventricular myocardium	reduced lung branching
TEF-1 (Chen *et al.*, 1994)	GT, null	day 12p.c	hypoplastic myocardium, reduced trabeculae	dilated brain vesicles
WT-1 (Kreidberg *et al.*, 1993)	T, null	day 14p.c.	hypoplastic myocardium	renal agenesis
Sp (Franz *et al.*, 1989)	S, various alleles	day 13p.c.	persistent truncus, congestive heart failure	neural tube and neural crest defects

T, targeted mutation; S, spontaneous mutation; GT, gene-trap integration; hypo, hypomorphic allele.

and comparison of different mutant phenotypes allows speculation of genes that may regulate this developmental process. The earliest effect on ventricle determination is observed in *MEF2C* null mutants (Lin *et al.*, 1997) which display no morphologically distinct right ventricle and concomitantly no expression of right ventricle specific markers e.g. *d-HAND* (Cserjesi *et al.*, 1995). *d-HAND* null mutants also display the loss of the right ventricle and trabeculation is lost suggesting both these genes are in the same pathway that determines the fate of the presumptive right ventricle. In other mutants the ventricles are determined but the maturation of ventricles is impaired. The myocardium and the ventricular septum are hypoplastic in *RXRα* mutants (Sucov *et al.*, 1994) while *gp130* null mutants only display hypoplastic myocardium (Yoshida *et al.*, 1996). In contrast, trabeculation is lost in *neuregulin* type I null mutants (Meyer *et al.*, 1997). These observations suggest that although the compact zone contributes to the myocardium, ventricular septum and trabeculation the molecular pathway determining each process can be separated into distinct developmental steps. This demonstrates that although the compact zone is not proliferative in *RXRα* and *gp130* null mutants it is responsive to neuregulin signals from the endocardium to induce trabeculation. Phenocopies of ventricular defects in other mutants potentially infers their involvement in the same ventricular maturation pathway e.g. TEF-1 (Chen *et al.*, 1994).

1.17 **Gene-Trap Technology**

Gene-trapping has been used successfully to identify novel developmentally regulated genes (Baker *et al.*, 1997; Chowdhury *et al.*, 1997; Forrester *et al.*, 1996; Gasca *et al.*, 1995; Gossler *et al.*, 1989; Skarnes *et al.*, 1992; Skarnes *et al.*, 1995; Takeuchi *et al.*, 1995). In mice the gene-trap technology relies on the ease of genetically manipulating embryonic stem (ES) cells *in vitro* and their potential to be reintroduced into an embryo to contribute to all tissues of resultant chimaeras. Gene-trapping is based on the

principle that when a promoterless reporter gene is introduced into an ES cell genome it will be inactive unless it integrates into an actively expressed transcription unit. As a consequence of integration reporter sequences, under the regulation of endogenous cis-acting regulatory sequences, allow direct monitoring of the endogenous gene by visualizing the reporter gene expression in ES cells and in mouse embryo chimeras. The reporter sequences can be used to clone the endogenous gene from a fusion transcript using PCR based strategies. The trapped gene may also be disrupted and inactivated by the insertion thus potentially generating a mutant. The mutation can be transmitted through the germline to analyse the function of the trapped gene in homozygous animals.

1.18 **Enhancer Trap Vectors**

The prototype trapping vector, the enhancer trap, used originally for identifying novel developmentally regulated genes was based on trapping regulatory elements which specified the spatial and temporal expression of genes. Enhancer trap vectors consist of a minimal promoter (ideally a TATA box and a transcriptional initiation site) driving *lacZ* expression. Theoretically when this vector integrates into the genome adjacent to an enhancer the expression of *lacZ* reflects the specific temporal and spatial expression pattern specified by the enhancer. The problem with using this approach is that enhancers have the potential to regulate expression bidirectionally over kilobases of DNA , therefore identifying and cloning a gene by this strategy has proven to be difficult. Additionally, because enhancer trap vectors do not necessarily integrate into the transcriptional unit of a gene they are unlikely to be mutagenic. This type of vector (P-element based) has been used successfully in *Drosophila* to identify and clone a number of genes [Mlodzik *et al.*, 1990 (*seven-up*); Fasano *et al.*, 1991 (*teashirt*); Doe *et al.*, 1991 (*prospero*); Bier *et al.*, 1990 (*rhomboid*)] but has proved more difficult in mammals.

1.19 Gene-Trap Vectors

The second generation of trapping vectors were designed to allow direct identification of the trapped gene. This was achieved by designing vectors that would only function if they integrated into an active transcription unit to generate a fusion transcript. Various types of DNA constructs have been engineered to include a reporter gene and a selectable marker (Cui *et al.*, 1994). The selectable marker (for example the drug resistance gene neomycin), can be placed under its own promoter eg. phosphoglycerate kinase (PGK-1) to select for all integration events. The reporter gene most often used is the bacterial *lacZ* gene, encoding ß-galactosidase, and its position on the construct can be designed such that integrations into active transcription units can be selected (see below). The expression of the *lacZ* gene is easily detectable using a histochemical substrate, X-gal, and allows direct visualisation of the expression patterns of trapped genes in chimaeric embryos. LacZ expressing ES cells can also be detected and isolated by fluorescence-activated cell sorting (FACS), and can still be successfully transmitted through the germ line (Reddy *et al.*, 1992). The subcellular localisation of ß-galactosidase activity can reflect potentially the function of the trapped gene (Gossler *et al.*, 1989; Friedrich *et al.*, 1991; Skarnes *et al.*, 1992). Furthermore the level of ß-galactosidase can be quantified by colourmetric measurements.

The first type of gene-trap vector, the promoter trap (or exon trap), consists of reporter gene lacking a promoter. This construct requires an in frame integration into an exon of a gene for the reporter to be expressed. These sites of insertion have been confirmed by cloning sequences 5-primed to the insertion sites by inverse PCR (polymerase chain reaction) (von Melchner *et al.*, 1992).

A second type of vector termed gene-trap (or intron trap) contains a splice acceptor upstream of a promoterless reporter gene. This construct requires integration into an intron of an endogenous gene and becomes functional through the splicing of an endogenous gene exon to the vector sequences via the splice acceptor site. Due to the

high proportion of intronic sequences in the eukaryote genome, 12 fold more *lacZ* expressing transformants arise from gene-trap vectors than from promoter trap vectors (Gossler *et al.*, 1989).

Integration of the promoter-trap and gene-trap constructs results in the generation of fusion transcripts between endogenous and vector sequences leading to a disruption of the trapped gene. This disruption potentially may inactivate the gene and effectively be regarded as a gene knock-out. Animals homozygous for these integrations derived from germline transmission may display observable phenotypes which aid analysing the function of the trapped gene. The fusion transcript also allows cloning the endogenous gene sequences using RT-PCR based cloning strategies.

The promoter trap and gene-trap vectors have superceded the enhancer trap vector as strategies to identify and clone novel genes. A number of studies using promoter and gene-trap vectors have identified developmentally regulated genes with specific spatial and temporal expression patterns, as determined from the generation of chimaeric embryos from ES cell lines (Friedrich *et al.*, 1991; Skarnes *et al.*, 1992). Some trapped genes have been cloned using RT-PCR cloning strategies from trapped gene-*lacZ* fusion transcripts (Skarnes *et al.*, 1992). Sequence analysis has confirmed that the vectors can behave as predicted i.e. promoter trap vectors can insert inframe into exons and gene-trap vectors can insert into introns and splice to upstream exons via the *en-2* splice site. In addition, germline transmission of gene-trap ES cell lines has shown that trapped genes can be disrupted and can lead to mutant phenotypes (Friedrich *et al.*, 1991; Gossler *et al.*, 1989; Skarnes *et al.*, 1992).

1.20 **Modified Gene-Trap Vectors**

1.20.1 **Reading Frame-Independent Gene-Trap Vectors**

The second generation gene-trap vectors have been modified to allow optimal detection of trapped genes. The basic gene-trap vector (PT1) consists of a splice acceptor from

the *en-2* gene upstream of a promoterless *lacZ* reporter gene followed by a selectable marker gene, neomycin, under the regulation of a promoter (PGK). A variation of PT1 termed PT1-ATG has a translation start (ATG) incorporated at the beginning of the reporter gene to allow translation of the reporter even when the gene-trap vector inserts into the 5' untranslated region of a gene or out of frame with the endogenous ATG. This modification has been shown to increase the number of LacZ expressing clones by about three fold (Hill *et al.*, 1993). Other modifications to ensure that the reporter is expressed include the use of MoMuLV splice acceptor from the *env* gene which allows splicing in all three reading frames (Schuster-Gossler *et al.*, 1994). Alternatively the addition of an internal ribosomal entry site (IRES) ensures reading frame-independent translation of the reporter gene (Chowdhury *et al.*, 1997; Mountford *et al.*, 1995).

1.20.2 **Direct Selection Gene-Trap Vectors**

To allow direct selection of true gene-trap events, neomycin has been fused in frame to the 3' end of the *lacZ* gene to generate a novel reporter, termed ß-geo. ß-geo encodes the enzyme activities of both LacZ and neomycin and can be positively selected in the presence of G418. The activity of neomycin has been found to be much more sensitive than ß-gal activity therefore gene-traps can be selected which express LacZ at undetectable levels in undifferentiated ES cells, but are found to express the reporter at higher levels in differentiated cells (Friedrich *et al.*, 1991).

1.20.3 **Gene Product Selective-Trap Vectors**

Gene-trap vectors modified to preferentially trap genes encoding particular classes of genes. The secretory trap vector has been generated by incorporating a transmembrane domain, from the CD4 type I transmembrane protein gene, upstream of the βgeo gene in the basic gene-trap vector and has been shown to preferentially trap genes encoding cell surface and cell membrane spanning molecules (Skarnes *et al.*, 1995). This vector

is being modified further to accommodate integration into any of the three reading frames and also to generate an exon-trap version of the vector to isolate genes with no or small introns.

Another novel vector termed poly A-trap has been designed to identify genes that are not expressed in ES cells (Niwa *et al.*, 1993). The vector was generated from a standard gene-trap vector but the neomycin resistance gene, *neo* was placed under the regulation of a β-actin promoter, expressed in ES cells, and the poly A signal following the neomycin resistance gene, *neo* was removed. Thus expression of the neomycin resistance gene, *neo* would require to trap an endogenous poly A tail. This vector was modified by the addition of a splice donor site to its 3'end to allow splicing to exons downstream of the neomycin resistance gene, *neo*. Examples of integrations into the 3'end of genes have been confirmed by the isolation of novel endogenous sequences using 3' RACE (Yoshida *et al.*, 1995).

1.21 **Transfection of ES Cells**

Exogenous DNA can be introduced into ES cells by a number of methods for example calcium phosphate DNA co-precipitation, electroporation, and retroviral infection (Lovell-Badge, 1987). Each method has its advantages, however retroviral based gene-trap vectors are less likely to cause rearrangements of the host genome on insertion (Friedrich *et al.*, 1993). Studies have indicated a preference of these vectors to insert at the 5' end of genes thus generating transcripts with a small contribution from endogenous gene sequences (Friedrich *et al.*, 1991; Sheridan *et al.*, 1990). In contrast, vectors electroporated into cells often integrate as tandem insertions with local rearrangements or small deletions. No preference for integration into the 5' has been observed for these vectors resulting in a larger range of fusion transcript sizes, which aids the cloning procedure (Chowdhury *et al.*, 1997, Gossler *et al.*, 1986; Robertson *et al.*, 1986).

1.22 Modified Gene-Trap Screens

Gene-trap screens have been designed to select for developmentally regulated genes. For example a recent screen has been conducted to select genes expressed in a restricted pattern in 8.5 d.p.c. mouse embryos (Wurst *et al.*, 1995). More recently screens have been adapted by the addition of *in vitro* selective steps. Changes in reporter gene expression in response to RA-exposure have been used to isolate genes which are differentially regulated during development (Forrester *et al.*, 1996). Another approach has taken advantage of *in vitro* ES differentiation protocols to screen cells with gene-trap integrations in particular cell lineages (Baker *et al.*, 1997).

Further lineage specific gene-trap screens are being developed by Forrester *et al.* (unpublished) by introducing *in vitro* differentiation steps in combination with defined growth factors to promote gene-trap ES cell line differentiation down particular lineages that are morphologically distinct cell types eg cardiac muscle, haematopoetic lineages, neurons, endodermal cells (Doetschman *et al.*,1985; Risau *et al.*, 1988).

1.23 Trapping in Practice

An attempt has been made to catalogue the known and novel genes that have been isolated from gene-trap screens carried out over the last 5 years from different laboratories (Table 1.2A and B). This exemplifies the working potential of the gene-trap approach in its infancy. Although examples of gene-trap integrations can be sited which follow the predicted theory (*jumonji*, Takeuchi *et al.*, 1995), some exceptions have accumulated (*cordon-bleu,* Gasca *et al.*, 1995). These exceptions have important implications on the design of future gene-trap vectors, gene-trap prescreens/screens and the analysis of gene-trap integrations. It is impressive to see gene-trapping is evolving at such a rapid pace many alternative vectors and prescreens have already been introduced, as documented above. In addition many advancements are being made to make gene-trap integration analysis more accessible, eg. direct sequencing of

Table 1.2A: Characterised Gene-Trap Integrations

Vector Design	Insertion site	Gene	Gene Product	βgal activity[a]	Homozygous phenotype	Reference
<u>Retroviral</u>						
Gene-Trap						
ROSAβgeo	5'UTR	TEF-1	transcription factor	widespread	embryonic lethal	Chen *et al.*, 1994
ROSAβgeo	intron	BTF-3	transcription factor	restricted	embryonic lethal	Deng *et al.*, 1995
ROSAβgeo	intron	CathepsinB	protease	N.A.i	N.A.i	Gogos *et al.*, 1996
ROSAβgeo	intron	ROSA26	novel	widespread	none	Zambrowicz *et al.*, 1997
ROSAβgeo	intron	Sec8	secretory protein	restricted	embryonic lethal	Friedrich *et al.*, 1997
Promoter-Trap						
U3βgeo	5'UTR	Eck	receptor tyrosine kinase	restricted	none	Chen *et al.*, 1996
U3βgeo	?	ArMT	arginine methyltransferase	N.D.	embryonic lethal	Scherer *et al.*, 1996
U3neo	5'UTR	fug1	novel	N.A.	embryonic lethal	DeGregori *et al.*, 1994
U3His	5'UTR	REX-1	transcription factor	N.A.	embryonic lethal	vonMelchner et al, 1992
<u>Plasmid</u>						
Gene-Trap						
pGT4.5	intron	Gt4-2	transcription factor	restricted	lethal[b]	Skarnes *et al.*, 1992
TV2	intron	Jumonji	DNA binding	restricted	embryonic lethal	Takeuchi *et al.*, 1995
pGT1.8βgeo	intron	E-catenin	cell adhesion molecule	restricted	embryonic lethal	Torres *et al.*, 1997
pT1-ATG	intron	Tfeb	transcription factor	restricted	none	McClive *et al.*,1998
pGT4.5	intron	Gt4-1	novel	widespread	perinatal lethal	Skarnes *et al.*, 1992
pGT4.5	intron	Gt10	novel	restricted	N.D.	Skarnes *et al.*, 1992
pGT4.5	intron	Cordonbleu	novel	restricted	none	Gasca *et al.* 1995
pT1-ATG	intron	R.140	novel	restricted	embryonic lethal	Forrester, *et al.*, 1996
pT1-ATG	intron	I.114	novel	restricted	none	Forrester, *et al.*, 1996
pT1-ATG	intron	I.163	novel	restricted	none	Forrester, *et al.*, 1996
pT1-ATG	intron	I.193	novel	restricted	none	Forrester, *et al.*, 1996
pGT4.5	?	Gt2	N.D.	widespread	none	Skarnes *et al.*, 1992
pT1-ATG	intron	I.23	N.D.	restricted	none	Forrester, *et al.*, 1996
pEN53	intron	cloneVI	novel	N.A.i	N.A.i	Menchini *et al.*, 1997
Secretory-Trap						
pGT1.8TM	intron	Netrin	secreted axon guidance cue	restricted	postnatal lethal	Skarnes *et al.*, 1995
pGT1.8TM	intron	PTPκ	protein tyrosine phosphatase	restricted	none	Skarnes *et al.*, 1995
pGT1.8TM	intron	LAR	protein tyrosine phosphatase	restricted	none	Skarnes *et al.*, 1995
Enhancer-Trap						
p3LSN	?	Etl-1	DNA binding	restricted	N.D.	Soininen *et al.*, 1992
p3LSN	intron	Etl-2	type-I cytokine receptor	restricted	N.D.	Neuhaus *et al.*, 1994
PolyA-Trap						
pPAT	intron	pat-12	novel	N.D.	none	Yoshida *et al.*, 1995

? unknown
N.D. not determined
N.A. not applicable as vector does not contain a reporter gene
N.A.i not applicable *in vitro* study
a βgal activity during development
b variable penetrance, some homozygotes viable with growth defect

Table 1.2B: Large Scale Gene-Trap Screens

Screen	Vector	Trapped Sequence Data	Homology (% of total)		
			Known Genes	ESTs	Novel
	Plasmid Gene-Trap				
Chowdhury *et al.*, 1997	pSAβgeo and pSAIRESβgeo	5'-RACE PCR	17 (30%)	11 (20%)	28 (50%)
Townley *et al.*, 1997	pGT1.8TM	Direct Sequencing	29 (51%)	11 (19%)	17 (30%)
	Retroviral Shuttle				
Hicks *et al.*, 1997	U3neoSV1	PSTs	42 (11%)	21 (5%)	337 (84%)*

PSTs, Promoter-Proximal Sequence Tags are derived from genomic DNA flanking insertion sites.
* Figures are derived from the comparison of PSTs to cDNA databases. Insertions of retroviral vector into exons would validate the of comparison PSTs to cDNA databases however this is not always the case and the vector may insert into promoter or intronic sites. As a consequence significant matches can be missed because the PST does not contain sufficient exon sequences. This results in an inaccurate over-representation of insertions into novel genes.

RACE products (Townley *et al.*, 1997). Modification of RACE-PCR techniques has led to the development of direct sequencing (Townley *et al.*, 1997). This technique allows selection of integration events at the sequence level without the need of cloning.

1.24 Generation and Isolation of Gene-Trap Cell lines R68 and R124

The gene-trap vector, PT1.ATG (Section 1.19), was electroporated into R1 ES cells (Simpson *et al.*, 1997). The PT1.ATG gene-trap vector is more precisely termed a intron trap vector and includes a splice acceptor derived from the *engrailed-2* gene, a β-galactosidase reporter gene with an inframe ATG translation initiation site at its 5' end and a neomycin selectable marker under the regulation of the phosphoglycerate kinase-1 (PGK-1) promoter. The presence of an ATG allows the translation of the reporter gene even when the vector has integrated into the 5' UTR (untranslated region) of a gene. The independently-driven selectable marker allows selection of all gene-trap integrations, including integrations into sites which are not active in undifferentiated ES cells. Integration of the vector into an intron of a functional gene allows a fusion transcript to be formed via the splice acceptor on the vector. This integration results in transcription of the reporter gene to be directly regulated by the promoter sequences of the endogenous gene.

ES cells transfected with the PT1.ATG were selected in the presence of G418 and replica plated. These ES cells were then induced with RA and stained for β-galactosidase (β-gal) activity to screen for cells carrying a correctly integrated and spliced vector. β-galactosidase-positive clones were then retested in the presence and absence of RA to select for integrations into RA-responsive genes. Twenty gene-trap cell lines were found to respond to RA, of these LacZ expression was induced in nine lines and repressed in eleven lines. Tetraploid mouse chimeras derived from eight of the eleven repressed gene-trap cell lines were expressed in the heart (Forrester *et al.*, 1996). This observation may be of significance as RA has been shown to inhibit

cardiogenesis in embryoid bodies by repression of cardiac specific genes, an *in vitro* model for early cardiogenesis (Wobus *et al.*, 1994).

Two of the repressed cell lines, R68 and R124, displayed reporter gene expression almost exclusively in the developing heart of 9-10 d.p.c. tetraploid chimeras.

1.25 **Aim of the Project**

The aim of this Ph.D project is to characterise two gene-trap integrations, R68 and R124 isolated from a gene-trap screen conducted by Forrester *et al.*, 1996. The characterisation of the gene-trap integrations includes identification of endogenous trapped gene sequences and analysis to determine the function of the trapped genes in heart development.

Chapter 2

2.0 Materials and Methods

2.1 Molecular Techniques

The preparation of standard solutions and general molecular techniques described in this section were according to Sambrook *et al.*, 1989 unless otherwise stated.

2.1.1 General Cloning Procedures

Restriction enzyme digests were performed as recommended by the suppliers (Boehringer Mannheim) and digest products were analysed by gel electrophoresis. Agarose gels at appropriate concentrations were prepared in 1xTAE buffer and stained with 0.5μg/ml ethidium bromide.

a. Preparation of Vectors for Cloning

Plasmid vectors linearised with restriction enzyme digestion for cloning purposes were treated with alkaline phosphatase to prevent recircularisation. Following digestion 1unit of calf intestinal phophatase (CIP - Boehringer) was added directly to the reaction and incubated at 37°C for 30 minutes. To inactivate the CIP, 5mM EDTA, 10μg proteinaseK and 0.5%SDS were added to the reaction mix and incubated at 56°C for 30 minutes. The reaction mix was then extracted with an equal volume of phenol/chloroform, precipitated with 0.1volume of 3M NaOAc pH5.2 and 2volumes of ethanol at -20°C for 1 hour. Following precipitation the plasmid DNA was pelleted by centrifugation, washed in 70% ethanol and resuspended in T.E.

b. Preparation of DNA Inserts

Restriction enzyme digest products were run on 0.8% low melt agarose gels. Bands of interest were excised under long wavelength UV illumination. An equal volume of dH_2O and sodium chloride to a final concentration of 0.1M was added to the excised band, heated to 70°C for 5 minutes and vortexed to melt the agarose. An equal volume

of Tris saturated phenol, prewarmed to 37°C, was added, mixed by vortexing and the suspension was centrifuged at 6.5K for 5 minutes into an upper aqueous phase, containing the DNA, and lower organic phase. The aqueous phase was precipitated with 2volumes of 100% ethanol at -20°C for 30 minutes Precipitated DNA was pelleted by centrifugation at 13K for 10 minutes, washed with 70% ethanol, air dried and resuspended in 20µl of T.E.

2.1.2 Ligations

Ligation reactions were set up with a vector to insert molarity ratio of 1:3 and included 1unit of T4 DNA Ligase and of 1x ligation buffer (66mM Tris.HCl pH7.5; 5mM $MgCl_2$; 1mM DTT; 1mM ATP) (Boehringer) in a total volume of 10µl. A control reaction containing vector alone was also set up to determine the background number of colonies due to self-ligation of vector. A successful ligation reaction usually produced an enrichment of more than 100-fold colonies over the vector-alone ligation. Cohesive-end ligations were performed at 16°C for a minimum of 1 hour, blunt-end ligations were incubated overnight at room temperature. Ligation reactions were precipitated (with 0.1 volume of 3M NaOAc pH5.2 and 2volumes of ethanol) and resuspended in 5µl of T.E before transformation by electroporation.

2.1.3 Transformation of Bacterial Cells

Electrocompetent cells were prepared by inoculating a single colony of the required bacterial strain, DH5α or XL1-Blue, into a 25 ml LB broth culture and was grown at 37°C overnight with vigorous shaking. 10ml of the overnight culture were used to inoculate 1litre of LB broth and grown to a specific density to ensure log phase (OD_{600}=0.6-0.9). The cells were chilled on ice for 15 minutes before centrifugation at 4K for 20 minutes at 4°C. The bacterial pellet was resuspended in a total of 1litre of ice cold water, re-centrifuged and resuspended in 500ml of ice cold water.

Centrifugation was repeated and the bacterial pellet resuspended in 40ml of ice cold 10% glycerol in water and after another round of centrifugation into 2ml of ice cold 10% glycerol. The electrocompetent bacterial cells were aliquoted into prechilled eppendorf tubes and snap frozen in liquid nitrogen before storage at -80°C.

Frozen aliquots of electrocompetent cells were thawed on ice for 30-60 minutes before electroporations. 35 µl of electrocompetent cells were mixed with 3-5µl of the ligation reaction in a prechilled cuvette with a 0.2cm electrode gap(Biorad). Cells were electroporated using Biorad's Gene Pulser (25 µFD, 200 ohm and 1.8 kV and a time constant of 3-4msec). 1.5ml of LB supplemented with 10% 2M glucose was immediately added to the cells following electroporation. The transformed cells were transferred to an eppendorf tube and placed at 37°C for 45 minutes to recover, with no shaking. Appropriate dilutions were plated onto LB-Ampicillin plates. When blue/white selection was required, standard 90mm agar plates were treated with 40µl of 20mg/ml X-gal and 4µl of 0.2mg/ml IPTG prior to plating. Transformation efficiencies of 10^8 transformants per µg of plasmid were typically acheived.

2.1.4 **Plasmid Preparation - "Miniprep"**

A single colony, from a freshly streaked plate, was used to inoculate 3ml of L broth containing the appropriate antibiotics. This culture was grown at 37°C overnight with shaking. 1.5ml of the overnight culture was transferred to an eppendorf tube and the cells were harvested by centrifugation at 3K for 10 minutes. The supernatant was discarded and the pellet resuspended in 100µl of solution1 (50 mM glucose; 25 mM Tris-HCl pH 8.0; 10 mM EDTA; 100µg/ml RNase A). 200µl of freshly prepared solution2(0.2 M NaOH;1% SDS) and 150µl of ice cold solution3 (3M KAcetate pH4.8 with acetic acid) were added sequentially, mixed and then placed on ice for 5 minutes. A precipitate formed which was removed by centrifugation at 13K for 10 minutes. 400µl of the DNA containing supernatant was extracted with an equal volume

of phenol/chloroform and precipitated with 240µl of isopropanol. Following centrifugation at 13K for 10 minutes, the pellet was washed in 70% ethanol, air dried and resuspended in 50µl of T.E. This protocol was scaled up a 100 fold to prepare larger quantities of DNA ("Midiprep"- 50ml culture and "Maxiprep"- 150ml culture). The quality of DNA from this preparation was sufficient for diagnostic digest analysis. To improve quality the DNA preparation was reprecipitated with an equal volume of 13% PEG 8000 in 1.6M NaCl on ice for 1 hour. The precipitated DNA was pelleted by centrifugation at 13K for 5 minutes, washed twice with 70% ethanol, air dried and resuspended in 50ul T.E. This procedure improved the quality of the DNA such that it could be used for all cloning procedures, probe fragment preparations and sequencing reactions. Alternative methods of plasmid preparations included the use of Qiagen and Promega Wizard kits. Reagents and protocols were supplied with the kits by manufacturers.

2.1.5 **Sequencing and Analyses of Double Stranded DNA Templates**

DNA was subcloned into the multiple cloning sites of commercially available plasmids flanked by T3/T7 or T7/Sp6 primer sites eg. pBluescript II KS-. Both strands of DNA were routinely sequenced to allow any ambiguities to be clarified. BLAST database search programs on the Netscape, provided by National Centre for Biotechnology Information(NCBI) (http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST), which search Genbank, EMBL, DDBJ, DDBJEST, PDB, SwissProt and PIR (Altschul et al, 1990) were used to perform searches with sequence data to identify any known homologies. Sequences were analysed and assembled using DNA Star computer software to identify ORF, map restriction enzyme sites and generate contigs of sequence.

2.1.5.1 **Manual Dideoxy-Termination Sequencing**

The dideoxy-termination method is based upon the elongation of a labelled DNA strand using DNA polymerases (Sanger *et al.*, 1977). Incorporation of terminating dideoxy nucleotides, that lack 3'-OH groups necessary for strand elongation, results in a series of nested DNA strands that can be resolved on a denaturing polyacrylamide gel. Four termination reactions, each for a different dideoxy nucleotide, are performed for each DNA template to give complete sequence information. DNA was sequenced using the Sequenase Version 2.0 DNA Sequencing Kit reagents (USB-Amersham) following the recommended protocol with modifications to improve annealing of the primer to the template DNA and to minimise re-annealing of template strands. This was done by alkali denaturing the plasmid in the presence of the primer and by the addition of DMSO.

2.5μg of template plasmid DNA (resuspended in T.E. pH7.5) mixed with 4pmoles of oligonucleotide primer in a volume of 20μl was denatured with 0.2M NaOH at room temperature for 5 minutes. The mixture was precipitated with 0.4M ammonium acetate (pH4.6 with acetic acid) and 2volumes of 100% ethanol on ice for 15 minutes followed by centrifugation at 13K for 15 minutes at $4^{o}C$ to pellet the DNA, washed in 70% ethanol and then resuspended in 12.5μl sequencing buffer (10%DMSO; 40mM Tris.HCl pH7.5; 20mM $MgCl_2$; 50mM NaCl). To this 1μl 0.1M DTT; 2μl labelling mix (1.5μM each of dGTP, dTTP and dCTP); 0.5μl α-^{35}S dATP (5μCi); 2μl(3.25units) DNA polymerase was added and the reaction incubated at room temperature for 5 minutes. 2.5μl of each nucleotide termination mix (8μM dideoxynucleotide; 80μM of the other three deoxynucleotides; 50mM NaCl; 10% DMSO) was aliquoted into a 96well plate and prewarmed to $37^{o}C$. 3.5μl of template was added to each nucleotide termination mix (ddGTP; ddATP, ddTTP and ddCTP) and incubate at $37^{o}C$ for 5 minutes. Termination reactions were stopped by adding 4μl of stop solution (95% formamide; 20mM EDTA; 0.05% bromophenol blue; 0.05%

xylene cyanol FF). Sequencing samples were denatured at 80°C for 10 minutes and ran on a 6% denaturing polyacrylamide gel (7M urea; 5.7% acrylamide; 0.3% bisacrylamide; 1xTBE; 0.06% ammonium persulphate; 35µl TEMED per 100ml mix) in 1xTBE. After electrophoresis the gel was soaked in 15% methanol; 5% acetic acid for 15 minutes, transferred to Whatman 3MM paper and dried at 80°C under vacuum before being exposed to autoradiographic film at room temperature. Overnight exposure was usually sufficient to give readable signals.

2.1.5.2 **Automated Cycle Sequencing**

Automated cycle sequencing is a PCR based Sanger dideoxy-termination method (Sanger *et al.*, 1977). Modifications include each of the four dideoxy terminators labelled with a different fluorescent dye. Thus all termination reactions are performed in a single tube. The wavelength of light emitted by each fluorescent dye can be detected and interpreted by ABI sequencing computer software. Cycle sequencing was performed using the Perkin-Elmer Taq DyeDeoxy Terminator Cycle Sequencing Kit according to the suppliers recommendations. 0.5µg of double stranded DNA resuspended in dH_20 was mixed with 3.2pmole primer and 8.0µl of terminator ready reaction mix (A, C, G, T - dye terminators; dITP; dATP; dCTP; dTTP; Tris.HCl pH9.0; $MgCl_2$; thermal stable pyrophosphatase; AmpliTaq DNA polymerase) in a volume of 20µl. This reaction was overlaid with 40µl mineral oil and cycle sequenced in a Hybaid Omnigene Thermocycler for 25 cycles. Each cycle consisted of:

96°C	30 seconds
50°C	15 seconds
60°C	4 minutes

Each step was preceeded by a rapid thermal ramp to achieve the required temperature. The sequencing reaction was precipitated with 0.3M sodium acetate pH4.6 and 2volumes of 100% ethanol on ice for 15 minutes followed by centrifugation at 13K for

25 minutes. The supernatant was discarded, the pellet washed in 70% ethanol, air dried and resuspended in 6µl loading buffer (5mM EDTA pH8.0; 10mg/ml Blue dextran in deionised formamide). Sequencing reactions were denatured at 95°C for 5 minutes and ran on denaturing polyacrylamide gel (7M urea; 5% acrylamide (29:1 Biorad); 1xTBE; 0.06% ammonium persulphate; 15µl TEMED per 50ml mix) in 1xTBE. The sequencing run, routinely 500-600bases, was processed on the ABI PRISM 377 DNA Sequencer combined with sequencing ABI Prism 2.1.1 computer software and presented as a nucleotide sequence with a profile of signal intensity, that allowed any ambiguities to be clarified.

2.1.6 **Isolation of Genomic DNA from ES Cells**

ES cells were grown to confluency in 25cm^2 flasks, rinsed twice in PBS and then lysed overnight at 37°C with 5ml of lysis buffer (100mM Tris.HCl pH8.5; 5mM EDTA; 0.2% SDS; 200mM NaCl) containing freshly added 100µg/ml Proteinase K. The lysate was extracted with an equal volume of phenol/chloroform/isoamyl (mixture 25:24:1) and centrifuged at 6K for 10 minutes to separate aqueous and organic phases. Extraction of the aqueous layer was repeated with chloroform/isoamyl (mixture 24:1). The aqueous phase was precipitated with an equal volume of isopropanol. Genomic DNA precipitates as the solutions are mixed by inversion. The DNA was spooled out of the tube using a sterile gilson tip, rinsed in 70% ethanol, and then resuspended in 50µl of T.E. To remove any ethanol carry-over the open sample tube was incubated at 60°C for 30 minutes. The sample was left overnight to complete the resuspension of the genomic DNA before determining the concentration by optical density measurements at 260nm.

2.1.7 **Isolation of Genomic DNA from Tissue**

Tail tips (0.5-1.0cm) were taken routinely from 3 week old animals to isolate genomic DNA for genotyping purposes. Tissue was digested overnight at 55°C in 0.5ml lysis buffer (100mM Tris.HCl pH8.5; 5mM EDTA; 0.2% SDS; 200mM NaCl) with freshly added 100μg/ml Proteinase K. The lysate was vortexed vigorously and centrifuged at 13K for 10 minutes to remove indigestible tissue eg. bones and hair. The resulting DNA containing supernatant was extracted and precipitated as described in 2.1.7.

2.1.8 **Southern Blot Analysis**

Genomic DNA (10μg) was digested overnight at 37°C with 40units of appropriate restriction enzymes and products analysed on 1% agarose gels. After electrophoresis the DNA was stained/destained with ethidium bromide and photographed. The gel was then denatured (0.5M NaOH; 1.5MNaCl) for 30 minutes and neutralised (1.5M NaCl; 0.5M Tris.HCl pH7.4) for 30 minutes. To set up the overnight capillary dry blot the gel was placed face down on Saran wrap. Hybond N^{+} membrane (Amersham) cut to the size of the gel was pre-soaked in 20xSSC and placed on to the gel followed by 2 pieces of Whatman 3MM paper (soaked in 20xSSC) and finally a dry piece of 3MM paper (as each layer was added all bubbles were carefully removed). A stack of absorbent paper towels and a weight was placed on top to encourage capillary flow. After overnight blotting the membrane was marked for orientation, rinsed in 2xSSC and baked at 80°C for 2 hours.

2.1.9 **Isolation of Total RNA from ES cells and Tissue**

Total RNA was isolated using a modified guanidinium isothiocynate and phenol RNA isolation protocol described by Chomczynski and Sacchi, (1987).

ES cells were grown to confluency in 90mm gelatinised plates, rinsed twice with PBS and then treated with 5ml of Solution D (4.4M Guanidinium Isothiocynate; 25mM

Sodium Citrate; 0.6% Sarcosyl; 100mM β-Mercaptoethanol). The contents of the plate were then scraped and transferred to centrifuge tube (Falcon 2059). Tissue for RNA isolation was freshly dissected and 0.5g were taken to be homogenised using a standard ground glass homogeniser. 5ml of Solution D was added to this and the lysate was transferred to a centrifuge tube (Falcon 2059). The following steps are common to the isolation of RNA from ES cells and tissue.

0.5ml 2M Sodium Actetate, 5ml Phenol and 2ml Chloroform were sequentially added to 5ml of the lysate, each reagent was thoroughly mixed with the lysate by inversion. The suspension was then placed on ice for 15 minutes followed by centrifugation at 6K for 15 minutes at 4°C (Sorval SS34 rotor) into a upper aqueous phase containing the RNA and a lower organic phase. RNA from the aqueous phase was precipitated by addition of an equal volume of isopropanol at -20°C for 1 hour followed by centrifugation at 8K for 20 minutes at 4°C (Sorval SS34 rotor). The pellet was resuspended in 0.3ml Solution D and reprecipitated with an equal volume of isopropanol at -20°C for 1 hour. The RNA was pelleted by centrifugation at 13K for 10 minutes at 4°C, washed with 70% ethanol, air dried and resuspended in 50μl of DEPC treated water. The RNA was placed at 4°C overnight before determining the concentration by absorbance at 260nm. Ratio of 260nm/280nm was used to estimate the purity of the sample. A ratio close to 1.8 indicated the sample was not significantly contaminated by protein.

2.1.10 **Northern Blot Analysis**

Total RNA (10μg) samples were mixed with 3volumes of sample buffer (66% deionised formamide; 22% formaldehyde; 1.2xMOPS) and denatured at 70°C for 10 minutes. Followed by cooling on ice and the addition of 2μl of RNA loading dye (50% glycerol; 1mM EDTA; 0.4% bromophenol blue; 0.4% xylene cyanol; 2μg ethidium bromide) the samples were separated on a 1% denaturing agarose gel

prepared in 1xMOPS (20mM MOPS; 4mM sodium acetate; 1mM EDTA) containing 17.5% formaldehyde. After electrophoresis the gel was denatured in 50mM NaOH for 30 minutes followed by neutralisation in buffer (1.5M NaCl; Tris.HCl pH 8.0; 1mM EDTA) for 30 minutes. An overnight capillary dry blot was then set up to transfer the RNA to Hybond N^+ membrane as described for Southern blot analysis (2.1.8).

2.1.11 Random-Primed Labelling

All experiments involving radioactivity were performed in designated areas following strict guidelines to minimise exposure and avoid contamination. DNA fragments to be used as ^{32}P-labelled probes for hybridisation to membrane blots were purified by elution from agarose gel bands (2.1.2). Random-primed labelling was performed using a "high prime" kit as recommended by Boerhinger: 25ng of DNA was mixed with dH_2O to a volume of 12μl, denatured at 100°C for 10 minutes, snap cooled on ice and then collected by centrifugation. On ice, 4μl high prime mix (1unit/μl Klenow polymerase; 0.125mM dATP; 0.125mM dGTP; 0.125mM; 0.125mM dTTP; 50% v/v glycerol) and 4μl of 50μCi [$\alpha^{32}P$]dCTP (3000Ci/mMol) aqueous solution were added to the denatured DNA and the reaction was then incubated at 37°C for 10 minutes. The reaction was stopped by adding 1μl 0.5M EDTA pH8.0 and 79μl of dH_2O. Unincorporated nucleotides were removed by centrifugation of the reaction through a G-50 Sephadex column at 2K for 2 minutes. An aliquot of 1μl from the labelled probe was used to determine the [$\alpha^{32}P$]dCTP incorporation.

2.1.12 Hybridisation Conditions

DNA and RNA blot hybridisations were performed overnight in rotating Techne hybridisation bottles in Techne Hybridiser HB-1 ovens. Blots were prehybridised for 1 hour in hybridisation buffer: Southern blots in 0.5M Na_2HPO_4 pH7.1 (with phosphoric acid); 15% deionised formamide; 1% BSA (w/v); 7% SDS; 1mM EDTA

pH8.0 at 65°C and Northern blots in 0.35M Na_2HPO_4 pH7.1; 30% deionised formamide; 1% BSA (w/v); 7% SDS; 1mM EDTA pH8.0 at 60°C.

The labelled probe was denatured at 100°C for 5 minutes, snap cooled on ice and centrifuged before adding to the blot with fresh hybridisation buffer. Routinely 10^7cpm of probe in 10ml hybridisation buffer were used for hybridising to a standard (13x11cm) midi-gel blot. Following hybridisation DNA and RNA blots were washed in a common wash buffer (150mM Na_2HPO_4 pH7.1; 0.1% SDS) 3 times for 20 minutes each at 60°C.

After washing the blots were wrapped in Saran wrap and exposed to autoradiographic film at -70°C for an appropriate length of time (1-7 days) for a signal to appear. Alternatively hybridised blots were exposed to Molecular Dynamics phosphor screens that were processed using a Molecular Dynamics phosphoimager and Imagequant Molecular Devices software to quantify signals. This computer software package gave a measurement of signal strength relative to background levels and to internal control signal bands.

Prior to rehybridising, DNA and RNA blots were stripped in boiling stripping buffer (0.01xSSC; 0.01xSDS) to remove bound radiolabelled probe. The buffer was changed every 10 minutes until the counts dropped to background levels. Stripped blots were then re-exposed to autoradiographic film to check the probe had been removed prior to rehybridisation.

2.1.13 **PCR amplification of genomic DNA**

The PCR reactions were set up on ice, in a total volume of 50μl, as follows: 0.5μg genomic DNA; 1xPCR buffer [50mM KCl/10mM Tris (pH8.3)]; 2mM $MgCl_2$; 0.2mM dNTP's; 0.6μM *Primer a1/b1; 0.6μM *Primer a2/b2; 28μl dH_20. The reaction was mixed gently, overlaid with 50μl of mineral oil and heated at 94°C for 5

minutes before the addition of 2.5units of Taq DNA polymerase (Promega). The PCR amplification was performed on a thermocycler through 30 cycles of:

94°C 1.5 minutes

60°C 1.5 minutes

72°C 3 minutes

The final step at 72°C was for 10 minutes to fill in single stranded ends. The combination of *Primer a1/a2 and b1/b2 give products of predicted size that were analysed by agarose gel electrophoresis and Southern blot analysis.

*Primers	Sequence (5' - 3')
a1(en2-intron)	ACTTGGCCTCACCAGGC
a2(en2-exon)	TGCTCTGTCAGGTACCTGTTGG
b1(KXE)	ATCCACCAATTTGAAGAACACC
b2(en2-intron)	TGAGCACCAGAGGACATCCG

2.1.14 **Rapid Amplification of cDNA Ends (5' RACE)**

The RACE protocol used is based upon the method described by Frohman *et al.*, 1988 with modifications as described in Townley *et al.*, 1997. 5'RACE allows the cloning of unkown 5'ends of mRNA. Reverse transcription was primed from a known sequence to generate a single stranded cDNA. This was tailed with dCTPs using terminal transferase. An anchor primer, complementary to the dCTP tail, was used to prime the second cDNA strand to generate double stranded cDNA. The double stranded cDNA was amplified using specific nested primers. Restriction sites engineered within the nested primers were used to directionally clone the amplified sequence into pBluescript II KS- plasmid. The Primers used were:

Primer(homology)	Sequence(5' - 3')	Restriction Enzyme site
1(*lacZ*)	GCAAGGCGATTAAGTTGGGT	-
2(anchor)	GGCCACGCGTCGACTAGTACGGGIIGGGIIGGGIIG	SpeI
3(*en-2*)	CCGTCGACTCTGGCGCCGCT	-
4(*en-2*)	TGCTCTGTCAGGTACCTGTTG	KpnI

1. First strand cDNA synthesis

1 μg of total RNA (isolated from ES cells as described in 2.1.10), 10ng primer 1 and sterile dH_2O, in a volume of 12μl, were mixed on ice and denatured at 70°C for 5 minutes followed by snap cooling on ice. The contents were collected by brief centrifugation and 1x1st strand buffer (50mM Tris.HCl pH8.3; 75mM KCl; 3mM $MgCl_2$); 10mM DTT; 0.5mM dNTP were added. This mix was heated at 37°C for 2 minutes and 200units of Superscript II (GibcoBRL) was added to a total volume of 20μl. The reverse transcription reaction was incubated at 37°C for 1 hour. After 1st strand cDNA synthesis, 0.1M NaOH was added and reactions incubated at 65°C for 20 minutes to hydrolyse RNA then neutralised with 0.1M HCl. Microdialysis was performed (to remove primers, buffers and free nucleotides) by loading all the 1st strand reaction onto a 0.025μm microdialysis filter (Millipore) and floating in a petri dish (~50ml) of T.E. and for 4 hours. Routinely, 6-8μl were recovered after dialysis and dH_20 was added to 20μl.

2. Poly C Tail Addition

1xTdT buffer (0.1M Potassuim cacodylate pH7.2; 10mM $CoCl_2$; 1mM DTT) and 0.13mM dCTP were added to 20μl of the 1st strand reaction and samples were incubated at 37°C for 2 minutes. To this 20units of terminal transferase(TdT GibcoBRL) were added and the reaction incubated at 37°C for 5 minutes, followed by heat inactivation at 70°C for 2 minutes.

3. 2nd Strand Synthesis

1x restriction buffer M(1mM Tris.HCl pH7.5; 5mM NaCl; 1mM $MgCl_2$; 0.1mM dithioerythritol) (Boehringer); 0.5mM dNTPs; 0.5ng/µl primer 2; 2units Klenow (Boehringer) were added to 15µl of poly C tailed cDNA. This reaction was incubated at room temperature for 30 minutes followed by incubation at 37°C for 30 minutes and inactivation at 70°C for 5 minutes. The cDNA sample was then microdialysed on 0.1 µm microdialysis filter as above. Routinely 8 -10µl were recovered after dialysis and dH_20 was added to 37µl.

4. 1st Round PCR amplification

The PCR reactions were set up on ice as follows: 5µl cDNA; 1xPCR buffer (50mM KCl/10mM Tris (pH8.3); 2mM $MgCl_2$; 0.2mM dNTP's; 2ng Primer 2; 2ng Primer 3; 28µl dH_20. The reaction was mixed gently, overlaid with 50µl of mineral oil and heated at 94°C for 5 minutes before the addition of 2.5units of Taq DNA polymerase (Promega).

The PCR amplification was performed on a thermocycler through 35 cycles of:

94°C 1min

60°C 2min

72°C 3min

The 1st round PCR products were dialysed on 0.1 µm filters as above which results in the retention of the largest PCR products. Routinely 15-18µl were recovered after dialysis and dH_20 was added to 40µl.

5. 2nd Round PCR amplification

The second round of PCR amplification reactions were performed as above using the anchor primer (Primer 2) and a nested primer in *en-2* sequence (Primer 4).

6. Specificity of PCR Amplification

The specificity of the first and second rounds of PCR amplificaton were checked by running 1/10th of the 40µl volume of PCR products after microdialysis on an agrose gel. The gel was Southern blotted and the blot hybridised with a ^{32}P-labelled probe

specific to engrailed sequences within the vector. Control PCR reactions, with no template/ no reverse transcriptase/ no polymerase/ no primers, which result in no amplification of the trapped sequences were also run.

7. Size Selection of PCR products

The second round PCR products were run on a 0.8% low melting point agarose gel. Products between 1.0 and 0.5kb were cut from the gel and purified by phenol extraction and precipitation (2.1.2).

8. Cloning PCR products

The PCR amplified cDNA products were digested and directionally cloned into a SpeI/KpnI site in pBluescript II KS- (See section 2.1.1 for general cloning procedures).

2.1.15 **RNase Protection**

1. Riboprobe Templates

Plasmids containing RACE products were linearised by restriction digest to allow the generation of antisense run-off transcripts of predicted sizes. Linearised plasmids were purified as described in 2.1.2.

2. Riboprobe synthesis

1µg of template DNA was mixed with: 10mM DTT; 0.4µg BSA; 1mM AUG (ATP:UTP:GTP at 1:1:1); 20units RNase inhibitor; 1X transcription buffer (40mM Tris.HCl pH8.0; 6mM $MgCl_2$; 10mM dithoithreitol; 2mM Spermidine); 250µci $\alpha^{32}P$ CTP; 20units polymerase (usually T3) in a total volume of 20µl.

The reaction mix was incubated at 37°C for 1 hour. Following the *in vitro* transcription, the reaction was treated with 20units RNase-free DNase at 37°C for 15 minutes to remove the DNA template. The reaction volume was made up to 100µl with the addition of dH_2O and centrifuged through a sephadex G-50/SpinX column (Sigma Costar) at 13K for 5 seconds to remove protein and unincorporated nucleotides.

3. Gel Purification of Riboprobe

96μl of riboprobe was mixed with 64μl of loading dye(95% formamide; 20mM EDTA; 0.05% bromophenol blue; 0.05% xylene cyanol FF) denatured at 80°C for 3 minutes and run on a 6% polyacrylamide sequencing gel(7M urea; 5.7% acrylamide; 0.3% bisacrylamide; 1xTBE; 0.06% ammonium persulphate; 35μl TEMED per 100ml mix) in 1xTBE for 1.5hours at 60W.

After the electrophoresis the wet gel was wrapped in Saran wrap and exposed to autoradiographic film (the film was placed on the gel and marked for orientation). The autoradiograph was aligned to the gel and the area of the gel containing the probe band excised. The gel was re-exposed to ensure that the correct region of the gel had been isolated. The probe was eluted from the gel slice by incubation in 1ml of probe elution buffer (0.5M ammonium acetate; 1mM EDTA; 0.2% SDS) at 37°C for 2 hours with vigorous shaking, the gel pelleted by centrifugation at 13K for 5 minutes and the probe-containing supernatant removed. 1μl of riboprobe, in 1ml of scintillation fluid (Ultima Gold), was used to determine the incorporation.

4. Hybridisation

10μg target RNA and 3.5 x 10^5cpm of eluted riboprobe in a total volume 100μl was precipitated with 0.3M sodium acetate, 20μg glycogen and 2volumes of 100% ethanol at -80°C for 30 minutes. RNA was pelleted by centrifugation at 13K for 10 minutes at 4°C, washed with 70% ethanol, air dried and resuspended in 30μl of hybridisation mix (0.4M NaCl; 40mM PIPES pH6.4; 1mM EDTA in deionised formamide). Hybridisation reactions were then denatured at 85°C for 15 minutes and hybridised overnight at 55°C.

5. RNase Digestion

Unhybridised single stranded RNA was digested in 350μl of digestion buffer (10mM Tris.HCl pH7.5; 5mM EDTA; 0.3M NaCl; 14ug RNase; 2000units Rnase T1) at 30°C for 30 minutes. 0.5% SDS and 0.13μg/μl proteinase K were added to stop the

digest with a further 10 minutes incubation at 30°C. Reactions were then extracted with phenol/chloroform and precipitated with 1ml 100% ethanol using 5µg tRNA as a carrier at -80°C for 30 minutes. RNA was pelleted by centrifugation at 13K for 10 minutes, air dried and resuspended in 4µl of loading dye. Samples were then denatured briefly at 95°C and run on a 6% polyacrylamide gel. The gel was dried before exposing to autoradiographic film at -80°C.

2.1.16 **Screening cDNA Libraries**

1. Library Titre

10ml of LBroth, supplemented with 0.2%maltose and 10mM $MgSO_4$, were inoculated with a single colony of C600Hfl cells from a freshly streaked plate and grown overnight at 37°C. Cells were harvested by centrifugation at 4K for 10 minutes and resuspended in 5ml 10mM $MgSO_4$, such that OD_{600}= 0.6. Serial dilutions of library phage lysate made in 100µl SM buffer (100mM NaCl; 10mM $MgSO_4$; 50mM Tris.HCl pH7.5; 0.01% gelatin) were mixed with 200µl of overnight bacterial culture and incubated at 37°C for 15 minutes. 3ml of 0.7% molten top agarose, supplemented with 10mM $MgSO_4$, were added to each dilution preparation and immediately poured contents onto prewarmed LB plates. The top agarose was allowed to set at room temperature and incubated at 37°C overnight. The titre was calculated from the number of plaques with respect to the dilution factor. The Clonetech Mouse Heart 5'- Stretch plus cDNA library gave a titer of $1x10^8$ pfu/ml.

2. Library Plating

i. A total of $1x10^6$ pfu were plated over four (20x20cm) agar plates. These plates were poured (250ml agar) the night before use and dried at 37°C for 4-6hours. For each plate $2.5x10^5$ pfu were mixed with 2ml C600Hfl cells in 10mM $MgSO_4$ (prepared as in 2.1.16.1) and incubated at 37°C for 30 minutes. Phage/bacterial preparations were mixed with 30ml of 0.7% molten top agarose supplemented with 10mM $MgSO_4$ and

immediately poured evenly onto separate prewarmed LB agar plates. The top agarose was allowed to set at room temperature and then the plates were incubated (inverted) at 37°C for 5-8 hours until lysis was observed. The plates were then placed at 4°C for the top agarose to harden before plaque lifts

3. Plaque Lifts

Hybond N^{+} nylon membrane (Amersham) was placed on to the surface of the agarose, avoid trapping bubbles, for 1min and marked for orientation by stabbing the agar through the overlaid membrane. A second duplicate lift was taken by placing a fresh piece of membrane on the agarose for 2 minutes and marking the initial alignment. Each membrane was then placed face up on Whatman 3MM paper soaked in: denaturing solution (1.5M NaCl; 0.5M NaOH) for 2 minutes; neutralising solution (0.5M Tris.HCl pH8.0; 1.5M NaCl) for 3 minutes; neutralising solution for 3 minutes and then rinsed in 2xSSC. The membranes were then air dried briefly and baked at 80°C for 2 hours.

4. Screening Library

Filters were prehybridised in prehybridisation buffer (5xDenhardt's; 6xSSC; 0.5% Sarkosyl; 100μg/ml denatured herring sperm) at 65°C for 2-3 hours. Overnight hybridisation was performed at 65°C with a ^{32}P-labelled probe (denatured and snapped cooled before addition) at 10^{6}cpm/ml hybridisation buffer (5xDenhardt's; 6xSSC; 0.5% Sarkosyl; 100μg/ml denatured herring sperm; 100μg/ml dextran sulphate). Filters were washed in buffer (2xSSC; 0.1%SDS) four times for 30 minutes at 65°C and exposed to autoradiographic film at -80°C.

The exposed film was developed and aligned to the library plate to identify positive plaques.

5. Isolation of Single Positive Plaques

Positive plaques from the first primary screen were cored with the wide end of a sterile blue Gilson tip and placed in 500μl of SM buffer. Elution of the phage was

encouraged by vigorous shaking for 2 hours. A secondary screen was then performed with serial dilutions of each primary plaque in 100μl of SM bufffer mixed with 100μl of cells plated upon 90mm agar plates in 3ml of top agarose. Thereafter steps 3, 4 and 5 were repeated, until a single positive plaque could be isolated to give a pure population of positive plaques.

6. Preparation of Phage DNA

Phage was eluted from a single positive plaque overnight in 500μl of SM buffer at 4^oC. 100μl of phage stock diluted to 200μl with SM buffer was added to 100μl C600Hfl cells, freshly cultured til OD_{600}=0.6, and incubated at 37^oC for 30 minutes for the phage to adsorb to the bacterial cells. The phage/bacterial cell mix was then used to inoculate a 15ml LBroth culture supplemented with 10mM $MgSO_4$ that was grown at 37^oC with vigorous shaking until lysis. For complete lysis 75μl of chloroform was added to the lysate and shaken at 37^oC for a further 10 minutes. The lysate was then centrifuged at 4K for 10 minutes and the supernatant containing the phage removed to a fresh tube.

10ml TM buffer (10mM $MgSO_4$; 10mM Tris.HCl pH7.5); 16μg/ml DNaseI; 16μg/ml RNaseA, were added to 10ml of lysate and incubated at room temperature for 15 minutes. 0.5M NaCl and 2.2g of PEG were added and the mix incubated at room temperature until the PEG dissolved and thereafter placed on ice for 15 minutes. The phage DNA was pelleted by centrifuge at 9.5K for 10 minutes at 4^oC, resuspended in 300μl TM and extracted twice with an equal volume of chloroform. 25mM EDTA and 0.5M NaCl were added to the aqueous phase and extracted with an equal volume of phenol followed by extraction with an equal volume of chloroform. The aqueous phase was precipitated with 2.5volumes of 100% ethanol on ice for 10 minutes. The DNA was pelleted by centrifugation, washed with 70% ethanol, air dried and resuspended in 50μl of TE.

7. cDNA Insert Subcloning

The size of insert contained in each phage clone was determined by restriction enzyme digestions. Southern blot analysis was performed to confirm the hybridisation of the clones to the desired sequence.

The cDNA inserts were subcloned into the EcoRI site of pZero-2 using the materials and methods provided with the Zero Background/Kan PLUS Cloning Kit (Invitrogen). The kit uses a positive selection mechanism based upon interruption of the lethal *ccdB* gene by the insertion of a DNA fragment to allow survival of the bacterial cell carrying the vector. cDNA inserts were sequenced using the flanking Sp6 and T7 primer sites.

2.1.17 **Plasmid Constructs used for Molecular Analysis**

Table 2.1 lists the plasmid constructs, polymerase sites, restriction enzyme sites and probe fragments used for analysing gene-trap integrations. PT1.ATG was the gene-trap vector electroporated into R1 ES cells to generate gene-trap cell lines R68 and R124 (Forrester *et al.*, 1996). RACE clones are RACE-PCR products derived from R68 and R124 ES cell RNA respectively cloned directionally into pBluescript II KS-. cDNA clones are cDNAs isolated from the Clonetech Mouse Heart 5'- Stretch plus cDNA library, screened with R124 RACE trapped novel sequence KXE, cloned into pZerO-2. Polymerase sites flanking multiple cloning sites were used to sequence cDNA fragments. Probe fragments were used for Southern and Northern blot analysis to show the presence of trapped and vector sequences. Riboprobes were used for RNase protection assays to confirm the presence of fusion (endogenous-reporter gene) transcripts. The plasmid maps are listed in appendix I.

		Probe Fragment Generation			Riboprobe Generation		
Plasmid	Seq.Sites	Fragment	RE site	Size (kb)	Line	Pol	Size(nt)
PT1.ATG	-	En-2 exon	BamHI	0.5	-	-	-
PT1.ATG	-	En-2 intron	BamHI/HindIII	1.0	-	-	-
PT1.ATG	-	lacZ	EcoRI/ClaI	2.2	-	-	-
pB.R68.13	T7/T3	-	Kpn/SpeI	0.5	SpeI	T3	444
pB.R68.32	T7/T3	-	Kpn/SpeI	0.3	SpeI	T3	307
pB.R124.4	T7/T3	-	Kpn/SpeI	0.5	SpeI	T3	500
pB.R124.11	T7/T3	KX	KpnI/SpeI	0.5	SpeI	T3	463
pB.R124.11	T7/T3	KXE	KpnI/SpeI/EarI	0.3	-	-	-
pZ.TRI	T7/SP6	TRI	EcoRI*	1.6	-	-	-
pZ.TnI	T7/SP6	TnI	NotI	0.6	-	-	-
pZ.VRI	T7/SP6	VRI	EcoRI*	0.8	-	-	-

Table 2.1: Plasmid constructs used for molecular analysis.

PT1.ATG gene-trap vector; pB, plasmids generated by cloning RACE products into pBluescript II KS(-); pZ, plasmids generated by cloning cDNA fragments into pZero-2. EcoRI* could also be digested with SalI and NotI.

2.2 ES Cell Methodology

General reference was made to Hogan *et al.*, 1994 for ES cell methodology. All ES cell manipulations were performed in laminar flow sterile hoods. Sterility was maintained by using strict microbiological techniques which included wiping the hood down and spraying all items entering the hood with 70% industrial methylated spirits (IMS). Solutions used for ES cell culturing were filtered for sterility and tested for contaminations. Sterile disposable tissue culture grade plasticware (Corning) was used for culturing.

2.2.1 Reagents for ES Methods

Differentiation Inhibiting Activity/Leukemia Inhibitory Factor (DIA/LIF) was prepared by Douglas Colby and Derek Rout at the CGR by transient expression of murine or human DIA/LIF expression plasmids in COS-7 cells using the previously described method (Smith, 1991). Serial dilutions of the supernatant were tested on ES cells for their ability to maintain pluripotency. Routinely 100x strength of the minimal dilution of the preparation required to keep ES cells undifferentiated was used as the working concentration.

- 15% FCS ES cell culture medium: 1xGlasgow MEM (BHK21) (Gibco); 0.25% sodium bicarbonate (Gibco); 0.1xMEM non-essential amino acids (Gibco); 4mM glutamine(Gibco); 2mM sodium pyruvate(Gibco); 0.1mM 2-mercaptoethanol (Sigma); 15% foetal calf serum (FCS) (Globepharm, Surrey).
- All-trans Retinoic Acid: All-trans retinoic acid (Sigma) prepared as 20mg/ml stock in DMSO, used at 10^{-6}M in 5% FCS ES cell medium.
- Trypsin solution: 0.025% trypsin (Gibco); 0.1% chicken serum (Flow Labs) and 1.3mM EDTA disodium salt (Sigma) in PBS.
- Pen/Strep: 50 units Penicillin; 50μg/ml Streptomycin (Gibco) in PBS
- Gelatin: 0.1% gelatin (Sigma) in PBS.

2.2.2 **ES Cell Growth and Maintenance**

Undifferentiated ES cell cultures were grown in gelatinised plastic flasks in ES cell culture medium supplemented with DIA/LIF, as described by Smith 1991. Cultures were maintained in humidified incubators at 37°C under 7% CO_2. Cultures were checked every day and not allowed to grow past confluency. In practice cells were passaged every two days. Media from the cultures was removed by aspiration and the cells were washed twice with 5mls of PBS, prewarmed to 37°C. Cells were then dissociated by incubation with 1ml of trypsin at 37°C for 2-3min. The flask was tapped to ensure complete dissociation of the cells from the surface of the flask. This was monitored using an inverted Olympus microscope CK2. When a single cell suspension was reached, 4mls of ES cell medium were added to stop trypsinisaton. This single cell suspension was centrifuged for 5 minutes at 1200rpm, excess media was removed and the pellet was resuspended in 5mls of fresh culture media. Cell numbers were determined and $1x10^6$ cells were seeded to a new $25cm^2$ gelatinised flask containing 10ml of ES cell medium supplemented with DIA/LIF.

2.2.3 **Freezing ES Cells**

ES cells from a $25cm^2$ flask were trypsinised into a single cell suspension and then pelleted and resuspended in 0.5ml of freezing mix (10% dimethyl sulphoxide in ES cell medium) and rapidly aliquotted into Nunc cryotubes and placed at -80°C overnight before being transferred to a liquid nitrogen cell bank for long term storage.

2.2.4 **Thawing ES Cells**

Frozen ES cell vials were taken directly from the liquid nitrogen storage and quickly thawed in a 37°C water bath. The contents of the vial were transferred to a centrifuge tube containing 10ml of prewarmed ES cell medium and centrifuged at 1200rpm for 3 minutes. Excess media was removed and the pelleted cells were resuspended in 1ml of

ES cell medium. This 1ml cell suspension was then used to seed a 25cm^2 flask containing 9ml of ES cell medium. The medium was changed after 8 hours of culture to remove remove any dead cells. Subsequently the culture was maintained as described above (2.2.1).

2.2.5 **Exposure of ES cells to Retinoic Acid**

Gene-trap ES cells were exposed to RA over a period of 48hours to induce differentiation. The response of the trapped gene was monitored by quantifying the reporter gene activity at the protein and RNA level. Cultures were set up: 3x10^5 ES cells per 6-well plate for β-galactosidase (β–gal) assays and 1x10^6 cells were plated in 90mm tissue culture plates for northern blot analysis in 15% FCS+LIF. The next morning the media was changed to 5% FCS-LIF and 10^{-6}M RA was added to the cultures for 0, 6, 12, 24 and 48 hours. All timepoints were done in duplicate and the media was changed at each time point in all cultures to keep the activity of RA optimal. Control cells were grown in 15% FCS+LIF.

2.2.5.1 **ß-galactosidase Assay**

The cells were harvested by trypinisation, washed in phosphate buffered saline (PBS) and resuspended in 0.25M Tris pH7.5. After three cycles of freeze/thawing in liquid nitrogen, the samples were centrifuged at 13K for 5 minutes and the supernatant was transfered to a fresh tube. The protein concentrations in the supernatant were determined using the Biorad protein assay kit and comparing the optical density (595nm) of samples to a standard BSA protein concentration curve. The ß-galactosidase assays were performed using equivalent amounts of protein (250µg) in 0.5ml of ß-gal buffer (60mM Na_2HPO_4, 40mM NaH_2PO_4, 10mM KCl, 1mM $MgCl_2$, 50mM B-mercaptoethanol) containing 0.2mg O-Nitrophenyl B-D-

Galactosidase (ONPG) at 37°C overnight. The reaction was stopped and optical density was measured at 420nm.

2.2.5.2 **Northern Blot Analysis**

Cultures were set up in parallel to the protein assay cultures to measure the levels of *lac Z* transcripts. Total RNA was isolated at the appropriate timepoints using a modified guanidinium isothiocynate and phenol RNA isolation protocol (2.1.10). 15μg of RNA from each sample were used for nothern blot analysis following standard procedures (2.1.11). Blots were hybridised (2.1.13) with ^{32}P-labelled *lacZ* probe (2.1.12) to detect the level of the *lac Z* fusion transcripts. The blots were stripped (2.1.13) then hybridised with ^{32}P-labelled *actin* probe to standardise the amount of RNA loaded on the gel. The blots were reprobed with ^{32}P-labelled *RARβ* probe to confirm the activity of RA, RARß is directly regulated by RA at the transcriptional level. Phosphoimaging analysis allowed the northern blot signals to be measured and quantified relative to *actin* signals.

2.2.6 ***In Vitro* Cardiac Differentiation**

Embryonic stem (ES) cells were grown on gelatin in the presence of LIF until confluent. The cells were then trypsinised and made into a suspension of 3x10^{4} cells per ml of ES cell media supplemented with DIA/LIF. 300 cells (10μl of suspension) were used to prepare a single embryoid body in a hanging drop - multiple drops of suspension were placed on the lid of a bacterial petri dish which was then placed on the base of the dish containing 10ml of sterile water. The cells aggregate together, by the force of gravity, to form embryoid bodies. After 2 days in hanging drops the embryoid bodies were transferred to a bacterial dish containing ES cell media supplemented with Pen/Strep (DIA/LIF was added) and grown in suspension for a further 5 days. The medium was changed every second day. After 5 days in

suspension the embryoid bodies were plated on gelatin coated 24-well plates (whenever possible a single embryoid body was placed in a well) in ES cell media supplemented with Pen/Strep. Thereafter plated embryoid bodies were checked daily for beating cells. After overnight plating beating cardiomyocytes were found in approximately 80% of embryoid bodies.

2.3 Biological Specimens

2.3.1 Animal Maintenance and Breeding

Mice were housed in a constant light-dark cycle, 14 hours light and 10 hours dark, environment. A supply of water and chow food was available to the mice at all times. In these conditions females ovulate every 4-5 days and males tend to mate females in estrus during the middle of the dark period. Thus natural matings were set up overnight and females were examined for vaginal plugs (coagulation of semen proteins) the next morning. The presence of the plug was taken as 0.5 days of gestation. The gestational period for a mouse is 20 to 21 days depending on the particular mouse strain. Litters born from natural matings were left with the parents as standard practise until they reach 3 weeks of age when they are weaned i.e. sexed and separated from the parents. Tail tips are also taken at this age for genotyping purposes (Southern blot analysis 2.1.9). At 6 weeks of age, male and female mice have reached sexual maturity and can be used for mating purposes.

2.3.2 Breeding Protocol for the R124 Gene-Trap Line

In this study inbred strains, 129CGR and C57BL/6, and the outbred strain MF1 were used to assess the *in vivo* function of the trapped gene. Backcross matings were set up with'pure'background females and males heterozygous for the gene-trap integration. Intercross matings were set up between heterozygous male and female siblings from

the same litter. Genotyping of backcross litters was done by detecting the presence of *lacZ* by Southern blot analysis. Intercross litters were intially genotyped by quantifying the number *en-2* copies relative to two endogenous copies in wildtype animals by Southern blot analysis combined with Phosphoimaging analysis. This was verified by quantifying the number of *lacZ* copies in homozygotes relative to heterozygotes. Upon isolation of endogenous trapped sequences by RACE-PCR, a RFLP was determined which allowed an unambiguous distinction between wild type, heterozygous and homozygous animals by Southern blot analysis.

2.3.3 **Animals Housed in Metabolic Cages**

Metabolic cages are designed to collect urine samples and allow the volume of water intake to be measured. Test and control animals were housed in metabolic cages for 24 hour periods with water but no food in the same light/dark cycle as all other animals. Urine secreted by each animal was collected for volume and osmolarity measurements. Osmolarity of urine is a measure of the salt concentration present in urine. A osmometer OM801, calibrated using salt concentration standards, was used to measure urine osmolarity (Osmol/kg).

2.3.4 **Specimens for Histology**

Timed matings were set up for the collection of embryos or fetuses at specific stages of gestation. Pregnant females were culled by a schedule one method i.e. cervical dislocation and offspring explanted into cold phosphate buffered saline (PBS). All extraembryonic tissues were removed from embryos and fetuses (i.e. decidua, yolk sac, placenta), in addition fetuses (11 days *post coitum* and older) were decapitated before any experimental procedure. All specimens collected were checked for normality by examining stage specific features eg. beating and looped heart; number of somites; turning of the embryo; neural tube closure; number of pharyngeal arches; fore

and hind limb development. Adult tissues, collected from animals culled by a schedule one method, were dissected free of any connective tissue and fat. Organs were examined for any gross abnormalities and weighed. In parallel, embryonic and adult, samples were taken from wild type siblings as controls.

2.3.4.1 **TESPA coating slides**

Microscope slides were treated with:10% HCl in 70% ethanol for 10 seconds; rinsed in dH_2O for 10 seconds; dehydrated in 95% ethanol for 10 seconds; baked dry at 150°C for 5 minutes; cooled down to room temperature; dipped in 2% TESPA in acetone for 10s; rinsed twice in 100% acetone for 10 seconds each; rinsed in dH_2O for 10 seconds and baked dry overnight at 42°C. TESPA coated slides were stored at 4°C with silca gel.

2.3.4.2 **Wax Histology**

Dissected specimens were fixed in Bouin's fixative (BDH) overnight at room temperature. After fixation the specimens were transferred to 70% ethanol for storage until the waxing procedure. For wax embedding specimens were dehydrated through a gradient of alcohols (70%, 95%, 95%, 100%, 100% ethanol) for 1 hour each and then cleared in xylene for 1 hour. Following clearing the xylene was replaced by molten wax and the specimens incubated at 55°C for a total of 3 hours. Wax was changed every hour (for specimens older than 11 days *post coitum* and adult tissues the last wax was extended to overnight at 55°C). The specimens were then orientated and embedded in fresh wax. This was left to set overnight before sectioning.

Sections were cut at a thickness of 7-10μm using a microtome, floated on a waterbath at 42°C to flatten and mounted on TESPA coated slides (2.3.4.1). The slides were then dried on a hotplate at 42°C for 30 minutes and baked overnight at 42°C.

Following this the sections were counterstained with haematoxylin and eosin (2.3.4.3).

2.3.4.3 **Counterstaining Wax Sections with Haematoxylin and Eosin**

Wax sections were treated as follows: dewaxed in xylene for 10 minutes; rehydrated through alcohol series (100% , 95%; 70% ethanol) for 5 minutes each; rinsed in water for 5min; stained with haematoxylin [Sigma supplied stock solution diluted 1/12 in water] for 5 minutes; rinsed in water till clear; stained with eosin (BDH)[0.1% solution in 70% ethanol] for 5 minutes; dehydrated through alcohol series (95%,100% ethanol) for 10 minutes each; xylene/histoclear for 10 minutes and coverslipped using DPX mountant (BDH).

2.3.4.4 **Cryostat Sections**

Specimens for cryostat sectioning were freshly dissected, placed in OTC, immediately frozen in liquid nitrogen and stored at -80°C. Specimens were removed from -80oC and placed in the cryostat at the cutting temperature (-15 to -20°C) for 1-2 hours to equilibrate before sectioning. 7-10μm thick sections were cut and lifted onto TESPA coated slides (2.3.4.1) prewarmed to room temperature. The sections were then dried onto the slides at room temperature (can be stored at -20°C if necessary at this stage) before staining for x-gal activity (as described below). Sections stained for x-gal activity were then counterstained with Haematoxylin and Eosin (2.3.4.6).

2.3.4.5 **X-gal Staining**

The x-gal staining protocol involves fixing tissue in a gluteraldehyde fixative followed by rinsing in a detergent phosphate wash buffer to remove the fix and finally staining with a x-gal (5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside) stain. This protocol can be adapted to stain cells, embryoid bodies, cryostat sections or whole embryos

(younger than 11.5 days *post coitum*) by varying the time for fixation and washing. Stated below is the protocol used for staining whole 8.5 days *post coitum* (d.p.c.) embryos. The times for fixing and washing cells, embryoid bodies, cryostat sections or whole embryos is tabulated.

8.5 d.p.c. embryos were rinsed in phosphate buffer (pH7.3) followed by fixation in fix solution (0.2% gluteraldehyde; 5mM EGTA (pH7.3); 2mM $MgCl_2$ in 0.1M sodium phosphate pH7.3) for 15 minutes at room temperature. The embryos were then washed in wash buffer (20mM $MgCl_2$; 0.01% deoxycholate; 0.02% nonidet in 0.1M sodium phosphate pH7.3) three times for 15 minutes each at room temperature and stained with x-gal stain (1mg/ml X-gal dissolved in di-methyl formamide; 250mM potassium ferrocyanide; 250mM potassium ferricyanide in wash buffer) overnight at 37^oC. After staining, stain was replaced with wash buffer and the samples stored at 4^oC.

Tissue	Fix Time (minutes)	Wash Time (minutes)
Cells	5	5
Cryostat sections	5	5
Embryoid Bodies	10	10
Embryos (d8.5-10.5 p.c.)	15 - 30	15 - 20

2.3.4.6 **Counterstaining Cryostat Sections with Haematoxylin and Eosin**

Cryostat sections were treated as follows: stained with haematoxylin [Sigma supplied stock solution diluted 1/12 in water] for 2.5 minutes; rinsed clear in water; stained with eosin (BDH)[0.1% solution in 70% ethanol] for 1 minute; dehydrated through alcohol (95%, 100% ethanol) for 10 minutes each; xylene/histoclear for 10 minutes and coverslipped counterstained using DPX mountant (BDH).

2.3.4.7 **Microscopy and Photography of Specimens**

Whole tissue specimens were examined using Olympus SZ40 and Olympus SZH10 dissection microscopes and photographed using an Olympus C-35AD-4 camera. Sections were examined using a Wild Leitz Laborlux S microscope and photographed using a Wild Leica MPS52 camera combined with a Wild Leica MPS46 exposure box.

Chapter 3

3.0 **Preliminary Characterisation of Gene-Trap Integrations R68 and R124**

Introduction

ES cell lines, R68 and R124, were isolated in a gene-trap screen designed to identify genes involved in the process of differentiation (Forrester *et al.,* 1996). *In vivo* analyses showed that the reporter gene was expressed in the hearts of tetraploid aggregation chimeras generated from both gene-trap ES cell lines and were therefore chosen for further analysis. This chapter describes the preliminary characterisation of the R68 and R124 gene-trap integrations. The preliminary characterisation involved assessment of the ES cell lines: *in vitro* by confirming their response to RA-exposure and their ability to differentiate into cardiomyocytes; *in vivo* by transmitting the integrations through the germline to determine whether the cell lines had retained their pluripotency and to characterise reporter gene activity; and molecular analysis to determine whether the endogenous trapped gene sequences could be cloned.

3.1 **Response of Reporter Gene Activity to Retinoic Acid**

The gene-trap ES cell lines were grown under the same conditions as in the original RA screen (Forrester *et al.*, 1996) and exposed to 10^{-6} M RA (Chapter 2, Section 2.2.4). Equal quantities (250μg) of protein were assayed to determine β-gal activity. As observed in the original screen both cell lines showed a repression of β–gal activity after 48 hours of RA exposure (Table 3.1). The data also revealed slightly elevated levels of β–gal activity after 12 and 6 hours of RA-exposure in gene-trap cell lines R68 and R124 respectively (Table 3.1).

Table 3.1: β-gal Response to RA-exposure

Cell Line	RA Treatment (hours) 0	6	12	24	48	Reduction 0 vs 48
R68	506 ± 14	561 ± 55	630 ± 45	388 ± 9	251 ± 3	0.50
R124	718 ± 16	767 ± 1	629 ± 19	433 ± 40	259 ± 9	0.64

Values are the mean of four independent OD_{420} (10^{-3}) measurements ± standard error
vs, versus.

RNA isolated from parallel cultures was analysed by northern blotting combined with phosphoimaging analysis. The presence of the fusion transcript was detected with a *lacZ* probe (northern blots not shown) and quantified relative to *β-actin* loading controls. Changes in transcript levels paralleled the changes observed at the protein level (Table 3.2). Elevated levels of RARβ transcripts were detected after 6 hours of RA-exposure (Table 3.2) confirming that the ES cell lines were responding to RA-activity (Manglesdorf *et al.*, 1994).

Table3.2: Quantitation of Fusion Transcripts

Cell Line	Transcript	RA Treatment (hours) 0	6	12	24	48	Reduction 0 vs 48
R68	*lacZ*	92 ± 6	139 ± 14	183 ± 11	88 ± 4	62 ± 4	0.33
	RARβ	15 ± 2	35 ± 3	na	na	na	na
R124	*lacZ*	16 ± 2	47 ± 3	25 ± 2	18 ± 2	8 ± 2	0.50
	RARβ	33 ± 5	89 ± 5	na	na	na	na

Values are the mean of two independent phosphoimaging measurements (10^{-2}) ± range.
na, not applicable; vs, versus.

3.2 *In Vitro* Differentiation

To assess the specificity of the reporter gene activity in R68 and R124 gene-trap integrations to cardiogenesis both ES cell lines were differentiated *in vitro* to generate beating cardiomyocytes.

ES cells differentiate into many lineages including beating cardiomyocytes (Sanchez *et al.,* 1991) which express cardiac specific genes in a temporal pattern resembling early cardiogenesis. In addition, electrophysiological measurements of these beating cells

show the presence of cardiac specific ion channels and characteristic cellular potentials, which confirm that the beating cells are cardiomyocytes (Wobus *et al.,* 1991).

3.2.1 **Differentiation of Embryoid Bodies**

The *in vitro* generation of beating cardiomyocytes was achieved using the hanging drop protocol to prepare embryoid bodies and then allowing these to differentiate in suspension by removing DIA/LIF (Chapter 2 Section 2.2.6).

After 5 days in suspension the embryoid bodies (EBs) were plated onto gelatin coated tissue culture dishes and monitored 1, 4 and 8 days after plating. One day after plating the EBs had adhered to the gelatin coated surface and cells could be observed spreading away from the large multilayered central embryoid body (Figure 3.1A). Two different waves of cells migrating from the EB were observed: the first wave were endodermal-like flat cells; the later second wave, (sitting on top of the endodermal-like cells), were slightly rounded cells. Beating cells were found in the central part of the EB usually in a single site covering up to 50% of the EB. In addition, multiple smaller sites were observed beating in the second wave of cells, often two sites close together could be found beating synchronously. Approximately 5 - 25% of the total embryoid body area was found to be beating in most cases. The morphology of beating cells and regions are different from surrounding cells, cells are more round and seem to overlap each other generating a network/spiraling pattern (Figure 3.2 and 3.3). Beating cardiomyocytes were observed in 80% of all embryoid bodies after overnight plating.

Four days after plating the embryoid bodies were found to have spread more and the central mound of cells was flatter (Figure 3.1B). The beating regions had also increased proportionally. No significant change in the number of embryoid bodies with beating cardiomyocytes was observed.

Figure 3.1: *In Vitro* Differentiation of ES cell derived EBs.

A. EB plated for 1 day, first wave of cells emigrating;

B. EB plated for 4 days, second wave emigrating.

Scale bar, 100μm.

Figure 3.1A

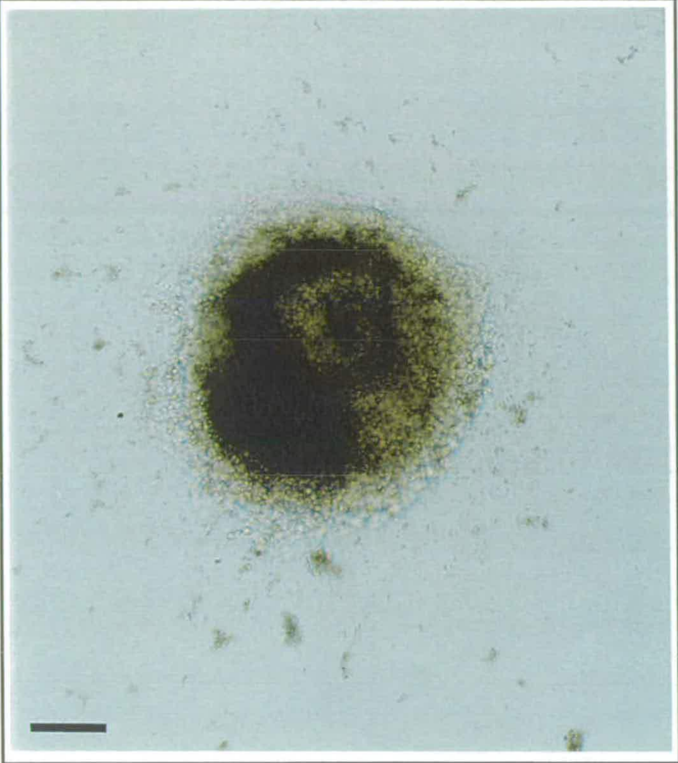
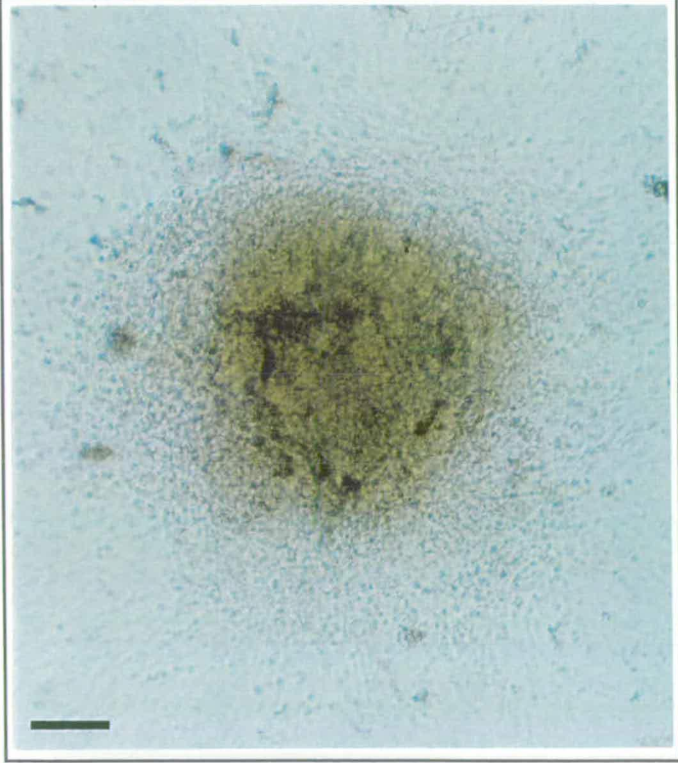


Figure 3.1B



Eight days after plating the embryoid bodies had spread further, the central region of embryoid bodies was much flatter, possibly only two to three layers thick. Again no significant change in the number of embryoid bodies with beating cardiomyocytes was observed but by this stage other differentiated cell types such as neuronal cells and haematopoetic cells could be distinguished morphologically.

3.2.2 β-gal Activity in the *In Vitro* Differentiated R68 Cell Line

In vitro undifferentiated R68 ES cells, grown in the presence of LIF, stained for β-gal activity display a single cytoplasmic spot of expression. Embryoid bodies in hanging drops generated from R68 ES cells, in the presence of LIF, also display reporter gene expression in all cells as a single deposit.

In suspension, in the absence of LIF, reporter gene activity becomes restricted to multiple clusters of cells within the embryoid body. One day after plating, β-gal activity was found to be coincidental with cells that were beating. A simple drawing of the cells which were beating in the culture prior to staining was used to correlate beating cells with stained cells. Expression was also found in some regions of unspread non-beating cells in embryoid bodies, it is most likely that these were undifferentiated ES cells. This could be tested by staining the cells for alkaline phosphatase activity. No beating or stained cells were observed in the first wave of endodermal-like cells. X-gal staining after four days of plating was the same as day 1, however the number of non-beating sites stained were reduced. After eight days of plating reporter gene expression was restricted exclusively to the beating cardiomyocytes derived from the R68 ES cells (Figure 3.2). No staining was observed in other differentiated cell types such as neuronal or haematopoetic cells.

In conclusion, reporter gene activity is detected in undifferentiated R68 ES cells and exclusively *in vitro* in R68 ES cell derived beating cardiomyocytes.

Figure 3.2: R68 ES cell derived Beating Cardiomyocytes.
Cardiomyocytes, stained for reporter gene activity, are morphologically distinct.
Scale bar, 10μm.

Figure 3.3: R124 ES cell derived Beating Cardiomyocytes.
Two adjacent groups of cardiomyocytes, beating in synchrony in culture, stained for reporter gene activity. Scale bar, 10μm.

Figure 3.2

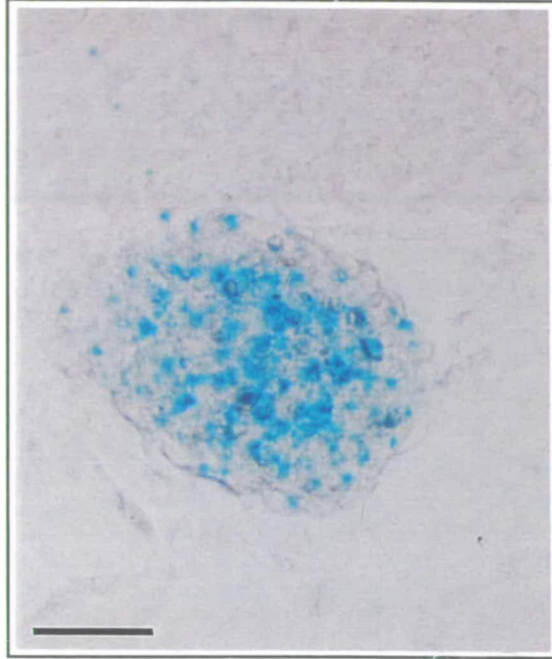
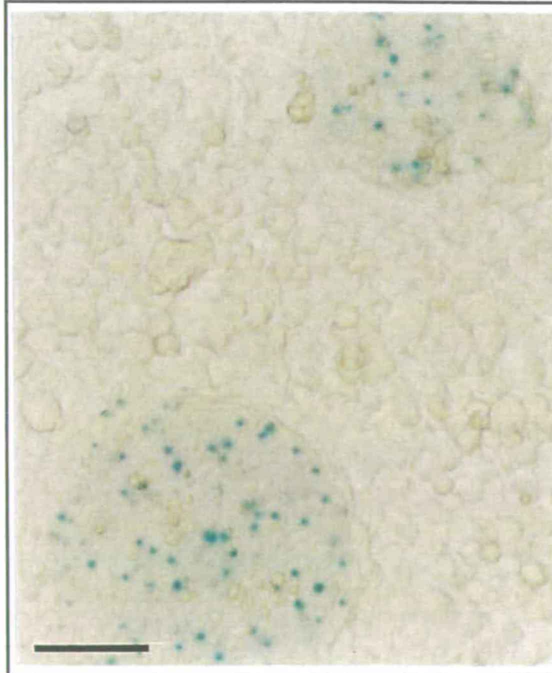


Figure 3.3



3.2.3 β-gal Activity in the *In Vitro* Differentiated R124 Cell Line

Undifferentiated R124 ES cells stained for β-gal activity displayed a single cytoplasmic spot of expression. The pattern of reporter gene expression observed in embryoid bodies derived from R124 ES cells over eight days of plating is similar to that observed with R68 ES cell line. After eight days, reporter gene expression is detected solely in the *in vitro* differentiated beating cardiomyocytes (Figure 3.3).

3.3 Germline Transmission of the Gene-Trap Integrations

3.3.1 R68 Gene-Trap Integration

Attempts have been made to transmit the R68 gene-trap integration through the germline to generate animals heterozygous for the integration but these have been unsuccessful. R68 ES cells were injected into a total of 440 blastocysts collected from C57BL/6 females, by Kathryn Newton and Jan Ure. Six founder chimeras (4 males and 2 females) resulted from a total of 86 animals born but none successfully transmitted the integration through the germline. Morula aggregations (46) gave 2 chimeras (1 male and 1 female) out of 17 animals born. The chimeras generated from morula aggregations were also unsuccessful in transmitting the integration through the germline. The low number of chimeras developing to term suggests that the condition of the ES cells is suboptimal, and that high chimerism resulted in midgestational death. The introduction of genetically engineered vectors into the ES cell genome can cause large rearrangements of the genome which lead to lack of germline transmission (Yu *et al.,* 1996). To eliminate this possibility R68 ES cells were karyotyped (by Derek Rout) and no gross disturbance of the ES cell genome was observed but subtle rearrangements cannot be excluded (See Section 3.5). As a last resort *in vitro* fertilisation with sperm isolated from R68 founders was attempted (Hogan *et al.*, 1994). Although the sperm derived from chimeras seemed normal and motile under the

light microscope this also proved to be unsuccessful. The lack of germline transmission therefore limits future experiments *in vivo* with this gene-trap line.

3.3.2 **R124 Gene-Trap Integration**

R124 ES cells were injected into C57BL/6 blastocysts by Jan Ure to generate chimeras. Three male chimeras 287, 288 and 289 (strong chimeras as predicted from coat colour) resulted from 16 independent blastocyst injections. Heterozygous embryos, collected from MF1 females mated to each chimera, displayed reporter gene activity restricted to the developing heart (Figure 3.4C), which had been anticipated from the reporter gene activity observed in tetraploid chimeras generated from R124 ES cells (Forrester *et al.,* 1996). Chimera 289 was backcrossed onto both inbred and outbred genetic backgrounds. To date the R124 integration has been backcrossed to C57BL/6, 129/CGR and MF1 to F6, F4 and F1 generations respectively.

3.4 ***In Vivo* Reporter Gene Activity in R124 Heterozygotes**

3.4.1 **Reporter Gene Activity Controls**

All expression studies were done on animals that had been backcrossed to at least the F2 generation. Patterns of reporter gene activity were verified by staining at least ten embryonic specimens and at least 3 tissue/organ specimens for each stage on inbred and outbred mouse backgrounds. For every heterozygous specimen stained at least one wild type sibling/sibling tissue was stained as a control for reporter gene activity. Background endogenous galactosidase activity was kept to a minimum by staining at pH 7.3 (Chapter 2, Section 2.3.4.5) which is the optimal for the bacterial reporter gene activity (Alam *et al.,* 1990), in contrast to mammalian galactosidase which has a pH optimum of 5.4. However, endogenous galactosidase activity was observed in wild type kidney, testis, ovaries and gut tissue in inbred and outbred heterozygous

specimens. This activity was observed as diffuse stain and could be distinguished from the punctate staining pattern displayed by the bacterial reporter gene.

3.4.2 Reporter Gene Activity during Mouse Embryogenesis

Whole embryos and cryostat sections of R124 heterozygotes were stained for reporter gene activity. During embryogenesis reporter gene activity was detected as early as 7.0 days*post coitum* (d.p.c.) (before any cardiac mesenchyme is determined) in single cells of the embryonic ectodermal layer (Figure 3.4A). These single cells could be pluripotent progenitor cells akin to ES cells and would be consistent with expression of the reporter gene in undifferentiated ES cells. At 7.5 d.p.c. reporter gene activity was detected in single cells within the cardiac mesoderm and along the whole length of the neuroectoderm (not shown). Once the primitive heart tube has formed and begins to loop (8.5 d.p.c.) an increasing number of cells within the developing heart display reporter gene activity (Figure 3.4B). By this stage the neural plate begins to fold and form the neural tube and it was observed that reporter gene activity was lost in the closed parts of the neural plate (not shown). At 9.0 d.p.c. of embryogenesis reporter gene activity was detected exclusively in the developing heart both in atria and ventricles (not shown). Exclusive expression of the reporter gene in the heart of midgestation fetuses was demonstrated by wholemount x-gal staining of a 10.5 d.p.c. fetus (Figure 3.4C). At 19 d.p.c. expression of the reporter gene can also detected in the choroid plexus of the brain. Cryostat sections of the heart (from 10 to 19 d.p.c.) reveal that reporter gene activity is detected in all cardiomyocytes in the myocardium and endocardium (not shown).

3.4.3 Reporter Gene Activity After Birth

After birth reporter gene activity was studied by staining 7-10μm cryostat sections of organs and tissues dissected from heterozygote males and females. Stages examined

Figure 3.4: Reporter Gene Expression during Embryogenesis

Sagittal cryostat sections stained for reporter gene expression:

A1. day 7.0 p.c. embryo, punctate expression in ectodermal cells and some staining in the mesoderm and extraembryonic amnion region ;

A2. orientation of A1, boxed;

B1. day 8.5 p.c. embryo, punctate expression in neural epithelia and some cells of developing looped heart;

B2. orientation of B1, boxed;

C. wholemount stained day 10.5 d.p.c. fetus, expression exclusive to the heart (expression in nasal placode background).

am, extraembryonic amnion; ec, ectoderm; me, mesoderm; ne, neural epithelia; dlh, developing looped heart; arrow, anterior to posterior orientation of embryo.

Scale bar, 50μm.

Figure 3.4A1

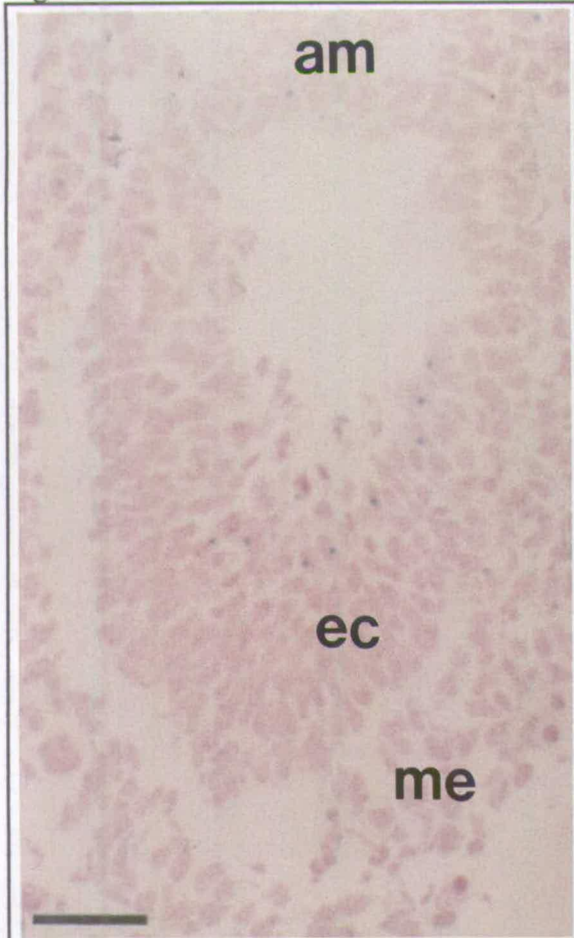


Figure 3.4B1

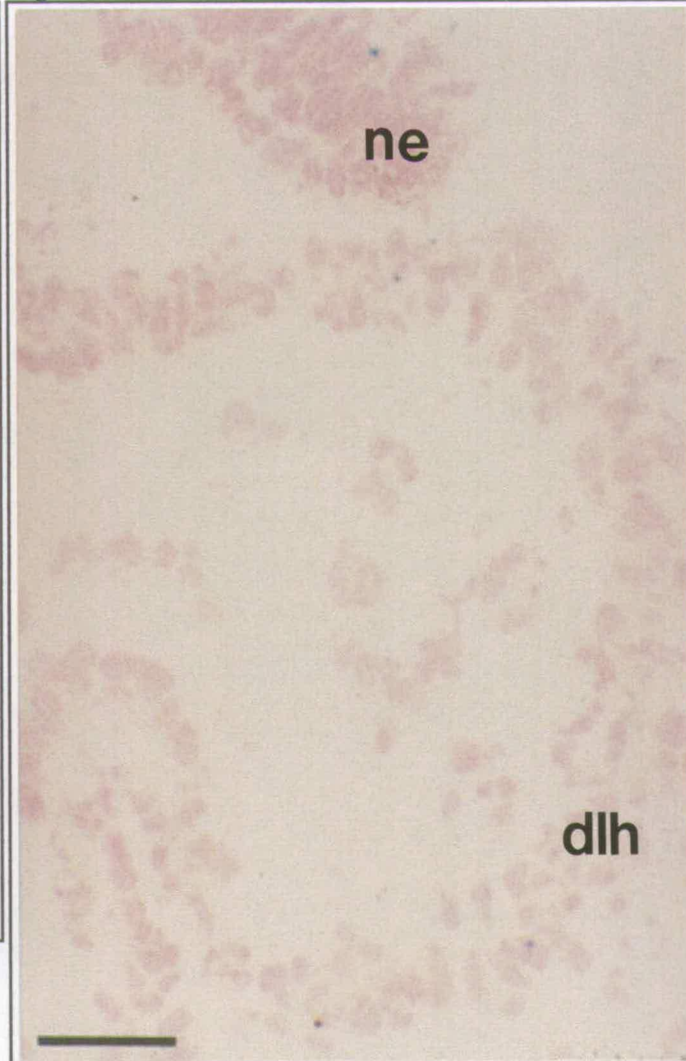


Figure 3.4A2

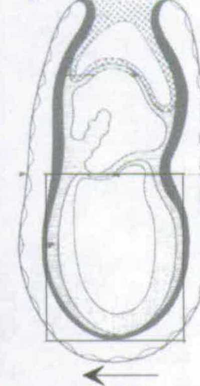


Figure 3.4B2

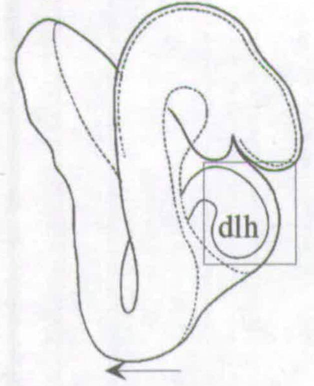
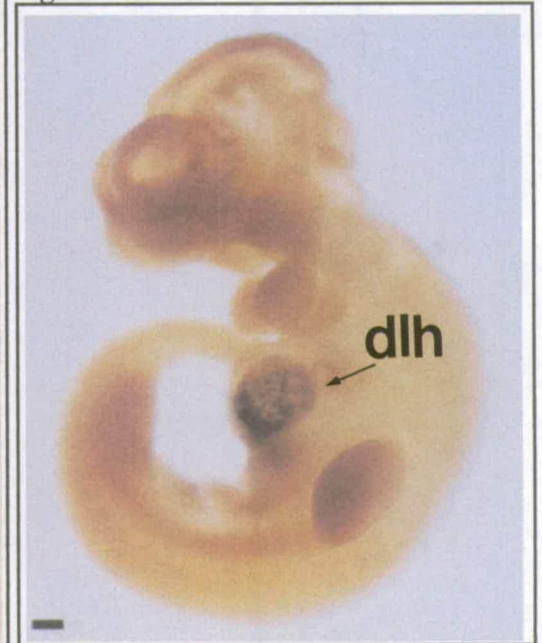


Figure 3.4C



included: 1 to 14 days*post partum* (d.p.p.) (pups collected daily); 21 to 49 d.p.p. (pups collected every 7 days) and 2 to 10 months p.p. (animals collected every month). To keep illustrations to a minimum temporal changes of reporter gene activity in each tissue have not been shown, instead adult tissues (10 weeks old) have been shown (Figure 3.5).

3.4.3.1 **Heart Expression**

Newborn pups, 1 and 2 d.p.p., continued to display reporter gene activity throughout the heart. In contrast, reporter gene activity could not be detected in the heart at 9 d.p.p. (2 males and 2 females from C57BL/6 background and 2 males from the MF1 background were examined). Examination of serial sections and whole hearts stained for reporter gene activity from 6 and 7 d.p.p pups showed increasingly patchy reporter gene activity in the heart up until 9 d.p.p. when no activity was detectable. At 11 d.p.p reporter gene activity reappeared as patches in the heart and by 14 d.p.p activity of the reporter gene was detected throughout the heart. Homogenous activity of the reporter gene in the heart remained constant after this stage and into adulthood (6 weeks to 18 months old tested) (Figure 3.5A).

3.4.3.2 **Kidney Expression**

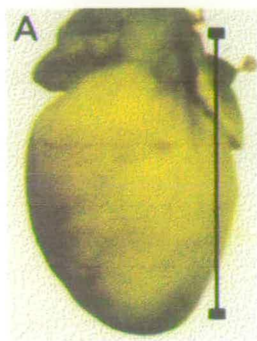
Reporter gene activity was first detected in the proximal tubules of the kidney at 6 d.p.p. and continued to adult stages (Figure 3.5C). Background diffuse staining due to the activity of endogenous galactosidases in the kidney was observed but this was eliminated by comparison to wild type controls.

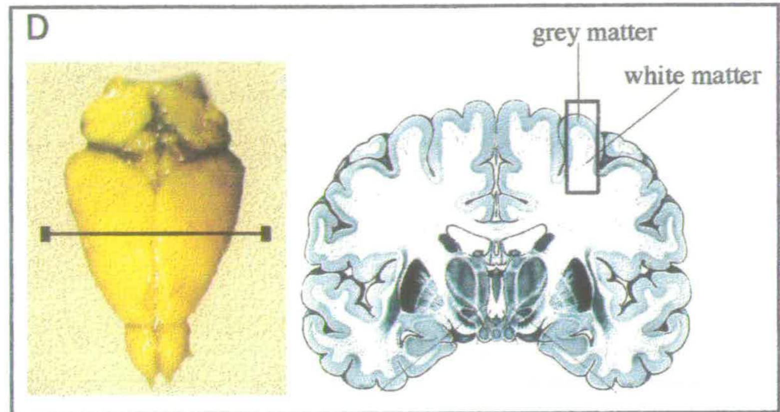
3.4.3.3 **Gonad Expression**

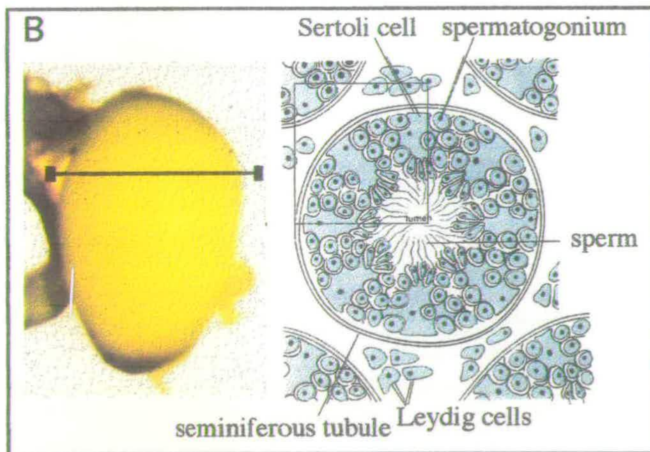
By 9 d.p.p reporter gene activity was detected in mesenchymal cells and primordial oocytes of the ovary and in mucosal cells of the oviduct adjacent to the ovary in female

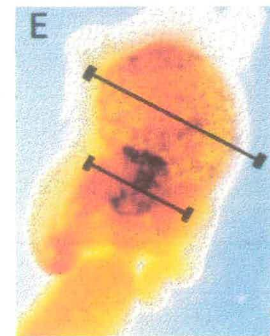
Figure 3.5: Reporter Gene Expression in Adult Tissues
Cryostat sections stained for reporter gene expression:
A. cardiomyocytes (arrowhead) in the heart ventricle, longitudinal section;
B. Sertoli cells (arrowhead) in the cortex of seminiferous tubules in the testis, transverse section;
C. proximal tubules (arrowhead) in the cortex of the kidney, longitudinal section;
D. grey (arrowhead) and white matter of the brain, coronal section;
E. mesenchymal cells (arrowhead) in the ovary, transverse section;
F. mucosal cells (arrowhead) in the oviduct, transverse section.
Scale bar, 50mm.

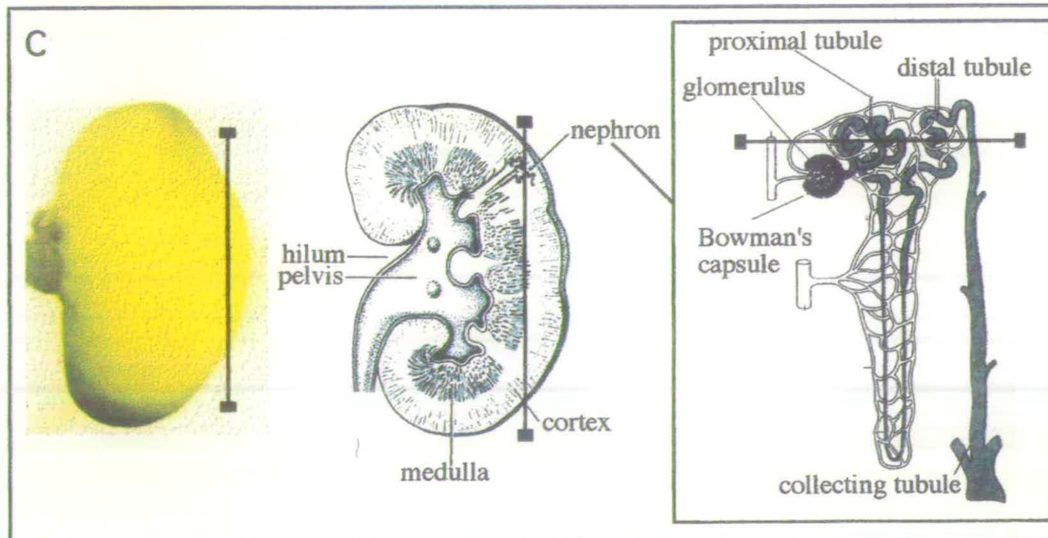
Orientation and plane of section (⊢⊣) for each organ. note: panel E includes oreintation and plane of section for Figure 3.5E and F, the ovary and oviduct stained for reporter gene expression here. Schematics representing histological section (box) through the brain, testis and kidney to clarify sites of reporter expression.











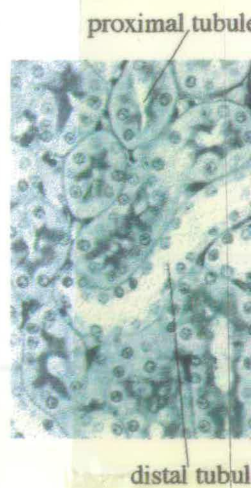


Figure 3.5A

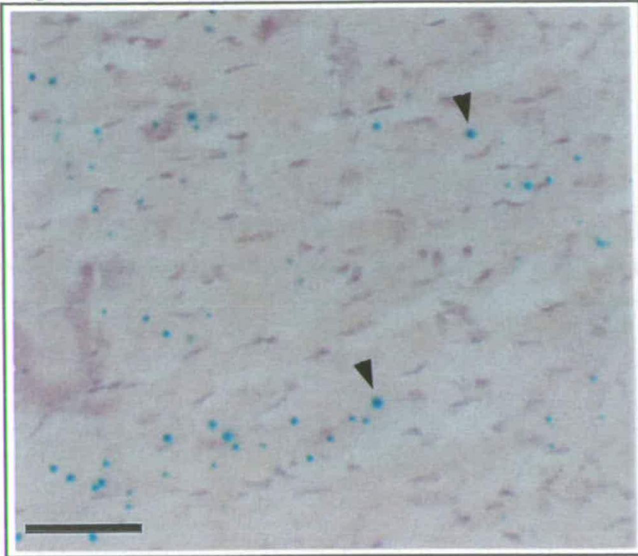


Figure 3.5D

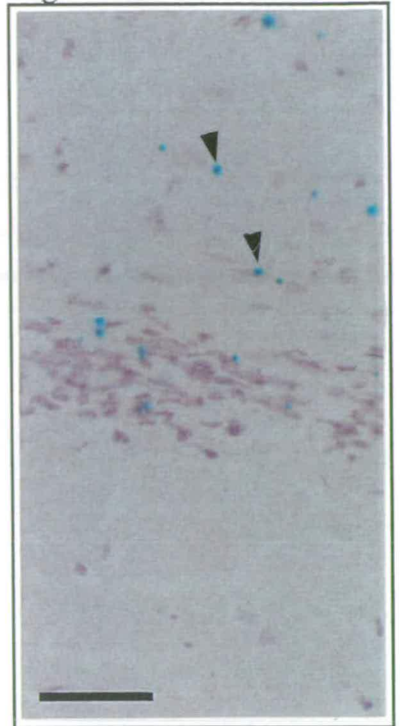


Figure 3.5B

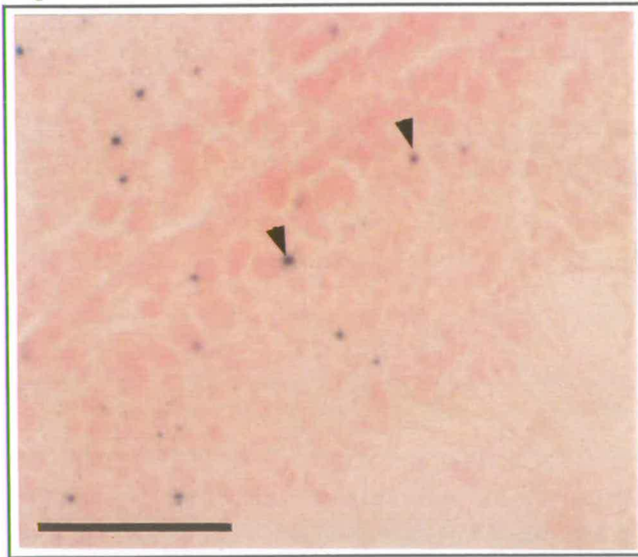


Figure 3.5E

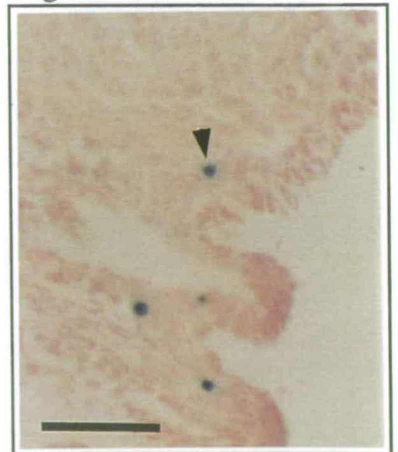


Figure 3.5C

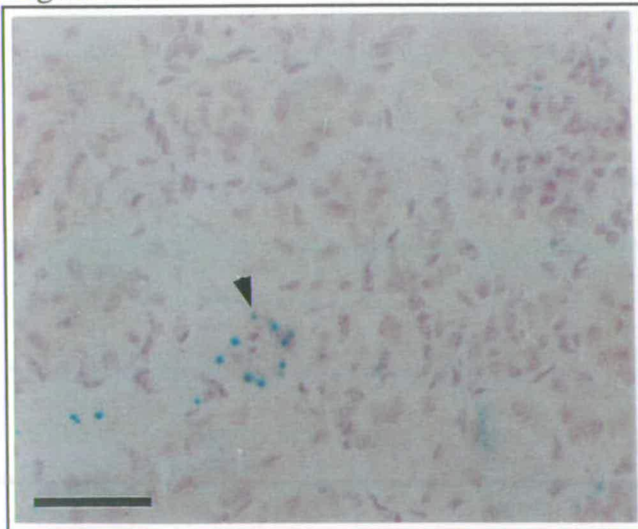
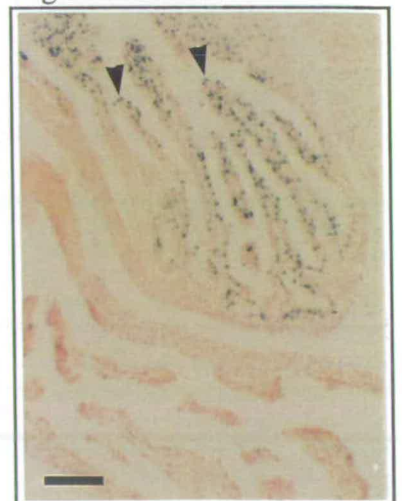


Figure 3.5F



pups. In males reporter gene activity was detected in the Sertoli cells of the testis at the same stage. Activity in the ovaries (Figure 3.5E), oviduct (Figure 3.5F) and testis (Figure 3.5B) continued to be detected in the adult.

3.4.3.4 Brain Expression

Reporter gene activity continued to be detected in the choroid plexus after birth. By 35 d.p.p. reporter gene activity was detected throughout the brain including the choroid plexus. The density of cells stained in the grey matter of the brain was greater than that in the white matter (Figure 3.5D).

3.4.3.5 Non-Expressing Tissues

No reporter gene activity was detected in striated skeletal muscle, smooth gut muscle, smooth bladder muscle, liver, thymus, lung or spleen (not shown).

3.5 Chromosomal Mapping

Using the PT1-ATG gene-trap vector (9.9kb) as a fluorescently labeled probe the gene-trap integrations R68 and R124 were mapped to mouse chromosomes 11 and 5 (band G2) respectively in ES cells by FISH (Fluorescent *in situ* Hybridisation) by Muriel Lee, MRC HGU. The chromosomes were initially identified by their G-banding pattern and then confirmed by chromosome specific painting (Rabbits *et al.*, 1995). FISH analysis of the R124 gene-trap integration is shown (Figure 3.6). This level of analysis revealed the vector had integrated into a single site in the ES cell genome in both gene-trap cell lines and no rearrangements of the genome had occured. Searching the Mouse Genome Database revealed no candidate genes on chromosome 5 or 11 which are expressed in the developing heart similar to the reporter gene in the R68 and R124 gene-trap integrations.

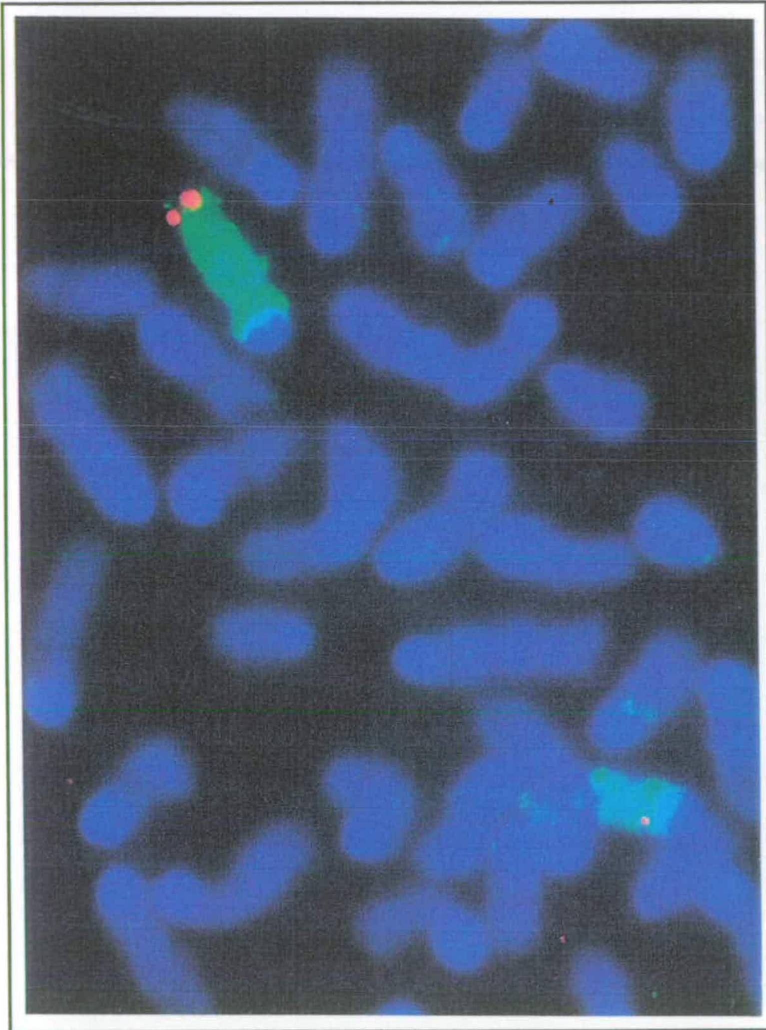
Figure 3.6: Chromosomal Mapping

Localisation of the R124 gene-trap integration by FISH analysis. Metaphase chromosomes hybridised with a fluorescent-labelled probe to detect integration of the gene-trap vector.

Orange spots, hybridisation signal;

green, chromosome paint specific to chromosome 5.

Figure 3.6



3.6 Detection of Fusion Transcripts

The expression of fusion transcripts in gene-trap cell lines R68 and R124 was detected using *lacZ* as a probe. A 4.4 kb fusion transcript was detected by northern blot analysis of total RNA isolated from both R68 and R124 ES cells (Figure 3.7). The *lacZ* transcript is approximately 3.3kb, expressed in the R121 gene-trap cell line, suggesting 1.1 kb of endogenous trapped gene transcript is fused to the *lacZ* transcript in R68 and R124 gene-trap integrations. The detection of transcripts on a northern blot indicates that the trapped sequences are actively transcribed.

3.7 Identification of Trapped Sequences using RACE PCR

5' rapid amplification of cDNA ends polymerase chain reaction (5'RACE PCR) allows the isolation of unknown sequences 5' to known sequences within the fusion transcript (2.1.14). It relies upon reverse transcription from a known stretch of sequence using a complementary primer to generate a cDNA which can subsequently be tailed with an adaptor sequence and amplified using PCR. The RACE PCR products can be size selected, purified and subcloned.

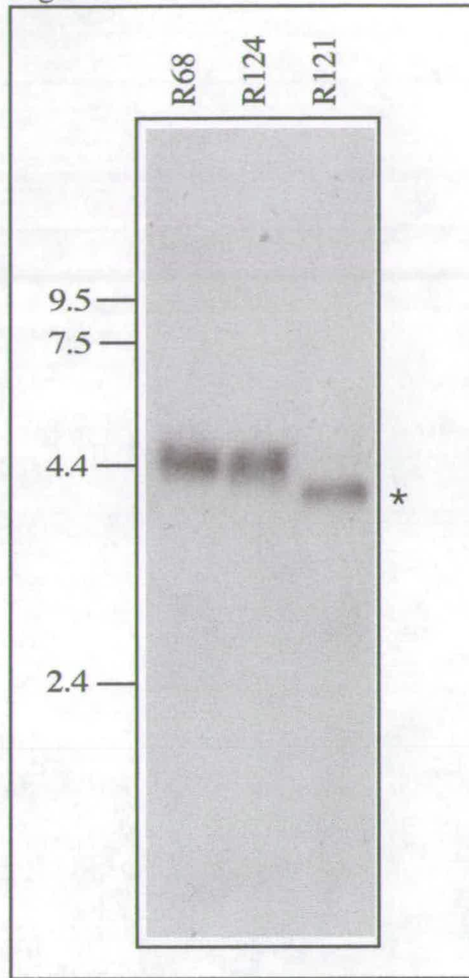
The products from RACE cloning of R68 and R124 gene-trap ES cell RNA were sequenced and grouped into groups I to IV according to sequence similarity 5' to the vector *en-2* SA sequence (Table 3.3).

Group I showed the intact splice acceptor sequence indicating no splice event had occured. This type of clone could arise by the amplification of pre-spliced mRNA. Group II showed vector-derived *en-2 SA* sequence spliced to vector-derived *en-2* intron at position 435. Group III showed vector-derived *en-2 SA* sequence spliced to vector-derived *en-2* intron at position 32. Detection of vector-derived *en-2* intron sequences fused to vector-derived *en-2* SA suggests the presence of cryptic SDs in the vector-derived *en-2* intron sequence. Activation of these cryptic SDs was consistent

Figure 3.7: Detection of Fusion Transcript

Detection of fusion transcripts by northern blot analysis of total RNA isolated from gene-trap cell lines, R68, R124 and *R121 using ^{32}P-labelled *lacZ* probe.
*expresses minimal *lacZ* transcript 3.3kb.

Figure 3.7



RACE Clone			Splice Junction Sequence		
Group	Name	Size (bases)	SJ: ← trapped sequence | *en-2* exon →	Trapped Sequence Homology	Splice Event
I	R68.14	200	gggaaagaggagtacacaaccagGTCCCAGGTCCCG	vector-derived *en-2* intron	none
	R124.9	700			
II	R68.19	400	gacccttttgggtttgccctttgGTCCCAGGTCCCG	vector-derived *en-2* intron	cryptic SD 435
	R68.32	307			
	R124.2	400			
	R124.4	500			
	R124.12	300			
III	R124.11	463	TTGGGCTGattcatgggaagaggaaccgaaaGTCCCAGGTCCCG	vector-derived *en-2* plus novel sequence	cryptic SD 32
IV	R68.1	600	ctattctgaccggccttcttcagGTCCCAGGTCCCG	endogenous *en-2* exon 2 [gb/L12705/MUSEN2AB]	N/A
	R6813	500			

Table 3.3: RACE products from gene-trap ES cell lines, R68 and R124, generated by electroporating the PT1.ATG vector into R1 ES cells Clones grouped (I to IV) according to homology of the trapped sequence 5' to the vector derived *en-2* SA sequence.
SJ, splice junction; SD, splice donor; novel sequence underlined; intron sequence lower case; [ ] genebank accession number.

with the gene-trap vector integrating into an exon (Chapter 4, Section 4.3) and has been observed in other gene-trap cell lines (personal communication Bill Skarnes). Integration of a gene-trap vector into an exon would presumably result in the lack of an immediate SD of the endogenous trapped gene upstream from the vector-derived *en-2* SA and as a consequence the splicing mechanism uses the nearest upstream SD. The cryptic SD at position 435 (TTGGTAAGA) has 67% identity and the cryptic SD at position 32 (AAAGTATGT) has 78% identity with the consensus SD sequence (C/A AGGTA/G AGT) making them potential targets for the splicing machinery and as observed they can be actively used (Figure 3.8). Cryptic splicing to position 435 results in the trapped sequence 435 bases 5' to the vector-derived *en-2* SA sequence. RACE clones derived from the R68 RNA were not long enough to identify the trapped sequence. However, cryptic splicing to position 32 results in the trapped sequence 32 bases 5' to the vector-derived *en-2* SA sequence making the identification of the trapped sequences more attainable. Thus sequences trapped by the R124 gene-trap integration have been identified (Chapter 4). In the R124.11 RACE clone 23bases of intron separate the endogenous trapped sequence and the vector-derived *en-2* SA sequence. Thus 9 bases of the 32 have been lost possibly due to cellular exonuclease activity on the 5' of the gene-trap vector before it integrated into the genome.

Group IV clones displayed sequences identical to endogenous *en-2* exon sequences which lie immediately 5' to the *en-2* primer sequence used for RACE cloning. This could arise by priming of the endogenous engrailed transcripts during PCR amplification of the RACE products suggesting RNA was not efficiently removed after first strand synthesis. Alternatively the gene-trap vector may have integrated into the endogenous *en-2* gene by homologous recombination, this is unlikely because the R68 gene-trap integration has been mapped to chromosome 11 by FISH analysis (Section 3.5) whereas *en-2* is located on chromosome 5. Furthermore reporter gene expression

Figure 3.8: Predicted Splice Events from RACE Products

A. Predicted integration into an intron, an endogenous 5' SD splices to vector derived *en-2* SA excluding all intervening sequences.

B. Integration into an exon, reveals two cryptic SD sites within the vector *en-2* intron sequence.

Group II and III refer to Table 3.3. SA, splice acceptor; SD, endogenous splice donor; cSD, cryptic splice donor

Figure 3.8A. Predicted Integration into an Intron

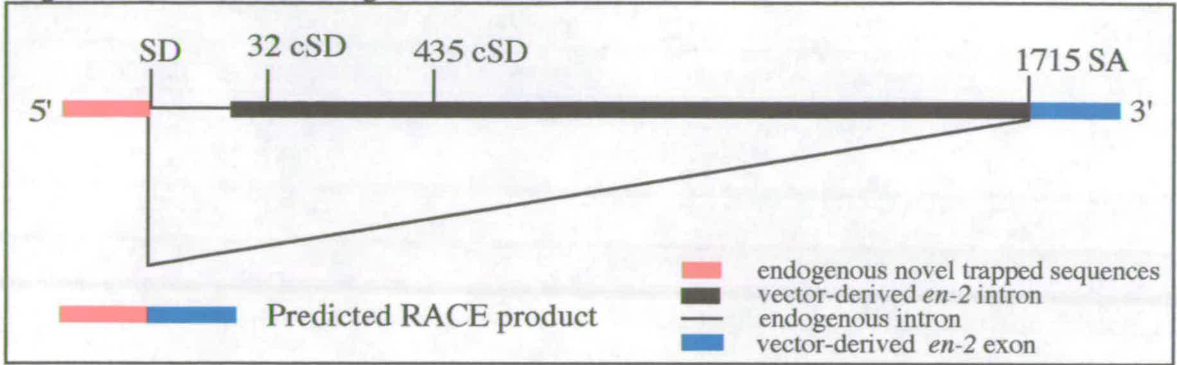
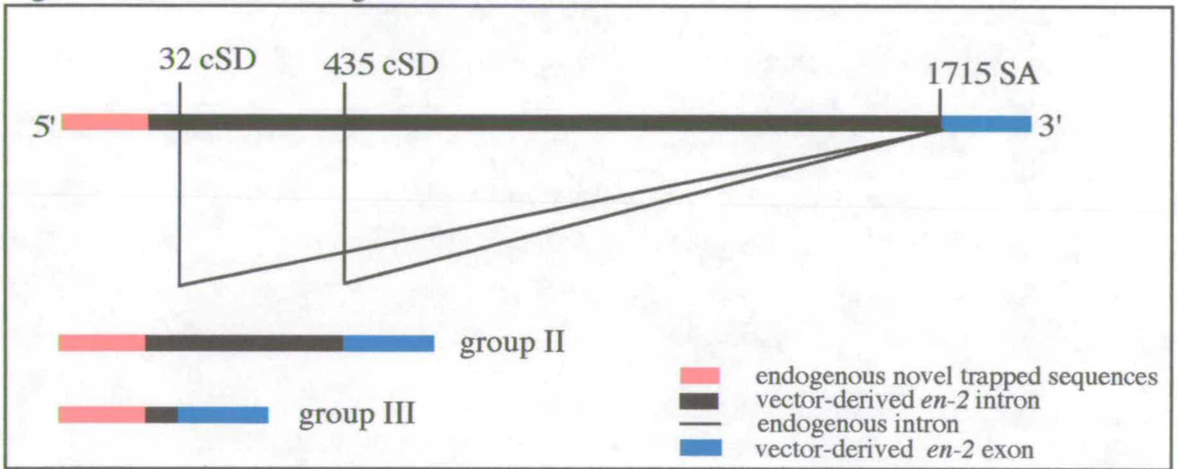


Figure 3.8B. Predicted Integration into an Exon



detected in tetraploid chimeras derived from the R68 ES cells did not mimic the expression of *en-2* (Joyner *et al.*, 1987).

3.8 RNase Protection Analysis of RACE Clones

RNase protection (2.1.15) was used to verify whether group II, group III and group IV clones were derived from the gene-trap integrations. Radio-labeled riboprobes from RACE clones (R68.32, R68.13 and R124.11) were generated to represent each group and hybridised to total RNA isolated from R68, R124 and R1 the parental cell line which contains no gene-trap integration (Figure 3.9). Group II RACE clone R68.32 protected a fusion RNA species of the predicted size (307bases) in the R68 and R124 ES cell RNA but this was not observed in the parental cell line. This indicates that group II clones are derived from the integrations. Endogenous *en-2* sequences (120 bases) are protected in all cell lines. The trapped sequence alone i.e. sequence 5' to vector-derived *en-2* SA (187 bases) was not protected in any cell line. This was consistent with the trapped sequence being derived from intron sequence, which is not normally expressed.

Group III RACE clone R124.11 protected three RNA species (463, 320 and 120 bases) corresponding to the predicted fusion (trapped sequence spliced to the vector-derived *en-2* SA equal to the size of the riboprobe), endogenous trapped sequence and endogenous *en-2* sequence respectively in the R124 cell line RNA. These observations are consistent with the isolation of a RACE clone containing trapped endogenous sequence that is expressed. Two RNA species (320 and 120 bases) were protected in R1 and R68 cell line RNAs corresponding to the endogenous trapped sequence and the endogenous *en-2* sequence. The discrepancy of 23 bases when the sum of these two RNA species is compared to the riboprobe can be explained by the derivation of the 23 bases from vector-derived *en-2* intron sequence that is not normally present in the mRNA population. The lack of a protected RNA species corresponding to the fusion in the R68 cell line RNA shows that the R124.11 RACE clone was specific to the

Figure 3.9: RNase Protection with RACE clones

RNA isolated from ES cell lines (R1 parental cell line; R68 and R124 independent gene-trap cell lines) was hybridised with ^{32}P-labelled riboprobes derived from RACE clones (trapped sequence 5' to vector-derived *en-2* SA fused to vector-derived *en-2* SA) representing groups II, III and IV. M13C, bacteriophage M13mp18 dCTP sequence, and M13A, bacteriophage M13mp18 dCTP sequence, used as size markers. Bands (*) observed in panel III do not correspond to protection of the other splice products identified by RACE.

Predicted size of fragments were protected by the riboprobes:

A, fusion transcript;

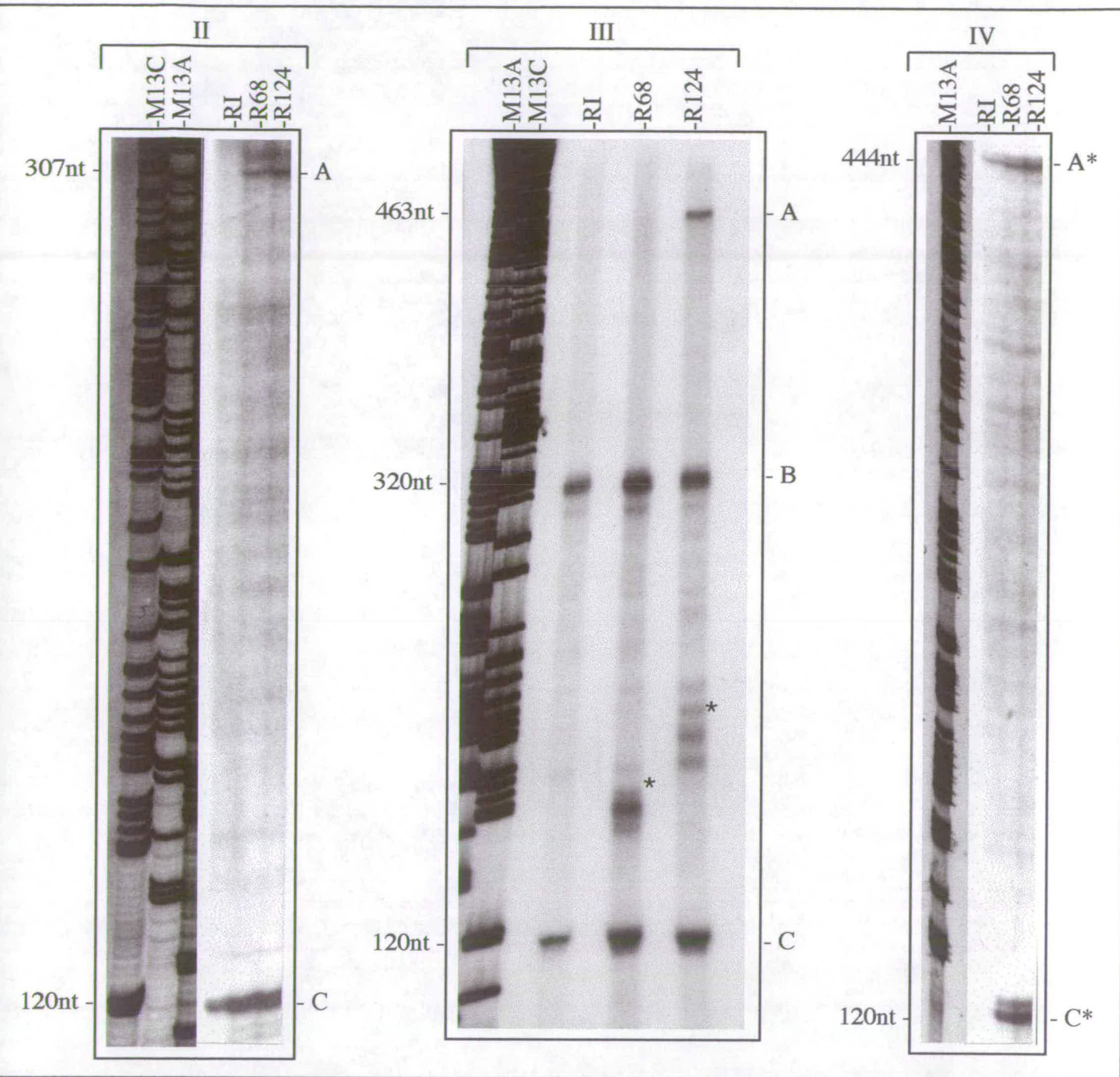
*A, endogenous transcript;

B, trapped sequence 5' to vector-derived *en-2* SA;

C, endogenous *en-2* sequence;

*C, vector-derived *en-2* SA.

Figure 3.9



R124 gene-trap integration. The presence of less prominant protected RNA species in R68 and R124 cell line RNAs, observed as additional bands, may correspond to RNA species with some sequence homology with the RACE clone derived riboprobe that is generated by the integration of the gene-trap vector because these species are not observed in the R1 cell line RNA.

Group IV RACE clone R68.13 protected a single RNA species equal to the size of the RACE clone derived riboprobe (444bases) in R1 ES cell RNA. This indicates that the clones in groupIV are derived from an endogenous *en-2* transcript. In the gene-trap cell lines, R68 and R124, the same RNA species (444 bases) and RNA corresponding to the size of vector-derived *en-2* SA (120 bases) is protected. The protection of the vector-derived *en-2* SA sequence was consistent with other sequences spliced to vector-derived *en-2* SA in the R68 and R124 gene-trap cell lines and with the detection of group II and III clones.

3.9 **Summary**

In vitro cardiac specificity of gene-trap integrations R68 and R124 was confirmed by detection of reporter gene activity detected exclusively in cardiomyocytes generated from both cell lines, respectively. *In vivo* reporter gene cardiac specificity during embryogenesis was shown in heterozygote embryos for the R124 gene-trap integration but this was not possible for the R68 integration, which could not be transmitted through the germline. Reporter gene activity was also detected in the heart, kidney, brain and the gonads of adult animals heterozygous for the R124 integration.

The data gathered from the preliminary characterisation of the R68 and R124 gene-trap integrations indicates that the "intron" gene-trap vector PT1.ATG did not integrate into the ES cell genome as predicted. Molecular analysis of the gene-trap integrations revealed cryptic splicing within the gene-trap vector and was consistent with integration into an exon. The use of the cryptic SD at position 32 in the R124 gene-trap integration allowed identification of the endogenous trapped sequence. In contrast the use of the cryptic SD at position 435 in the R68 gene-trap integration displaced the endogenous trapped sequence further from the vector-derived *en-2* SA and was not identified. However this does not seem to have perturbed the functionality of the vector. Northern blot analysis shows that reporter gene activity was expressed as a fusion mRNA with endogenous sequences and RA-responsive at the transcriptional level.

3.10 **Discussion**

In Vitro Reporter Gene Analysis

Repression of reporter gene activity in gene-trap cell lines R68 and R124 observed in the original screen after 48 hours of RA-exposure was confirmed and shown to be regulated at the transcriptional level. The data also revealed slightly elevated levels of reporter gene activity and transcripts, after 12 and 6 hours of RA-exposure, in the R68 and R124 gene-trap cell lines respectively. This RA-induction was not as dramatic as the RA-repression observed at 48 hours and thus further data will be required to show statistical significance. However, the speculative interpretation of this RA-induction of the reporter gene suggests that the trapped genes maybe directly and positively regulated by RA at the transcriptional level in undifferentiated ES cells, but as differentiation proceeds over the 48 hours the response of the genes reflects the differentiation state of the cell. *In vitro* the RA response time documented for genes regulated directly by RA e.g. *RARβ* and containing a RARE (retinoic acid response element) in their promoters is 6 hours (Manglesdorf *et al.,* 1994). Thus RA-transcriptional induction, if significant, would have important implications on the transcriptional regulation of the endogenous trapped genes.

In vitro cardiac specificity of gene-trap integrations R68 and R124 was shown by correlating reporter gene activity with beating cardiomyocytes derived from both cell lines, respectively. This could be further verified by double labeling reporter gene active cardiomyocytes with known cardiac specific markers e.g. MHCα.

Cardiogenesis has been shown to be inhibited in EBs treated with RA, for the first 2 days of embryoid formation, via the repression of cardiac specific genes e.g. MHCα (Wobus *et al.,* 1994). It is intriguing to compare the RA-repression of cardiac specific genes in EBs to the RA-repression of the 'cardiac specific' R68 and R124 gene-trap integrations after 48 hours of RA-exposure. The question that arises is whether it is valid to compare the differentiation of ES cells in a monolayer to differentiation of ES

cells in EBs. The period of RA-exposure in both systems is comparable, undifferentiated ES cells are exposed to RA for the first 48 hours of differentiation. In the case of the monolayer the ES culture was harvested after 48 hours. In contrast the EBs were allowed to differentiate into morphologically distinct cell types. Thus to validate this comparison one would have to compare the changes in the transcripts of known cardiac specific genes after the 48 hour period. If this was comparable in both systems of differentiation it would suggest the cell fate decision was made within the 48 hour period of RA-exposure to inhibit cardiogenesis via repression of cardiac specific genes or conversely the ES cells were pushed towards an alternative differentiation pathway. This data would support differentiation of a monolayer of ES cells was comparable to differentiation of ES cell derived EBs.

RA-activity affects the *in vitro* cardiogenesis in EBs in a dose and stage/time dependent manner resembling the *in vivo* effects of RA-activity. (Wobus *et al.,* 1994; Wilson *et al.,* 1953). RA is further implicated in cardiogenesis by the cardiac phenotypes observed in knock-out mutants of genes in the retinoid pathway (Dickman *et al.,* 1997; Sucov *et al.,* 1994). The effect of RA on differentiation of EBs and ES cells cannot be strictly considered in isolation with respect to cardiogenesis because the fate of other lineages, via transcriptional regulation of specific genes, may also be determined during the period of RA-exposure (Wobus *et al.*, 1994). The timing and dose of RA-exposure are critical in determining differentiation status. Thus one may extrapolate from this that RA-repression of reporter gene activity in ES cells may not solely select cardiac specific gene-trap integrations.

However, if used in combination with *in vitro* cardiac specific reporter gene activity in gene-trap ES cell derived cardiomyocytes, confirmed with colocalisation of known cardiac markers, this may prove a useful screen to select cardiac specific gene-trap integrations.

Attempts to Transmit the R68 Integration through the Germline

The difficulties in generating chimeras which transmit the R68 gene-trap integration through the germ line may reflect the quality of the R68 ES cells. This highlights a major disadvantage of gene-trapping technology that only one ES cell clone for an integration event is available for studying. The identification of the trapped gene would relieve the difficulties of trying to generate chimeras capable of transmitting the integration through the germ line with R68 ES cells because it would then be possible target the gene by homologous recombination.

The EBs derived from the R68 ES cell line could be differentiated successfully *in vitro* to show cardiac lineage specificity. Thus the R68 gene-trap integration could be used as a cardiac specific marker *in vitro*. The selection of R68 ES cells to homozygosity *in vitro* could be used in combination with the generation of beating cardiomyocytes to investigate the potential phenotype caused by the R68 integration. Studies have shown that cardiac specific markers and electrophysiological data can be gathered from single beating cells in culture (Sanchez *et al.,* 1991; Wobus *et al.,* 1994)).

In Vivo Reporter Gene Activity in R124 Heterozygotes

During development reporter gene expression is restricted exclusively to the heart from 9 d.p.c., when the heart is undergoing looping, until just before birth when it is also detected in the choroid plexus of the brain. In the adult, expression continues to be detected in the heart and is also detected in the kidney, gonads and brain. The early expression of the reporter in the cardiogenic crescent region is restricted to a few cells, this pattern of expression is unlike that documented for the earliest known cardiac markers *Nkx2.5* and *Mef2c*, which are expressed homogenously throughout the cardiogenic region. This early expression implicates the trapped gene in commitment to and maintenance of the cardiac lineage. As the heart develops into two bilateral primitive heart tubes which fuse into the single beating linear heart tube the number of

cells expressing the reporter increases, however the expressing cells are not restricted to any region of the tubes but seem to be scattered throughout the myocardial and endocardial layers. Other cardiac markers at this stage e.g. *eHAND*, *dHAND* are expressed in an asymmetric pattern in the linear heart tube and have been implicated in chamber specification and heart looping (Biben *et al.*, 1997). As looping proceeds chamber specification/regionalisation of the heart is being determined, expression of contractile proteins and cardiac markers can be assigned to the presumptive chambers, however expression of the reporter gene is detected homogenously throughout the heart. By completion of heart looping (9 d.p.c.) and at later stages the expression of the reporter gene is detected exclusively in and throughout the heart, in the epicardium, myocardium and endocardium.

On the basis of reporter gene expression pattern during embryogenesis, the gene trapped by the R124 integration may play a specific role in heart development but does not seem to be involved in regional specification of the heart.

After birth the reporter gene continues to be expressed in the heart but is undetectable at 9 d.p.p., however detection of expression resumes by 11 d.p.p.. The precise window of switch off/down regulation of the reporter is difficult to estimate because β-gal is a relatively stable protein and may therefore remain in the cell for a few days after transcriptional switch off (Cui *et al.*, 1994). In addition expression of the reporter is not observed as a clean switch off/switch on in all cells in the heart but as 'patches' of cells. This may in some way reflect the transcriptional/translational mechanisms involved in regulating expression of the trapped gene taking into consideration that the expression of the reporter during embryogenesis also initially appeared in single cells. Thus the 'patches' of cells may reflect some clonal intra-cellular communication event. Another possibility may be that the reporter gene activity was restricted in some cells leading to mosaic expression (Cui *et al.*, 1994).

The switch on/off of the trapped gene in the heart is very intriguing and one could speculate that it is marking a neonatal to adult stage transition which could involve the expression of an alternative gene isoform eg. switching of *FGFR1* isoforms during heart development (Jin *et al.,* 1994). This is also the period in which reporter gene expression is first detected in the kidney and gonads. A number of examples have been documented which demarcate a "neonate to adult transition" in the mouse between 1 to 3 weeks after birth. In the first week of birth the heart stops undergoing growth via hyperplasia and any growth after this period is achieved through hypertrophy of the polyploid adult cardiomyocytes (Brodsky *et al.,* 1980). The predominating MHC isoform switches from beta, in the embryonic and fetal heart, to alpha in the adult heart (Ng *et al.*, 1991). Ventricular cardiomyocytes begin to exhibit a transient outward K+ current resembling the adult (Wang *et al.,* 1996; Wang *et al.,* 1997) and during this period the expression of enzymes reach levels required for adult cardiac function eg. cyclic AMP-dependent protein kinase (Haddox *et al.,* 1979); lactic dehydrogenase, creatine phosphokinase (Courtney *et al.,* 1978). In the kidney changes in the renin-angiotensin system take place within this period achieving adult levels by five weeks after birth (Gomez *et al.,* 1995). This maturation period of the kidney is associated with a phase of corticosteriod-binding globulin (Scrocchi *et al.,* 1993). Processes are also initiated in the gonads within the first weeks of birth to achieve levels required for sexual maturity at 6 weeks of age. Steroidogenesis in the male and female gonads is characterised by different profiles of steroid hormones (Mannan *et al.,* 1991; Sheffield *et al.,* 1988). Spermatogonia enter spermatogenesis in the second week and produce the first wave of spermatozoa by the fourth week after birth. This process is associated with expression of markers such as *HSP70-2* (Dix *et al.,* 1997), *Mos*, *Abl*, *actin* and *Hox-1.4* (Propst *et al.,* 1988) in the testis.

In the kidney expression of the reporter gene is restricted to the tubules that are involved in resorption of water and salt from urine. This process is strictly regulated

and is critical in the balance of water to salt in blood. Water/salt ratio is a primary factor in blood pressure maintenance and inbalance contributes significantly to hypo- and hypertension.

Expression of the reporter gene in the gonads would suggest that the trapped gene may play a role in germ cell maturation. Reporter gene expression in the male gonads reflects the distribution of Sertoli cells in seminiferous tubules, this will be shown definitively by double labelling cells with a Sertoli cell marker e.g. *WT1* and *lacZ*. These cells are involved in providing maturation signals during spermatogenesis (Jegou, 1993). In the female ovaries, the reporter gene is not expressed, in the follicle cells, the homologous structure of the male Sertoli cells but is restricted to the primitive oocytes and in mesenchymal 'packer' cells, the function of which is ill-defined in literature. The oviduct tract adjacent to the ovaries also expresses the reporter gene in mucus forming cells.

Integration of a gene-trap vector into an active gene in the ES cell genome should place the reporter gene under direct regulation of endogenous cis-acting promoters of the trapped gene. Thus the activity of the reporter should reflect the expression of the endogenous trapped gene and this has been shown for several gene-trap integrations (Gasca *et al.*, 1995; Skarnes *et al.*, 1992). To complement these expression studies it will be necessary to study the diversity of transcripts arising from the trapped gene during embryogenesis and to perform *in situ* hybridisation studies to characterise the expression of the trapped gene using endogenous sequences.

Molecular Analysis of Gene-Trap Integrations R68 and R124

The RACE products derived from R68 and R124 ES cell RNA indicate that integration of the gene-trap vector into an exon leads to a pattern of unpredicted splicing. Integration into an exon has been shown for the R124 integration (Chapter 4, Section 4.2) however for the R68 integration it has been assumed on the basis of the RACE

products. Two cryptic SD sites (32 and 435) have been identified within the vector-derived *en-2* intron sequence. RNase protection has shown that the cryptic splicing between vector engrailed intron (435 site) and engrailed splice acceptor is derived from the R68 gene-trap integration. Thus in the R68 integration the endogenous trapped sequence lies 435 bases 5' to the splice site. It would be possible to isolate the endogenous trapped sequence by the selection of larger RACE clones or by performing RACE using 5' nested primers complementary to the engrailed vector intron sequences. The time constraints of a Ph.D did not permit these experiments to be attempted. In the R124 gene-trap integration cryptic splice events to position 32 and 435 were identified indicating that both can occur when the vector has integrated into an exon. Comparison of each site to the consensus SD has shown the cryptic SD at position 32 has a higher percentage identity with the consensus, potentially making it a stronger candidate for a SD. In contrast the RACE products, from both the R68 and R124 integrations, suggest that site 435 was preferentially used as a SD. Thus it seems that the preference of one cryptic splice site over another cannot be based on sequence homology alone but may be determined by the context of the integration in the genome. However, splicing to position 32 allows identification of the endogenous trapped sequence, as observed for the R124 gene-trap integration.

Thus the "intron" gene-trap vector PT1.ATG can act as a functional "exon" gene-trap vector with an active reporter gene that will allow characterisation of the endogenous trapped gene and allow the isolation of the trapped sequences. In addition integration into an exon may have increased the the possibility of disrupting the endogenous trapped gene.

Implications for Future Gene-Trap Screens

RA-repression as a screen to select for cardiac specific gene-trap integrations may not be ideal, as discussed above, but combined with *in vitro* reporter gene cardiac

specificity in beating cardiomyocytes may prove more fruitful. Gene-trap integrations R68 and R124 fulfill the predictions of this screen, reporter gene activity was RA-repressed and detected exclusively in *in vitro* beating cardiomyocytes generated from both gene-trap cell lines. *In vivo* reporter gene activity was detected exclusively in the developing heart of embryos heterozygous for the R124 gene-trap integration. This was consistent with *in vitro* differentiation of EBs recapitulating early cardiogenesis (Sanchez *et al.,* 1991). In the adult the reporter gene was expressed more widely but was cardiac muscle specific.

PT1.ATG gene-trap vector, although it was designed as a intron-trap vector, has been shown to be a functional exon-trap vector through the activation of cryptic SDs within the gene-trap vector. The cryptic splicing leads to vector-derived *en-2* intron sequence incorporation into the fusion transcripts and complicates RACE analysis by the presence of multiple products. Thus, screens have been used to remove such events by the detection of intron containing RACE products e.g. Townely *et al.,* 1996. An alternative approach would be to modify the PT1.ATG gene-trap vector to remove the cryptic SD sites, but this rather than alleviate the problem may result in the activation of other uncharacterised cryptic SD sites when the vector integrates into an exon. An example of this is documented for human thalassemia syndromes where loss of a splice site results in more, rather than less, alternative splice events (Orkin *et al.*, 1984). Thus to avoid this, it may be optimal to modify the cryptic SD at position 32 to enhance splicing activity upon integration of the vector into an exon and to remove the cryptic SD at position 435 thus eliminating RACE products which contain 435 bases of vector-derived *en 2* intron and make the identification of endogenous trapped sequences difficult. This modification would result in optimization of the vector to be functional as a 'exon' and 'intron' gene-trap vector.

3.11 **Conclusion**

The R124 gene-trap cell line fulfilled both criteria of the analysis, the integration was transmitted through the germ line and endogenous trapped sequences were isolated.

In contrast, characterisation of the R68 gene-trap cell line was unsuccessful, the integration could not be transmitted through the germline and no endogenous trapped sequence was isolated. Other strategies could allow the characterisation of the R68 gene-trap integration but time constraints of a Ph.D did not permit the possible experiments discussed above to be attempted and so no further studies were continued with the R68 gene-trap cell line.

RACE cloning data from R68 and R124 integrations revealed the gene-trap vector had not integrated in the ES cell genome as predicted, into an intron, but into an exon. PT1.ATG gene-trap vector was shown to function as an exon-trap vector, via the activation of cryptic SDs within the vector, by expression of the reporter gene as a fusion transcript with endogenous sequences and isolation of endogenous trapped sequence from the R124 gene-trap integration. Integration of the vector into an exon may have increased its potential to generate a mutation. *In vitro* reporter gene analysis confirmed that gene-trap integrations R68 and R124 were RA-responsive and specific to cardiomyocytes generated from the *in vitro* differentiation of both cell lines, respectively. *In vivo* reporter gene activity characterised in embryos heterozygous for the R124 integration showed cardiac specificity during embryogenesis. In adult heterozygotes reporter gene activity was detected in the kidney, gonads, brain and specifically in cardiac muscle. This indicates that the gene trapped by the R124 integration may be involved in the development of the heart and function of the adult tissues. Further characterisation of the R124 gene-trap integration and identification of endogenous trapped sequences are presented in Chapter 4 and the phenotypic consequences of disrupting the endogenous trapped gene are presented in Chapter 5.

Chapter 4

4.0 **Molecular Analysis of the R124 Gene-Trap Integration**

Introduction

RACE clone analysis of the R124 gene-trap integration suggests that the gene-trap vector has integrated into an endogenous exon and that endogenous trapped sequences are expressed as a fusion transcript. The reporter gene analysis of the R124 gene-trap integration suggests that the trapped gene may play a role in heart development and in the function of the heart, kidney, gonads and brain. This chapter describes the PCR and RFLP analysis to confirm that the vector has integrated into an exon and that the trapped sequence is linked to the vector at the genomic level. The isolation of more sequence to identify the trapped gene and the detection of endogenous transcripts is also presented.

4.1 **Sequence Analysis of R124 ES Cell RNA derived RACE Clone 11**

Sequencing revealed that clone 11, isolated from RACE-PCR of R124 ES cell RNA and confirmed by RNase Protection Assay (Chapter 3 Section 3.8), contained 320 bases of novel endogenous sequence (referred to as KXE), 23 bases of vector derived *en-2* intron and 120 bases of vector derived *en-2* exon sequence (Figure 4.1). Translation of the novel trapped sequence revealed a short open reading frame (ORF), in frame 2 with a methionine at position 266, that could potentially be the start of the coding sequence this start does conform to a Kozak consensus (Figure 4.1). The methionine at position 266 was out of frame with the *lacZ* reporter gene ATG start (frame 1). This trapped sequence showed no significant homology to known genes or ESTs in database searches. The lack of a ORF and homology suggests this sequence may be non-coding 5' UTR (untranslated region) sequence. Thus it will be essential to isolate more cDNA sequence 3' to the integration to identify the gene function. BLAST database search programs, provided by National Centre for Biotechnology Information

Figure 4.1: R124 RACE clone 11

Full RACE clone sequence and endogenous trapped sequence translated in all three forward frames.

Box, KXE sequence;

black, endogenous trapped sequence;

red, vector derived *en-2* intron sequence;

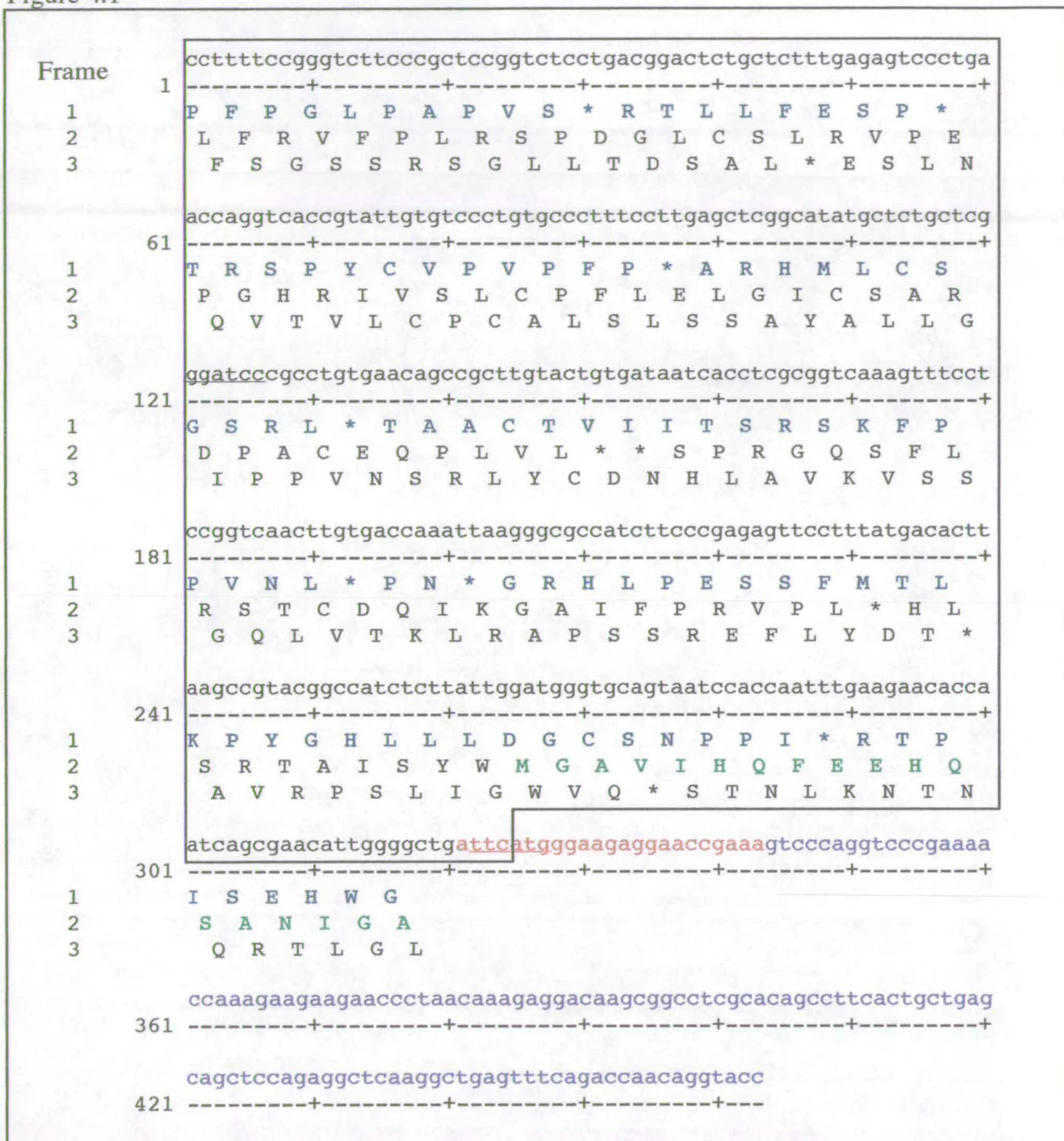
blue, vector derived *en-2* exon sequence;

underlined, restriction sites (BamHI, GGATCC; EarI site TTCATG);

dark blue, frame 1 is in frame with the downstream *lacZ* reporter gene;

green, frame 2 has a potential open reading frame, with a methionine at position 266.

Figure 4.1



(NCBI) on the Netscape (http://www.ncbi.nlm.nih.gov/cgi-bin/BLAST), were used to perform searches on databases which included Genbank, EMBL, DDBJ, DDBJEST, PDB, SwissProt and PIR (Altschul et al, 1990).

4.2 Vector integration into an endogenous gene exon

The use of the cryptic SD (Chapter 3, Section 3.7) suggested that the vector had integrated into an exon. This was confirmed at the genomic level using PCR analysis (by Peter McClive) (Chapter 2, Section 2.1.9). Primer pair b1 and b2, complementary to the endogenous trapped sequence and to the vector sequence respectively amplified a PCR product of predicted size (282 bases) from heterozygote and homozygote genomic DNA confirming that there were no intervening intronic sequences between these primer sites and therefore that the vector had integrated into an exon (Figure 4.2). Control primers, a1 and a2 complementary to the *en-2* splice acceptor, intron and exon sequence respectively, amplified a PCR product of predicted size (672 bases) in all genotypes. The band generated from the wild type genomic DNA sample corresponds to the endogenous *en-2* genomic sequences and the more intense heterozygous and homozygous bands include the additional amplification from 1 and 2 copies of vector derived *en-2* sequences, respectively.

4.3 Isolation of cDNA

In order to obtain more sequence information the novel trapped sequence, KXE (Figure 4.1), was used to screen a heart cDNA library (Clontech, adult Mouse Heart 5'- Stretch plus cDNA Library) (Chapter 2, Section 2.1.16). Three cDNA clones containing the trapped sequence were isolated. Restriction enzyme mapping showed that two clones (TRI, 1.6kb) were identical and these overlapped with a third clone (VRI, 0.8kb) (Figure 4.3). Clone VRI(0.8kb) has 131 bases overlap with the 5' end of the trapped sequence isolated by RACE. Restriction enzyme analysis and sequence

Figure 4.2: PCR analysis of the R124 genomic integration site.

Detection of PCR products of a predicted size verifies the integration of the gene-trap vector is into an exon.

A. Control primer pair (a1 x a2) amplifies a 672bp fragment in wild type, heterozygote and homozygote genomic DNA. Test primers (b1 x b2) amplify a 282bp fragment only in heterozygote and homozygote genomic DNA.

B, primer sites determined from vector (*en-2* SA intron, exon) and RACE cloned KXE sequence.

Figure 4.2A.

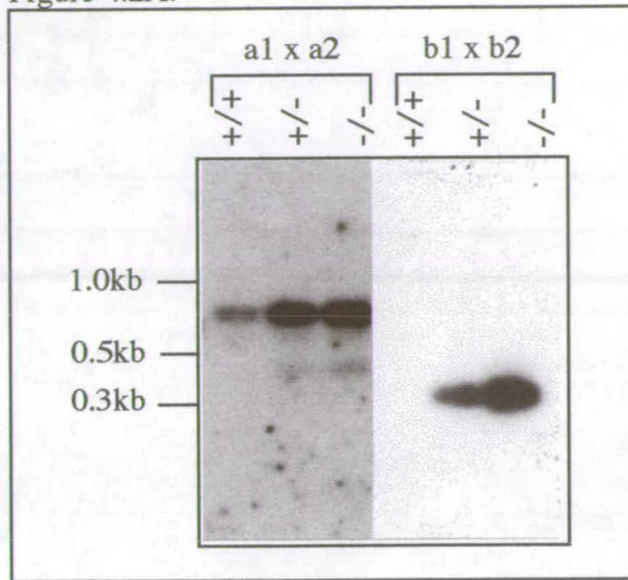


Figure 4.2B

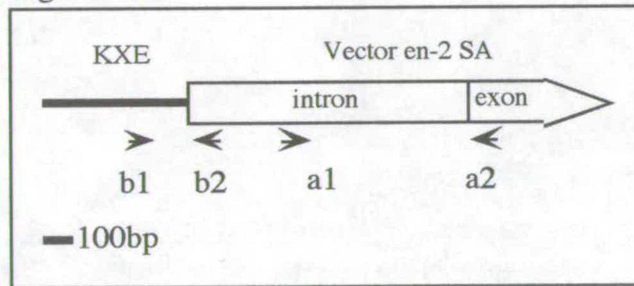


Figure 4.3: cDNA clones isolated from a Mouse Heart cDNA Library Screen

cDNA clones relative to endogenous RACE cloned sequence KXE (320 bases). Restriction sites used to align sequences shown.

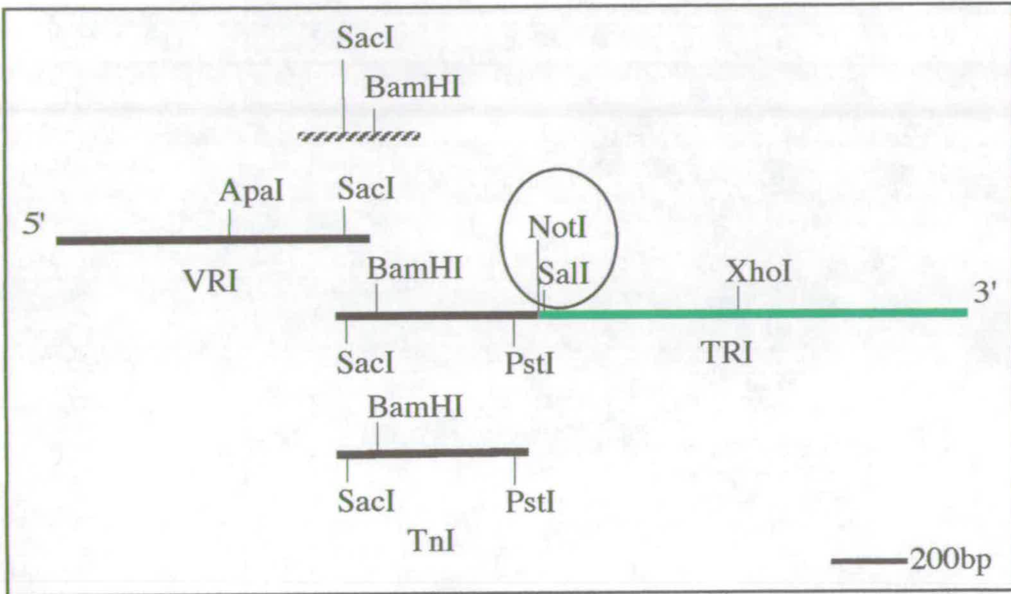
VRI (0.8kb) cDNA clone.

TRI (1.6kb) hybrid cDNA clone.

TnI (0.6kb), NotI subclone of TRI.

Hatched bar, KXE; circled, restriction sites in cDNA library adaptor; green bar, homology to mitochondrial sequences.

Figure 4.3



analysis revealed that the TRI clones contained restriction sites present in the adaptor sequences used when preparing the cDNA library indicating that these were hybrid clones. The fragment 5' of the adaptor sequence (TnI, 0.6kb) contained 176 bases that were identical to the 3' region of the RACE cloned KXE sequence. Database searches using sequences 3' of the adapter sequence clones showed high homology to mitochondrial DNA in database searches and were irrelevent to the characterisation of the gene-trap integration. Thus, in total a contig of 1004 bases of cDNA 5' and 306 bases 3' to the site of integration has been isolated (Figure 4.4).

4.4 Sequence Analysis of cDNA Clones

No homology with known genes in database searches was revealed, however homology with a number of EST was revealed indicating the VRI cDNA clone sequence is transcribed (Table 4.1). The sequence analysis of the TnI fragment showed the presence of an inverted repeat (25 bases) separated by 255 bases that potentially generates secondary structure and disrupts sequencing (Figure 4.4). Subcloning the TnI clone to remove one repeat allowed a total of 559 bases of sequence to be determined. Database searches have revealed the inverted 25 base sequence and the most 3' sequence (TnT7, 306 bases) of the TnI cDNA clone to be highly conserved among myosin heavy genes (Table 4.1). Approximately 41 bases of the TnI clone are unknown, this sequence will be required to show the overlap between the 5' (TnSp6) and 3' (TnT7) ends of TnI and confirm the presence of the RACE clone sequence KXE. The complete sequence of the clone will provide definitive proof that the TnI clone arises from a single transcript and was not the result of some homologous recombination event between the highly conserved inverted repeats.

Translation of the cDNA clone VRI sequence in all frames revealed no continuous ORF suggesting that it may be 5'UTR. Translation of the TnI sequence revealed ORF2

Figure 4.4: Contig of RACE and cDNA Sequences

Boxed, RACE cloned sequence;

black, VRI cDNA sequences;

pink, overlap between RACE and VRI sequences;

red, overlap between RACE, VRI and TnI sequence;

blue, overlap between RACE and TnI sequence;

orange, TnI sequences;

..\\.., 41bp of unknown sequence;

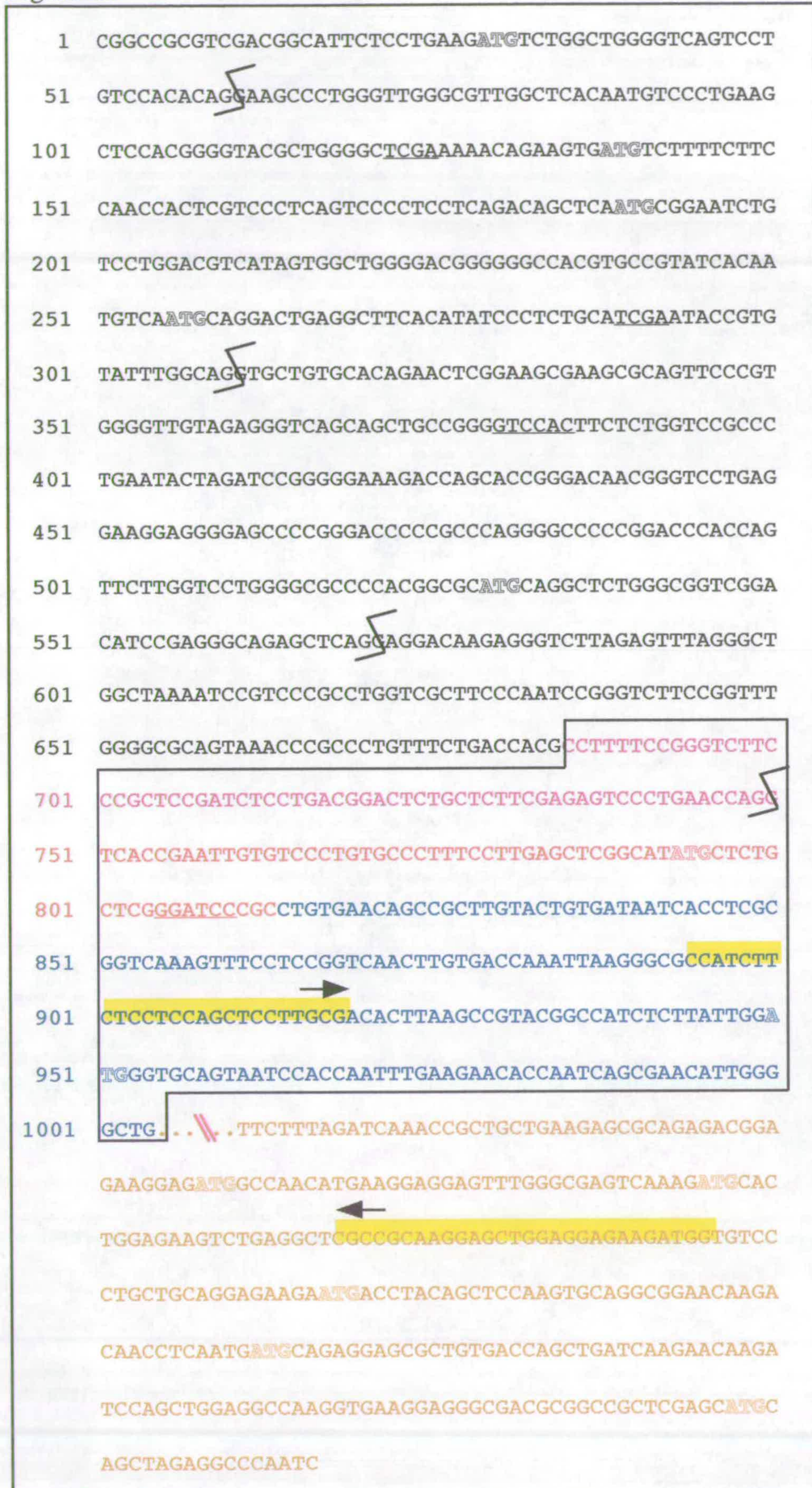
yellow, inverted repeats, arrow shows direction;

outlined, potential ATG starts;

underlined, restriction sites (BamHI, GGATCC; PvuII, GTCCAC; TaqI, TCGA);

↳, potential splice junctions.

Figure 4.4



matched the MHCα sequence at the protein level, this was anticipated because of the high homology at the nucleotide level (Table 4.1).

4.5 **Detection of endogenous transcripts by northern blot analysis**

Northern blot analysis was used to detect the distribution of endogenous transcripts in wild type tissues using a probe generated from the cDNA clone VRI isolated from the heart library. This probe detects a heart specific transcript (4.4kb) and a ubiquitously expressed transcript (1.4kb) in R1 ES cells and in wild type tissues (heart, kidney, brain, testis, skeletal muscle also detected in, but not shown, lung, spleen, thymus and liver) (Figure 4.5A). Thus the distribution of endogenous transcripts seems to be in a wider range of tissues than detected with reporter gene activity.

The smaller RACE cloned KXE probe detects a more complex pattern of transcripts that overlap with the transcripts detected with the VRI probe. This could be explained because the RACE clone contains sequences which extend more 3' than the cDNA and includes the conserved 25 base repeat. Thus one could infer that the shorter probe may contain sequences in common with more transcripts. KXE used as a probe detects transcripts of three sizes (4.4, 2.0 and 1.4kb) which are differentially expressed in wild type tissues (Figure 4.5B). 4.4kb and 2.0kb transcripts are expressed in all tissues expressing the reporter gene but are also detected in tissues in which reporter gene activity cannot be detected by histological staining i.e. skeletal muscle, lung and thymus. These transcripts run close to 18S and 28S ribosomal RNA (rRNA). Samples of rRNA ran with test samples did not hybridise with the probe (data not shown) excluding the possibility that the probe was detecting rRNA. A 1.4 kb transcript is detected in heart, kidney, testis and brain.

Table 4.1: Database Search Results

Clone	Database	Access. no.	Homology	%match (no.bases)
VRI	DBEST	gb|AA472888	Soares mouse lymph node	98 (263)
	DBEST	gb|W93516	Soares fetal heart NbHH19W Homo sapiens	82 (301)
	DBEST	gb|AA437126	Soares tDBESTis NHT Homo sapiens	82 (302)
TnSP6	nr	gb|M76598|M USACMHCA	Mouse alpha cardiac myosin heavy chain mRNA	90 (32)*
	nr	emb|X15938|R NMHCB	Rat alpha cardiac myosin heavy chain mRNA	90 (32)*
		emb|X15939|R NMHCB	Rat beta cardiac myosin heavy chain mRNA	90 (32)*
	nr	dbj|D00943|H UMCAMHC	Human cardiac alpha-myosin heavy chain mRNA	90 (32)*
	DBEST	gb|AA153146	Stratagene mouse heart (#937316) cDNA clone 604443 5' similar to gb:Z20656_rna1 myosin heavy chain, cardiac muscle alpha isoform(human); gb:M76601 Mouse alpha cardiac myosin heavy chain mRNA,	90 (32)*
TnT7	nr	gb|M76599|M USACMHCB	Mouse alpha cardiac myosin heavy chain mRNA	100 (260)‡
		emb|X15938|R NMHCA	Rat alpha cardiac myosin heavy chain mRNA	95 (260)‡
		emb|X15939|R NMHCB	Rat for beta cardiac myosin heavy chain mRNA	95 260)‡
		dbj|D00943|H UMCAMHC	Human cardiac alpha-myosin heavy chain mRNA	93 (260)‡
	DBEST	gb|AA153146	Stratagene mouse heart (#937316) cDNA clone 604443 5' similar to gb:Z20656_rna1 myosin heavy chain, cardiac muscle alpha isoform(human); gb:M76601 Mouse alpha cardiac myosin heavy chain mRNA,	99 (260)‡
	DBEST	emb|F00516|H SB07B072	H. sapiens partial cDNA sequence	90 (232)
25mer#	nr	gb|M76598|M USACMHCA	Mouse alpha cardiac myosin heavy chain mRNA	100 (25)
		emb|X15938|R NMHCA	Rat alpha cardiac myosin heavy chain mRNA	100 (25)
		emb|X15939|R NMHCB	Rat beta cardiac myosin heavy chain mRNA	100 (25)
		emb|Z20656|H SCAMHCA	Homo sapiens cardiac alpha-myosin heavy chain	100 (25)
	DBEST	gb|AA153146	Stratagene mouse heart (#937316) cDNA clone 604443 5' similar to gb:Z20656_rna1 myosin heavy chain, cardiac muscle alpha isoform(human); gb:M76601 Mouse alpha cardiac myosin heavy chain mRNA	100 (25)
		emb|F00516|H SB07B072	H. sapiens partial cDNA sequence	100 (25)
		gb|AA106052	Stratagene mouse kidney(#937315) cDNA	91 (24)
		gb|AA207961	GuayWoodford Beier mouse kidney day 7	91 (24)
		gb|AA619356	Barstead mouse myotubes MPLRB5 cDNA	91 (24)

List restricted to most highest homologies
TnSP6 and TnT7, opposite ends of TnI clone
* homology over same region of clone
‡ homology over same region of clone, includes database homology to *
# sequence present in TrSP6 and TRT7

Figure 4.5: Expression of Endogenous Transcripts

Wild type adult tissue RNA northern blots hybridised with ^{32}P-labelled probes derived from:

A. cDNA VRI minus sequences overlapping with KXE (600bases), R1 ES cell sample on blot.

B. RACE cloned KXE (320 bases), R1 and R124 ES cell samples on blot.

β-Actin loading controls presented below each panel.β-Actin cDNA (610bases) (Tokunaga *et al.*, 1986).

[note: loading of heart sample less than other tissue samples and exposure time not long enough to detect smaller and less abundant 1.6kb transcript in heart tissue samples in Figure 4.5B].

Figure 4.5A

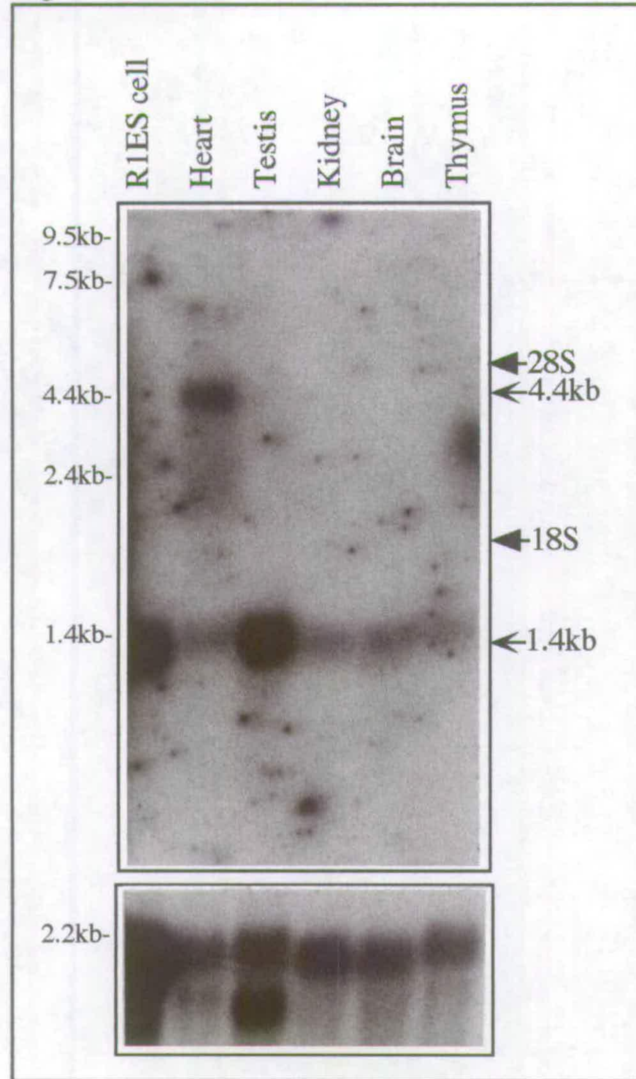
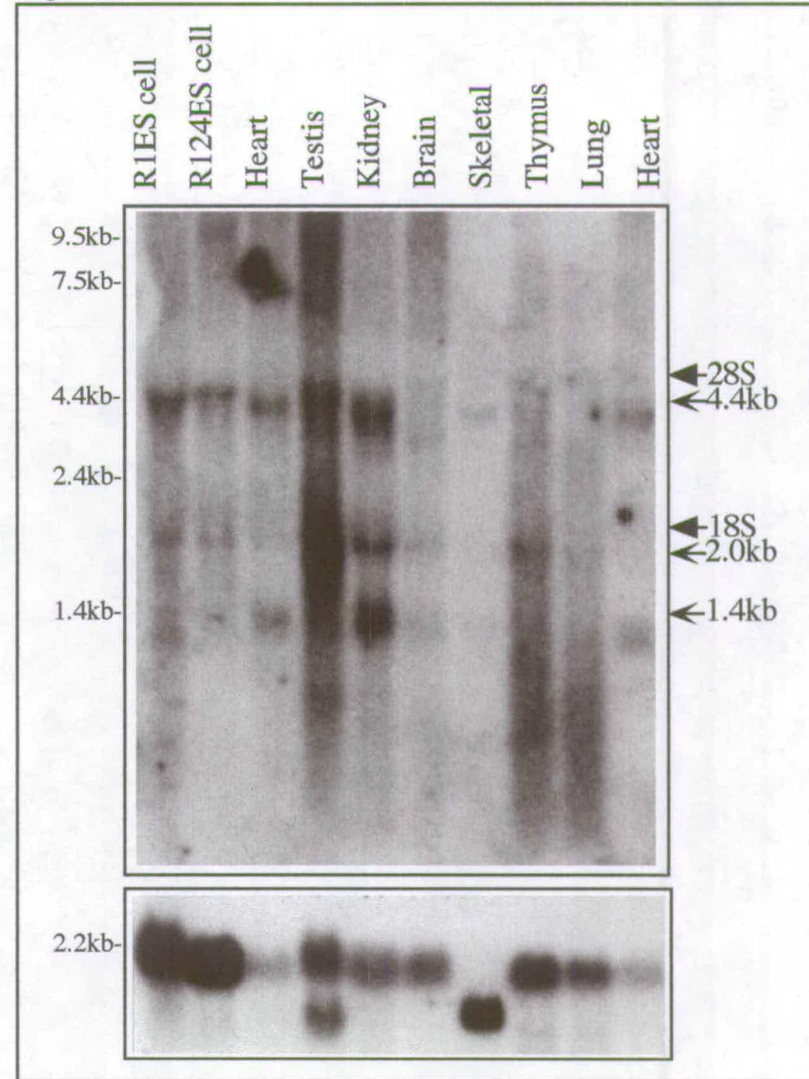


Figure 4.5B



4.6 **Genomic Structure of the R124 Gene-Trap Integration**

Identification of animals heterozygous or homozygous for the R124 integration is important for expression and phenotype analyses. This was initially done by the detection and quantification of the number of gene-trap vector copies integrated into the genome using a *lacZ* probe. However relying on a signal intensity was not ideal so attempts were made to analyse the genomic structure of the integration and identify a RFLP (restriction fragment length polymorphism) which allowed unequivocal distinction between wild type and trapped alleles of the endogenous trapped gene.

4.61 **Detecting and Quantifiying the Number of Integrated Vector Copies**

Backcross and intercross litters (Chapter 2, Section 2.3.2) were initially genotyped by Southern blot analysis of genomic DNA digested with EcoRI and hybridised to detect the presence of *lacZ* (Figure 4.6C). Figure 4.6A shows the banding pattern obtained from a litter derived from mating heterozygous males and females (intercross). No bands were detected in the wild type genomic DNA and two bands (6kb and 10kb) of equal intensity are detected in the heterozygote and homozygote genomic DNA samples. Bands (6kb and 10kb) in homozygotes were twice as intense as the heterozygote bands. The banding pattern indicates that two copies of *lacZ* are present, this results from two copies of the vector integrating into the genome in tandem (Figure 4.6C).

The tandem integration of the vector was confirmed using *en-2* exon sequences as a probe, which detects both the endogenous (12kb) and vector-derived (6kb and 10kb) *en-2* sequences (Figure 4.6B). The vector-derived bands are the same as those detected with the *lacZ* probe and were anticipated from the restriction map of the vector. The intensity of bands on a Southern blot hybridised with the *en-2* probe were quantified by phosphoimaging analysis (ImageQuant, Molecular Devices) to estimate the number of vector associated *en-2* copies relative to the two copies of endogenous

Figure 4.6: Southern blot analysis of genomic DNA

Genomic DNA digested with EcoRI and hybridised sequentially with:

A. ^{32}P -labelled *lacZ* probe and;

B. ^{32}P -labelled vector-derived *en-2* exon probe.

C. Predicted fragments generated by EcoRI digest of genomic DNA and detected by vector derived probes. +ve, *lacZ* positive; -ve *lacZ* negative; +/+, wild type genotype; +/-, heterozygous genotype; -/- homozygous genotype.

Figure 4.6A

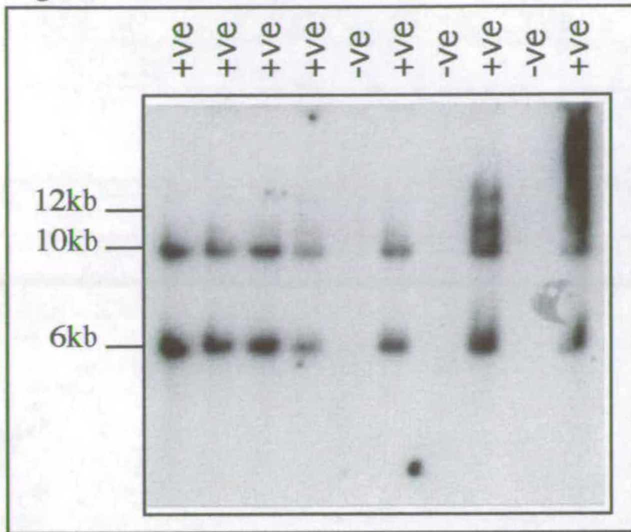


Figure 4.6B

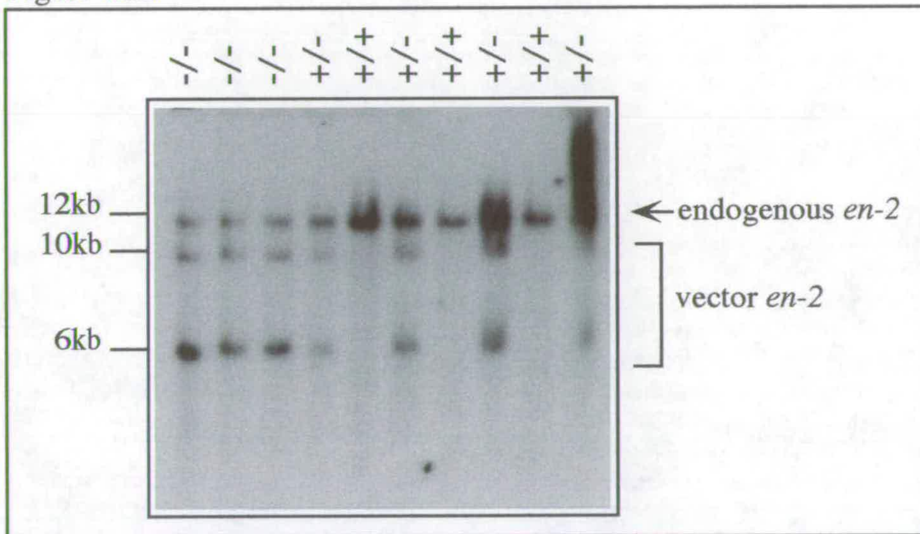
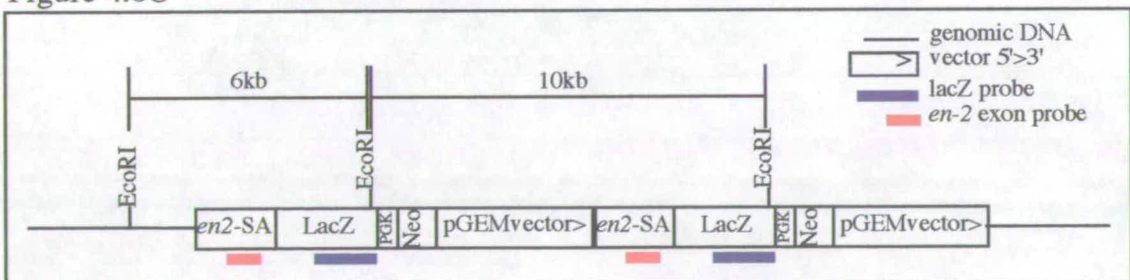


Figure 4.6C



en-2 in a genomic DNA sample. In samples from animals heterozygous for the R124 gene-trap integration the ratio between endogenous *en-2* and vector *en-2* was 1: 1, confirming two copies of the vector had integrated. A proportion of samples generated from intercross litters showed a ratio of 1:2 between endogenous *en-2* and vector-derived *en-2* suggesting that 4 copies of vector-derived *en-2* were present and that these samples were derived from animals homozygous for the gene-trap integration. This always correlated with samples that had been identified as homozygotes based on the *lacZ* intensity.

4.62 **Restriction Fragment Length Polymorphism (RFLP) Analysis**

The size of DNA fragments generated by restriction endonuclease activity from a single genetic locus can vary from one allele to another because of the variation in nucleotide sequences. The variation observed is termed restriction fragment length polymorphism (RFLP) and can be used as a molecular marker to identify one allele from another. A RFLP has been determined which allows distinction between wild type and trapped alleles. Southern blot analysis of genomic DNA digested with the restriction enzyme BamHI and hybridised with the RACE clone KX, that contains novel sequence and a region of the vector-derived en-2 exon sequence, generates a characteristic banding pattern (Figure 4.7A).

The RFLP reliably confirmed the genotype of 52 animals which had been genotyped as wild type, heterozygous or homozygous with *lacZ* and *en-2* assays described above. Interestingly, litters backcrossed onto 129/CGR and C57BL/6 backgrounds showed strain-related RFLPs (Figure 4.7A). Southern blot analysis with genomic DNA isolated from the pure 129/CGR and C57BL/6 strains has confirmed that a polymorphism at the endogenous locus exists between the two strains.

The RFLP has shown the direct linkage of the trapped sequence to the integrated gene-trap vector at the genomic level by the detection of common bands by sequential

Figure 4.7: RFLP analysis of the R124 integration

A. Genomic DNA digested with BamHI and hybridised to 32P-labelled RACE clone KX (contains endogenous trapped sequence and vector derived *en-2* SA sequence) probe results in a banding pattern which allows distinction between wildtype and trapped alleles.

B. Southern blot analysis of genomic DNA digested with BamHI and sequentially hybridised with a probe derived from 1. the RACE cloned endogenous trapped sequence KXE (contains no vector-derived *en-2* sequence) and 2. vector-derived *en-2* intron sequence shows linkage of KXE to the gene-trap vector at the genomic level.

+/+, wild type genotype; +/-, heterozygous genotype; -/- homozygous genotype; A, trapped allele; B and C wild type alleles [polymorphic between 129CGR (B) and C57Bl/6 (C)]; E, a 300bp band common to wild type and trapped alleles; D, vector derived *en-2* exon sequence; Ia, endogenous *en-2* intron sequence; Ib, vector-derived *en-2* intron from second vector.

Figure 4.7A

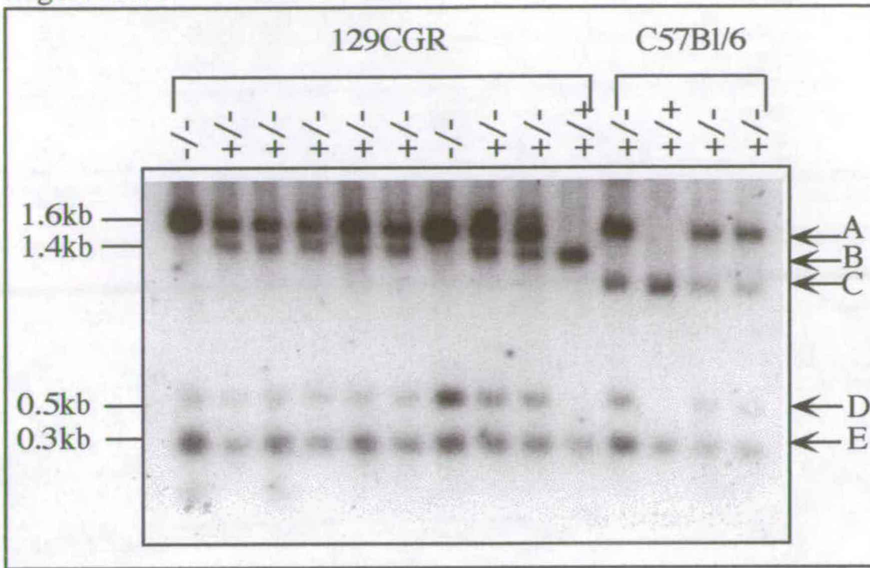


Figure 4.7B1

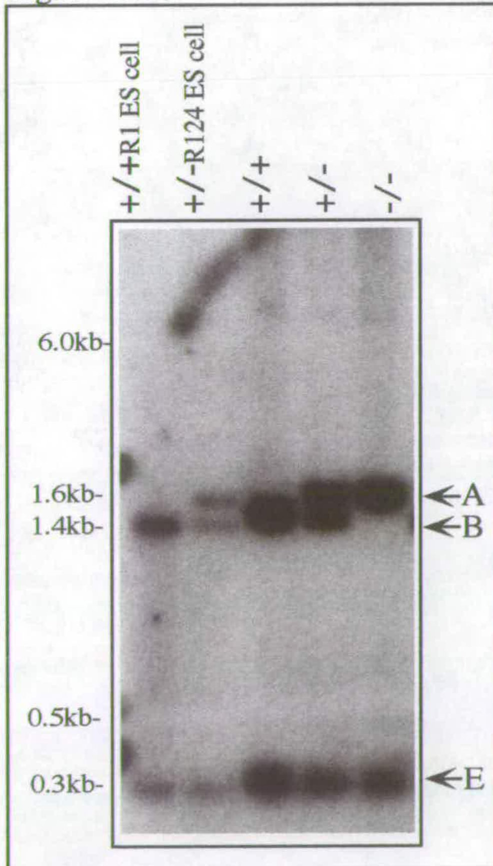
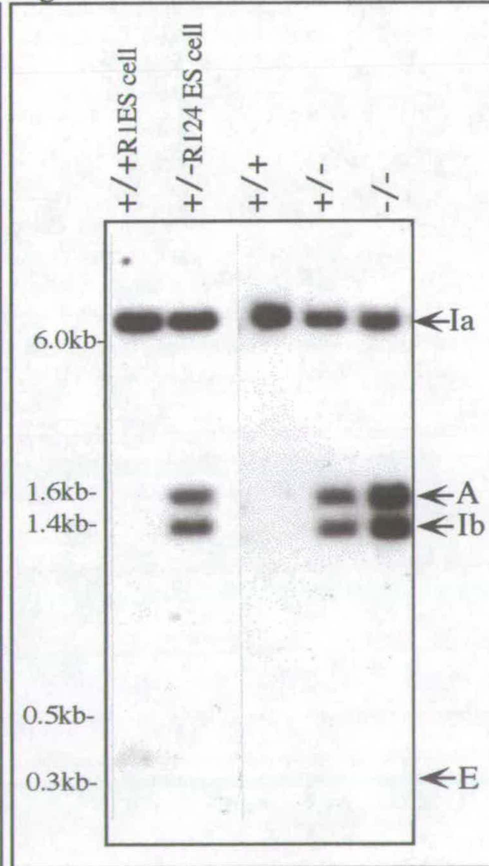


Figure 4.7B2



hybridisation of Southern blots with trapped sequence specific and vector-derived engrailed specific probes (Figure 4.7 B1 and B2). Band A (1.6kb) arises from the trapped allele and can be detected by probes generated from KXE and vector-derived *en-2* intron. Bands B (1.4kb) and C (1.2kb) represent the endogenous wildtype allele, which is naturally polymorphic between 129/CGR and C57BL/6 mouse strains and can be detected by KXE. Band D (0.5kb) is specific to the vector and is not present in wild type samples. Band E (0.3kb) is endogenous sequence that is constant between wild type and mutant alleles, and can be detected by KXE and is consistent with the presence of a BamHI site in the probe. Band Ia represents endogenous *en-2* sequences and can be detected with a vector-derived *en-2* intron. Band Ib represents vector-derived *en-2* intron sequence from the second copy of the vector integrated into the genome (data summarised in Figure 4.9A). Further RFLP analysis of genomic DNA (wild type, heterozygous and homozygous samples) digested with a range of restriction enzymes, BamHI, BglII, KpnI, PvuII, and TaqI (not shown) and sequentially hybridised with VRI cDNA, RACE KXE and vector specific probes has shown bands in common indicating that the cDNA, RACE trapped and gene-trap vector sequences are linked at the genomic level (Figure 4.8).

The RFLP analysis of genomic DNA and cDNA clones with the different probes, the sequence analysis of the RACE clone and VRI cDNA, and the phosphoimaging analysis has allowed the prediction of the R124 gene-trap integration in the genome, 2 copies of the gene-trap vector integrated into a single site in a head-to-tail tandem repeat on chromosome 5, (Figure 4.9A) and a local exon/intron structure of the endogenous trapped gene (Figure 4.9B). The isolation of a genomic clone for the endogenous gene will confirm the exon/intron structure by comparison to cDNA sequences.

Figure 4.8: Linkage of Endogenous Sequences to the Vector at the genomic level

Southern blot analysis of genomic DNA digested with BglII, KpnI and PvuII, and sequentially hybridised to:

A. KXE (endogenous RACE cloned sequence);

B. VRI (cDNA);

C. vector derived *en-2* intron probes.

The trapped allele (arrow head) was detected in common with all probes.

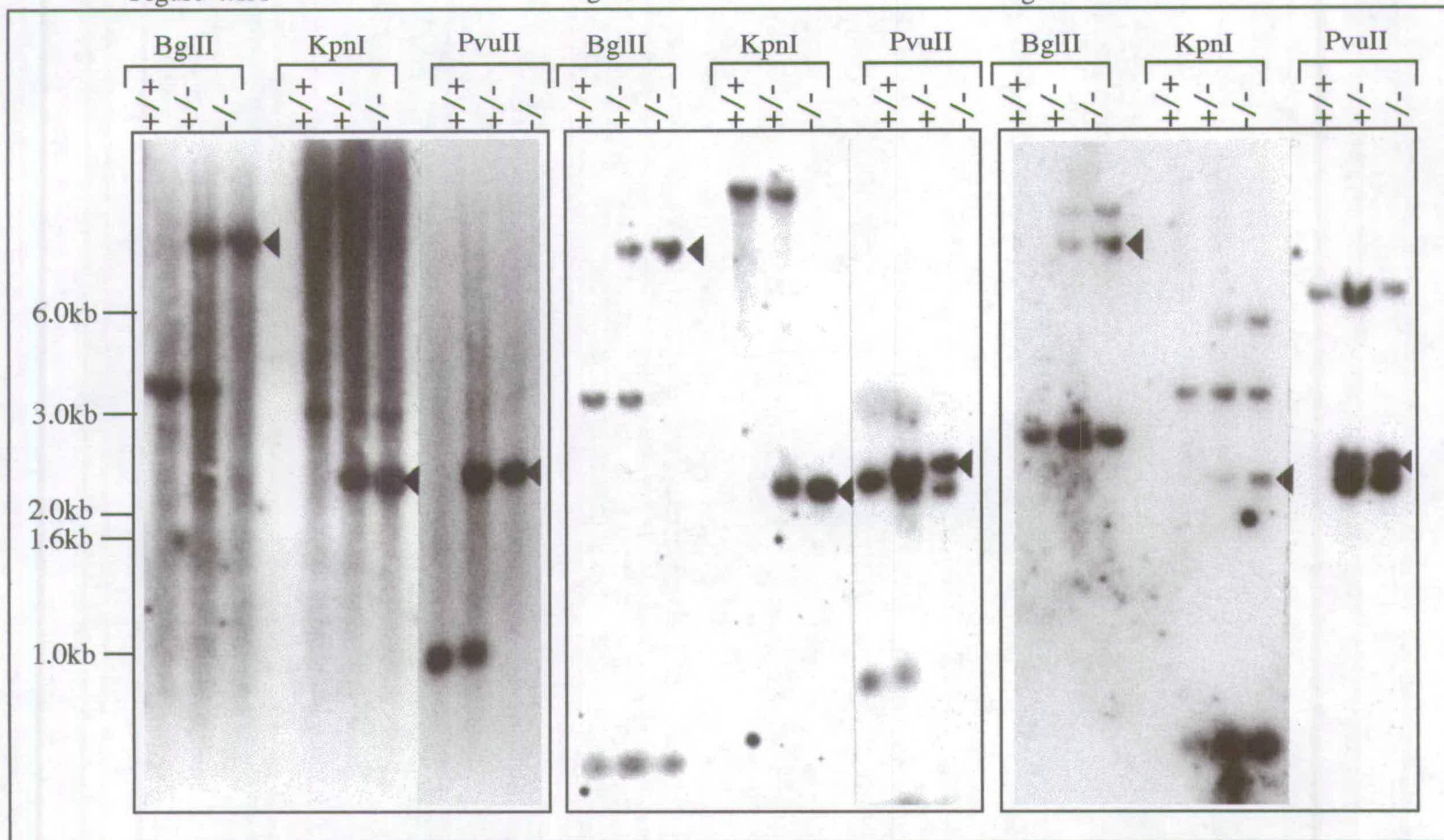


Figure 4.8A
Figure 4.8B
Figure 4.8C
BglII
KpnI
PvuII
BglII
KpnI
PvuII
BglII
KpnI
PvuII
6.0kb
3.0kb
2.0kb
1.6kb
1.0kb

Figure 4.9: Predicted Structure of the R124 gene-trap integration

A. Tandem integration of gene-trap vectors into an endogenous exon, detection of bands A - E, Ia and Ib, restriction fragments generated from genomic DNA digested with BamHI shown in Figure 4.7.

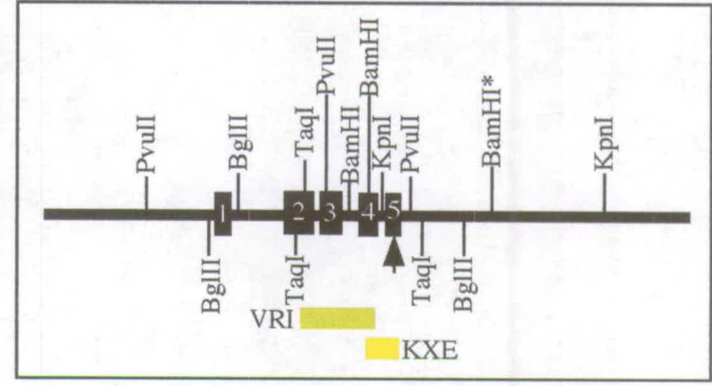
B. Preliminary local endogenous exon/intron structure and restriction sites surrounding integration site (not to scale). Preliminary data suggests VRI generated from exons 1, 2 and 3, and KXE generated from exons 3 and 4. Arrow, site of integration into exon 5.

Figure 4.9A

endogenous exon - size unknown
unknown genomic DNA structure
KXE probe
vector-derived *en-2* exon probe
vector-derived *en-2* intron probe
vector-derived *lacZ* probe
vector 5'->3'
BamHI* polymorphic site

Wild type allele
B - 1.4kb 129CGR
[B=C - 1.2kb in C57Bl/6]
E - 0.3kb
Trapped allele
A - 1.6kb
E - 0.3kb
Vector seq.
D - 0.5kb
Ib - 1.4kb

Figure 4.9B



4.7 **Summary**

PCR analysis of genomic DNA, (from wild type, heterozygotes and homozygotes), was used to confirm the integration of the gene-trap vector into an endogenous exon. The endogenous sequence trapped by the R124 gene-trap integration was used to screen a mouse heart cDNA library to isolate cDNA clones. Sequencing analysis was performed on the isolated sequences (1004 bases 5' to the integration) to classify the trapped gene. No ORF, no functional motif and no homology with known genes in database searches was found suggesting the isolated sequence was 5'UTR and novel. Northern blot analysis confirmed the trapped sequences were expressed. Restriction fragment length polymorphism (RFLP) analysis was used to show linkage of the gene-trap vector to the trapped gene at the genomic level. This aided in the prediction of the structure of the R124 integration in the mouse genome. In addition, animals could be genotyped on the basis of distinct RFLP patterns for wild type and trapped alleles.

4.8 Discussion

Integration into an Exon

PCR analysis and RFLP analysis of the trapped allele genomic DNA verify integration into an exon and that the endogenous trapped gene sequences are directly linked to vector sequences at the genomic level. Direct linkage of vector sequences and endogenous trapped sequences allows the determination of RFLPs that distinguish the wild type and trapped alleles. As a mutagen an exon integration has the theoretical advantage over intron integrations that it may directly disrupt DNA coding sequence whereas intron integrations are dependent upon splicing. Examples of intron integrations which give rise to wildtype transcripts from trapped alleles via splicing around vector integrations have been documented eg. *cordon-bleu* (Gasca *et al.*, 1995), R213 (McClive *et al.*, in press 1998). However this is weighed against the generation of multiple splice variants caused by the juxtaposition of two splice acceptor sequences in exon integrations. It remains to be seen whether in practice intron or exon integrations are more effective as mutagens.

Endogenous Trapped Sequences

The trapped gene sequence (320bp) isolated using RACE shows no ORF and the size of the fusion transcript (1.1kb+*lacZ* 3.3kb) would suggest the gene-trap vector may have integrated into the 5' untranslated region (UTR) of the endogenous gene. Furthermore the cDNA (813bp) isolated from a heart library using KXE as probe, lies 5' to the integration site, does not reveal an ORF. Although no ORF has been identified inframe with the *lacZ* sequence, translation of β-gal protein can be initiated from the ATG engineered 5' to the *lacZ* in the vector. An ORF has been identified out-of-frame with *lacZ* but this does not conform to the Kozak consensus thus is unlikely to be the initiating methionine (Kozak, 1987).

The sequences isolated from the endogenous trapped gene by the R124 integration show no homology to any known genes but a number of ESTs have been identified which show high homology (Table 4.1). These ESTs are derived from different tissue cDNA libraries and can be used for "virtual northern blot" analysis i.e. gives an idea of tissues in which the trapped gene may be expressed. The gene sequences trapped by the R124 show homology to ESTs isolated from heart (fetal and adult), testis and kidney tissues which is consistent with tissues in which reporter gene activity was detected for the R124 integration (Chapter 3). In addition, EST sequences can be used to extend the sequence information of the trapped gene. However care must be taken to ensure the EST is not from a splice isoform or related gene, or that it does not arise from a chimeric cDNA clone.

It would appear that clone TRI is chimeric. Whether the subclone of TRI, TnI, is chimeric is yet to be determined. This cDNA contains homology to KXE and to myosin heavy chain gene alpha (MHCα), which is expressed exclusively in the heart. However, the trapped gene maps to chromosome 5 while the MHCα gene has mapped to chromosome 14 (Beisel *et al.*, 1989). It is not inconceivable that another related MHC has been isolated but northern blot analysis with the region of homology displays transcript sizes corresponding to MHCα transcripts (data not shown) and not to transcript sizes detected in common with KXE and VRI cDNA clone. This suggests strongly that TnI is also a chimeric cDNA, but differs from the chimeric TRI clone in that it has no adaptor sequences within it. The 25bp homology found in the KXE and TnI is found to be highly conserved in all MHCs from this it can be inferred that TnI may have been generated by some homologous recombination event during the preparation of the cDNA library. The reason why this 25mer is conserved in MHCs throughout the animal kingdom is unknown and whether this homology is significant to the identity or the function of the trapped gene is also still unknown. However, in database searches MHC genes showed a single site of homology with this 25mer,

none displayed an inverted repeat suggesting that although the sequence is conserved the inverted repeat motif is not.

The identification of ESTs with homology and detection of transcripts by northern blot analysis with the KXE and VRI indicate that the gene-trap vector has integrated into the transcribed region of the trapped endogenous gene and is thus likely to be mutagenic. To proceed further it will be important to isolate the full cDNA of the trapped gene and identify any functional motifs of the gene. A 4.4kb transcript exclusive to heart tissue is detected by northern blot analysis with the cDNA clone VRI thus a heart cDNA library could be screened using VRI as a probe to identify the remaining sequences of this heart specific cDNA.

Distribution of Endogenous Transcripts

Northern blot analysis of wild type RNA, isolated from adult tissues, was performed to determine the expression of the endogenous trapped gene sequences using probes generated from the VRI cDNA and KXE RACE cloned endogenous sequence. Transcripts of a common size (4.4kb and 1.4kb) were detected in the heart RNA and a 1.4kb transcript was detected in RNA with both probes. Alternative transcripts were detected in other tissues, VRI detected a 1.4kb transcript in all tissues tested in contrast KXE detected a 1.4kb transcript in the heart, kidney, testis and brain while 4.4kb and 2.0kb transcripts were detected in all tissues tested. Detection of the additional transcripts with the KXE probe may be due to the sequences which extend 3' to the VRI homology and are exclusive to KXE. The detection of alternative transcripts suggests that alternative splicing may be involved in generating different transcript isoforms from the same gene cistron. This can be inferred because the linkage of the VRI and KXE in the genome was shown by RFLP analysis. The probes detecting transcripts from related homologous genes is unlikely because Southern blot analysis indicates the trapped gene is a single copy gene.

Diversity of transcripts could be dissected by the isolation of tissue specific cDNA sequences and comparison of these to isolate transcript specific probes. Identification of tissue specific transcripts probes will be essential to characterise the distribution of the endogenous gene in the adult and during embryogenesis, using northern blot and *in situ* hybridisation analysis. Specific probes would allow RNAse protection assay experiments, a technique that allows detection at a higher specificity and thus is more conclusive than northern blot analysis, to be performed to confirm the distribution of transcripts. Knowing the distribution of specific transcripts will allow speculation of the function of the trapped gene and interpretation of any phenotype that may arise due to the integration of the gene-trap vector.

The distribution of the endogenous transcripts was in a range of tissues wider than that observed with reporter gene analysis. One could infer that detection of the reporter gene activity was not optimal and thus it did not reflect the expression of endogenous trapped gene. In attempt to address this, northern blot analysis, to detect fusion (trapped endogenous and *lacZ* sequence) transcripts using *lacZ* specific probes, was performed (data not shown). The primary problem encountered was fusion transcripts could not be detected in RNA from tissues, such as the kidney, that express the reporter in restricted cell types, as determined from histological staining. In contrast, a 4.4kb fusion transcript was detected in RNA from tissues displaying homogenous reporter gene activity e.g. heart. A simple explanation would be that RNA from non-expressing cells dilutes the fusion transcripts from expressing cells in tissues with restricted reporter gene activity, such as the kidney, to levels undetectable by northern blot analysis. Abundance and stability of the fusion transcript seem critical in the detection of fusion transcripts. Thus it may prove more reliable to characterise the distribution of fusion transcripts by *in situ* hybridisation or β-gal fusion protein detection using antibodies.

Alternatively the problem is not of detection but because the reporter has not "tagged" all the transcripts transcribed from the trapped gene i.e. multiple alternative transcript isoforms are generated from the trapped gene by splicing all of which do not contain the exon into which the vector has integrated, thus reporter expression is not detected in the full range of tissues. This would be consistent with trapping for example, the 1.4kb transcript detected, by the KXE probe, in tissues in which the reporter gene activity was detected by histology. To test this postulation one would have to identify the 1.4kb transcript specific sequence. This could be done by performing 5' RACE on RNA isolated from these tissues using vector specific primers. Protection of fusion (vector and transcript specific sequence) transcripts in RNA from these tissues by RNAse protection analysis would allow this postulation to be verified.

4.9 Conclusion

The gene-trap vector has integrated into the exon of a novel gene in the R124 gene-trap integration and is likely to have disrupted the function of the endogenous gene. A total of 1004 bases of the endogenous trapped gene sequence 5' to the integration site were isolated. These sequences show no homology to any known genes but a number of ESTs with homology were identified in database searches. Thus, it will be essential to isolate more sequence to identify functional domains of the gene. Northern blot analysis confirmed the isolated sequences were expressed and that they detected multiple transcripts in a range of adult tissues wider than that detected by reporter gene activity. However the distribution of one transcript, detected by the RACE cloned KXE sequence, resembled the distribution of reporter gene activity suggesting that the R124 integration may have trapped an alternatively spliced transcript of an endogenous gene.

Chapter 5

5.0 **Phenotype Analysis**

Introduction

The molecular analysis of the R124 gene-trap integration has shown that the gene-trap vector has integrated into an exon of a gene and may have disrupted the function of the endogenous gene (Chapter 4). Therefore the phenotype of homozygous animals has been analysed, paying particular attention to tissues where reporter gene expression was detected (Chapter3).

5.1 **Neonatal Lethality**

Genotyping intercross litters from the inbred, C57BL/6 and 129/CGR, backgrounds at 3 weeks of age revealed a statistically significant ($p<0.005$) deficit (57%) in the number of animals homozygous for the R124 gene-trap integration (Table 5.1). Genotype analysis of 19 d.p.c. litters (collected by caesarean) and neonatal animals (collected in the first 6 hours after birth) showed the expected Mendelian ratio of 1:2:1 suggesting homozygote animals are dying after birth and being cannabilised by the mothers. Close monitoring of litters immediately after birth resulted in the retrieval of eight dead pups, between 12-24 hours after birth, that were subsequently genotyped as homozygotes. Upon dissection it became apparent that the dead homozygous pups suffered from a localised dilatation affecting the right ventricle wall of the heart (Figure 5.1B). The localised dilatation was found in the same region of the right ventricle in all eight dead pups. Histological sectioning of these hearts revealed a localised thinning of the ventricle wall at that point where the dilatation appears, although the trabeculation in that region does not seem to be affected (Figure 5.1A). The thinning and localised dilatation of the right ventricle is suggestive that the pups are dying because of a right ventricle dysfunction. This defect has never been observed in any wildtype or heterozygote animals (to date 29 animals from backcross litters and 37 wild

Figure 5.1: Neonatal Phenotype

A. Section of a dead homozygote neonatal heart with a right ventricle dilatation.

B. A heart from a dead homozygote neonatal heart with a right ventricle dilatation stained for reporter gene activity.

C. Section of control neonatal kidney.

D. Section of a dead homozygote neonatal kidney showing blood congestion in the veins of kidneys .

E. Section of a dead neonatal lung reveals no abnormalities or congestion of blood. arrow, dilatation; rv, right ventricle; rvw, right ventricle wall; at, atrium; ve, vein; gl, glomerulus; al, alveoli; br, bronchiole; scale bar, 100μm.

Figure 5.1A

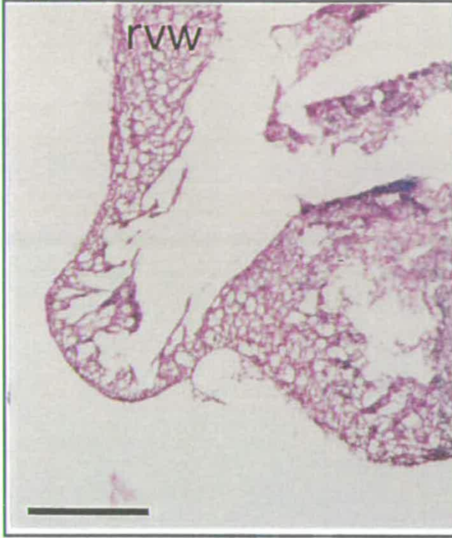


Figure 5.1B

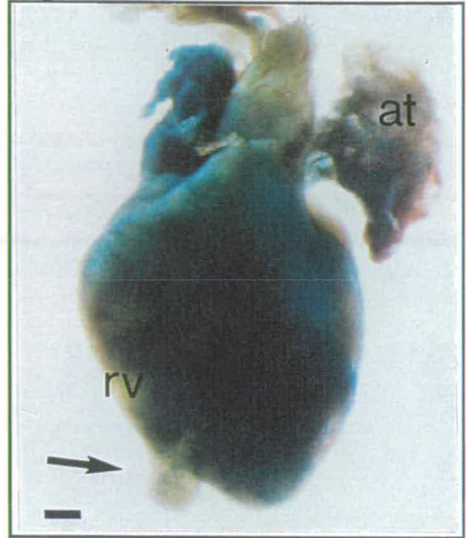


Figure 5.1C

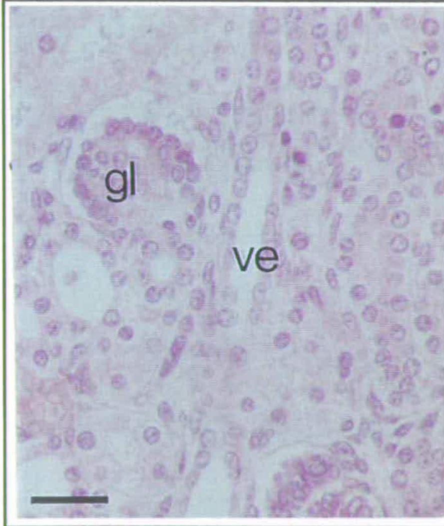


Figure 5.1D

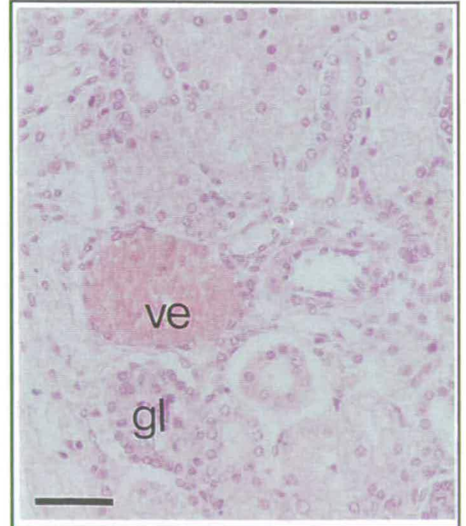
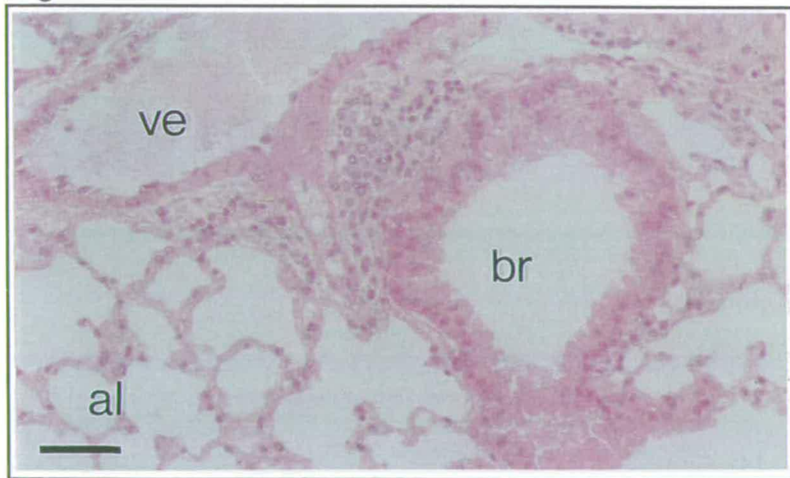


Figure 5.1E



type/heterozygote animals from intercross litters have been analysed), nor in eight live homozygotes sacrificed in the first 24 hours after birth. Congestion of the veins in the kidney of dead homozygotes has also been observed (Figure 5.1D) which would be consistent with right ventricular dysfunction. Conversely, no abnormalities or congestion of blood have been observed in the lungs (Figure 5.1E) of dead neonate homozygotes ruling out any left ventricle dysfunction.

When the R124 integration was bred onto a MF1 outbred background no loss of homozygote animals was observed at 3 weeks of age (Table 5.1). Examination of neonatal hearts collected from thirteen outbred homozygote pups sacrificed between 12-24 hours after birth revealed no dysmorphogenesis of the right ventricle.

	Genotype			% loss of		
Stage	+/+	+/-	-/-	-/-	χ^2 Test	P-Value
19 d.p.c.	20	39	15	24	1.13	0.56836
newborn (0 - 6 hours)	16	45	20	2	1.39	0.49908
3 weeks - Inbred						
- 129/CGR	45	79	20	52*	11.48	0.00322
- C57BL/6	26	54	9	66*	11.76	0.00280
- Total	71	133	29	57*	22.56	0.00001
- Outbred						
- MF1	50	121	45	21	3.82	0.14808

*Statistically Significant

Table 5.1: Genotyping of R124 Intercross Litters

Numbers of animals genotyped as wild type (+/+), heterozygous (+/-) or homozygous (-/-) from R124 intercross litters. Inbred intercross numbers are the sum of F1 to F5 generations, a similar proportion of homozygotes were lost in each generation. Outbred intercross numbers are from the F1 generation.

5.2 Adult Phenotype

The neonatal lethality is not fully penetrant and 40% of animals homozygous for the R124 gene-trap integration survive to adulthood. The ratio of males to females surviving (1:1) was as expected suggesting equal numbers of each were lost at the neonatal stage. At the gross morphological level, homozygotes surviving to adulthood did not show any major dilatation affecting the right ventricle of the heart but dysmorphogenesis of the right ventricle wall in the same region has been observed in 3 out of 6 male and 2 out of 10 female surviving homozygotes (Figure 5.2C). Upon sectioning two hearts, with right ventricle dysmorphogenesis, disarray of the myofibrils was observed in this region (data not shown).

In addition, all tested R124 homozygote males (4 C57BL/6 and 8 129/CGR background) surviving to adulthood displayed enlarged hearts (65% increase) (Figure 5.2B and C) and kidneys (52% increase) (Figure 5.3A) compared to control male sibling (heterozygote and wildtype) organs (Table 5.2). No overall difference in body weight between siblings was observed. This fully penetrant enlargement phenotype in inbred homozygous males was not observed in the inbred homozygous females (10) nor in homozygous males or females from the outbred MF1 background.

Gross histology of the enlarged hearts and kidneys displayed no obvious defect but presented a general overgrowth syndrome due to hypertrophy and hyperplasia. Detailed histology of kidney tissue [carried out in collaboration with Dr.Stewart Fleming] revealed the presence of membrane bound vesicles within the cytoplasm of the proximal tubules (Figure 5.3C), which are never observed in normal kidney tissue (Figure 5.3B). In addition, obliteration of the capillary tuft and expansion of the mesangial cells associated with glomeruli was observed (Figure 5.3E). The pathology of the glomeruli is often associated with hyperfiltration. This suggested that the "overgrowth syndrome" may be due to a defect in water to salt balance in kidney function and or blood pressure. This led to preliminary experiments to address

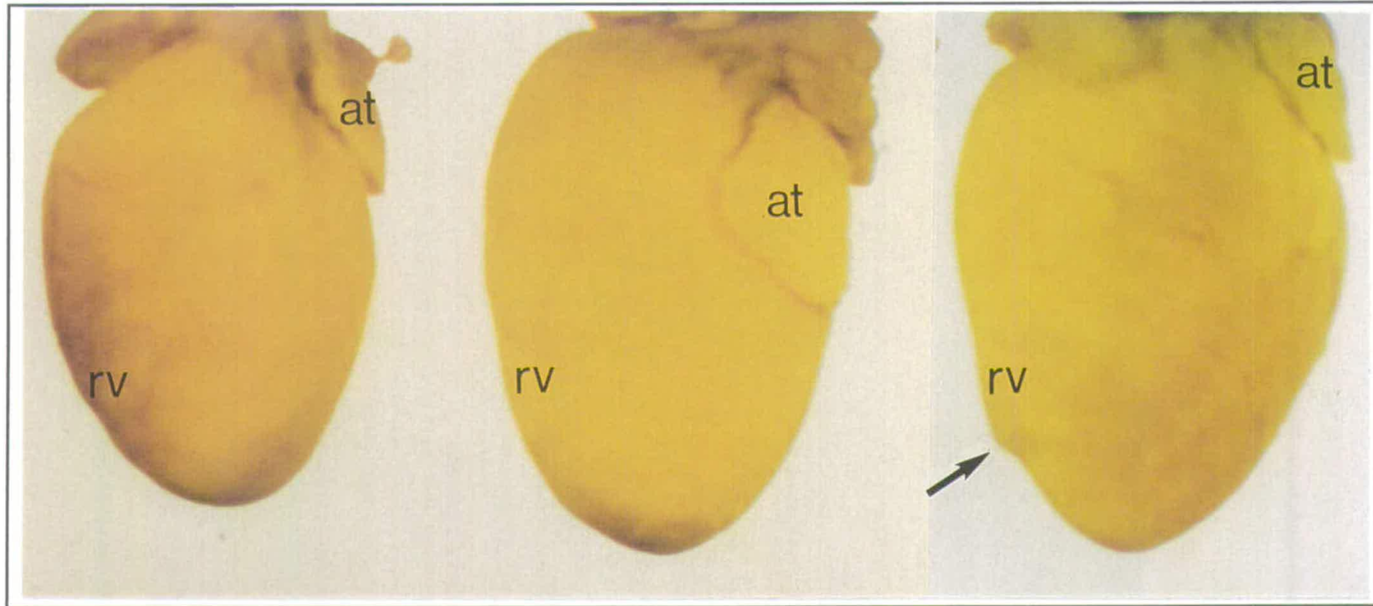
Figure 5.2: Enlarged Hearts

Hearts collected from adult inbred males A. control, B. and C. homozygous.

C. demonstrates dysmorphogenesis of the right ventricle wall observed in 30% of homozygotes.

arrow, dysmorphogenesis; rv, right ventricle; at, left atruim.

Figure 5.2



Right ⟷ Left

1mm

Figure 5.3: Kidney dysmorphogenesis

A. Kidneys collected from inbred adult wild type (+/+) and homozygote (-/-) males .

B. Section of control kidney proximal tubules (scale bar, 50µm).

C. Section of homozygous kidney showing membrane bound vesicles in proximal tubules (scale bar, 50µm).

D. Section of control kidney glomerulus (scale bar, 10µm)

E. Section of homozygote kidney glomerulus showing obliteration of the capillary tuft of the glomeruli and disorganization of associated mesangial cells (scale bar, 10µm). arrow, vescicles in proximal tubules; ct, capillary tuft; me, mesangial cells.

Figure 5.3A

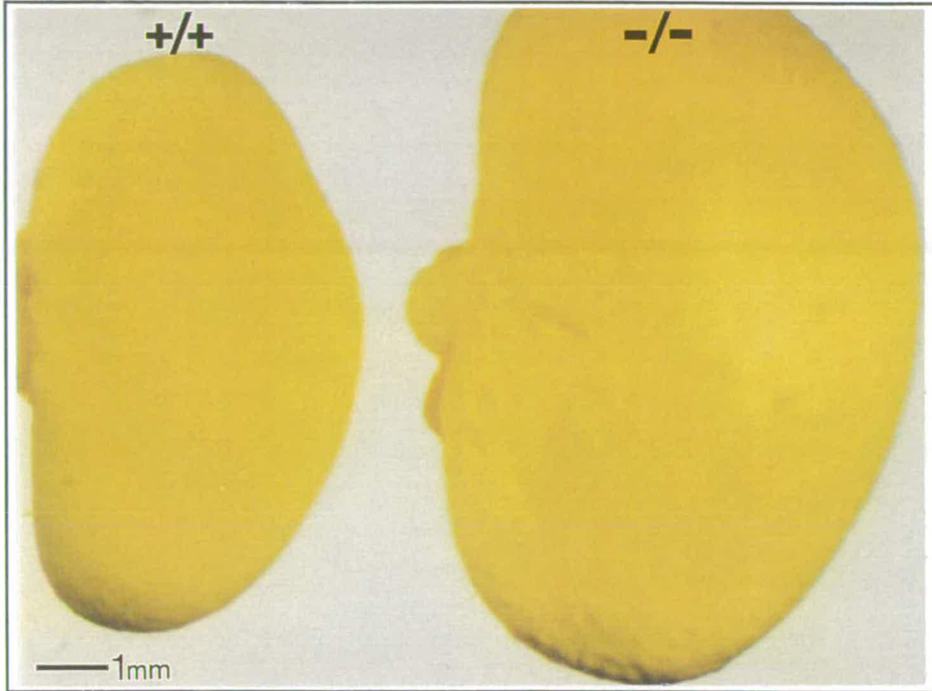


Figure 5.3B

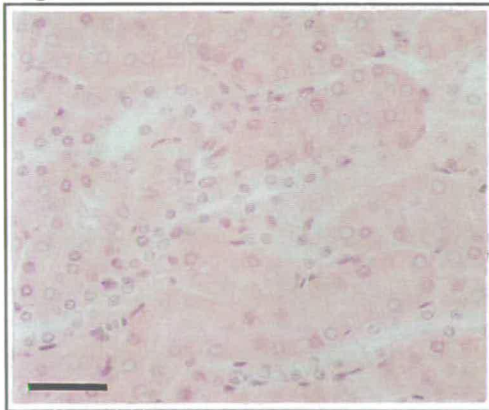


Figure 5.3C

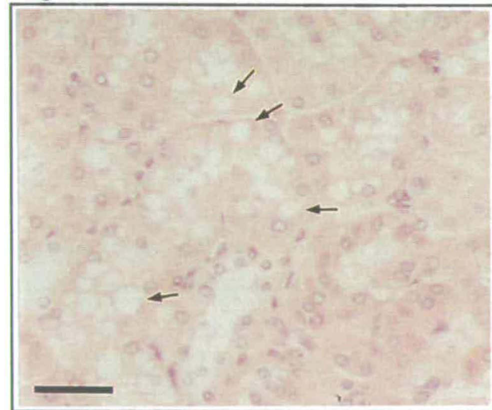


Figure 5.3D

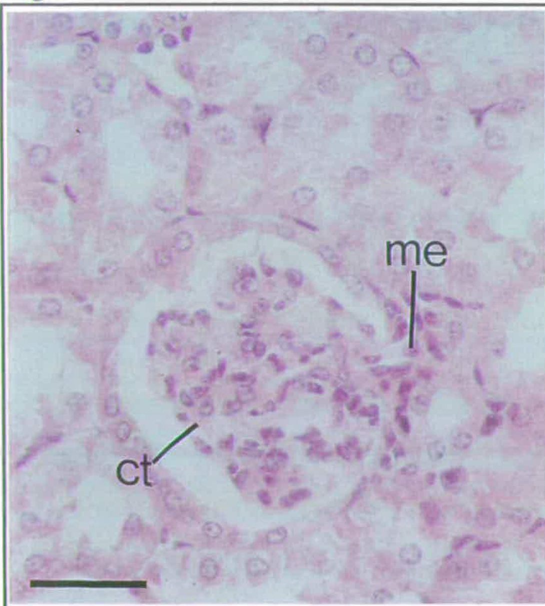
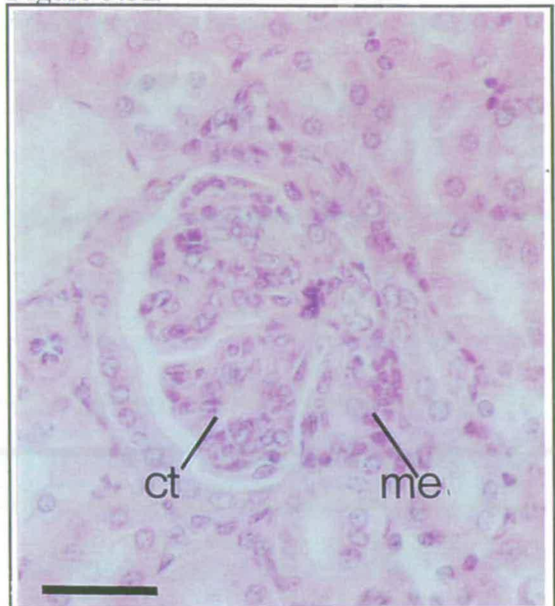


Figure 5.3E



Animal	Genotype(n)	Mean Weights ± Standard Error Heart(mg)	Kidney(mg)	Body(g)	Weight Ratio (10^{-3}) Heart/Body	Kidney/Body	Increase (%) Heart	Kidney
inbred males	-/- (10)	277±18	607±36	35.17±1.85	8.1±0.7	17.6±1.4	65*	52**
	+/- (10)	172±11	401±35	36.41±2.08	4.9±0.5	11.6±1.5		
inbred females	-/- (9)	192±9	393±13	30.70±1.03	6.3±0.4	12.8±0.6	0	-7
	+/- (9)	193±6	412±15	30.22±0.67	6.3±0.3	13.8±0.6		
outbred males	-/- (5)	195±11	566±45	37.78±2.78	5.2±0.5	15.0±0.7	0	-5
	+/- (5)	216±12	607±14	39.22±1.21	5.2±0.2	15.8±0.7		
outbred females	-/- (6)	152±7	417±17	29.68±0.77	5.2±0.2	14.2±0.3	4	1
	+/- (6)	151±7	432±23	30.68±1.08	5.0±0.4	14.0±0.6		

T-Test $*p < 0.0001$, $**p < 0.005$

Table 5.2: Organ and Body Weights

Organ and body weights of homozygote (-/-) and heterozygote (+/-) males and females on inbred and outbred genetic backgrounds.

whether kidney function and blood pressure regulation was impaired in surviving R124 homozygous males.

5.3 **Kidney Function and Blood Pressure Evaluation**

To evaluate the function of the kidney, homozygous males (10) and their control wild type littermates (10) were placed in individual metabolic cages overnight (with water but no food) to collect urine samples. The volume of water intake and urine produced per animal was recorded and urine osmolarity was measured (Chapter 2 Section 2.3.3) (Table 5.3). A significant increase in the osmolarity of urine produced by homozygote males when compared to controls was observed. No significant difference between control and homozygotes in volume of fluid intake or secretion was observed. This increase in osmolarity would suggest that there was a higher than normal salt concentration in the urine, implying less salt was being resorbed by the kidneys.

The change in osmolarity may be an indication that blood pressure is altered in the homozygous animals and so experiments are underway to compare the blood pressure in wildtype, heterozygote and homozygote siblings.

	Number of Animals	Osmolarity (Osmo/Kg)	Water Intake (ml)	Urine Secretion (ml)
Controls	10	1.21 ± 0.16	7.49 ± 0.29	2.34 ± 0.25
Homozygotes	10	1.98 ± 0.32*	6.74 ± 0.55	2.09 ± 0.23

* T-Test $p = 0.0047$

Table 5.3: Metabolic Measurements
Comparison of osmolarity, water intake and urine secretion measurements from control and homozygous adult inbred males.

5.4 Male Infertility

Homozygous males (3 C57BL/6 and 5 129/CGR background) for the R124 integration surviving to adulthood were mated to MF1 females and although they each plugged several females, none sired any offspring indicating that they were infertile. Histological analysis of testis collected from all inbred homozygotes, surviving to adults, showed there were no mature elongated spermatozoa within the semiferous tubules (Figure 5.4B and 5.4D). In contrast many round immature spermatocytes could be seen in the tubules and epididymis. This indicates that the production of sperm, spermatogenesis, was not affected but maturation of sperm, spermiogenesis, was arrested. *In situ* hybridisation analysis [carried out by Philipa Saunders] with a stage specific sperm marker (Rat-*TP2*) has shown that no sperm past stage 7 of maturation are present within these seminiferous tubules (Figure 5.4C and 5.4D). The phenotype observed shows that the supporting role of the Sertoli cells in sperm maturation was affected in homozygote males carrying the R124 integration. Morphologically the Sertoli cells appear normal, in contrast generalised leydig cell hyperplasia and an elevated number of germ cells undergoing apoptosis was observed (Figure 5.4B).

Surviving homozygote females (2 C57BL/6 and 4 129/CGR) from the inbred background were mated with wildtype males to check fertility. They were found to be fertile and gave birth to average size litters, which when genotyped as heterozygotes confirmed the homozygous status of the mother.

Homozygote adults from the MF1 outbred background were also tested for fertility, both males (7) and females (7) were found to be fertile.

Figure 5.4: Male Infertility

A. Section of inbred wild type male testis.

B. Section of inbred homozygote male testis.

C. Section of inbred wild type male testis hybridised to dig-labelled *TP-2* riboprobe.

D. Section of inbred homozygote male testis hybridised to dig-labelled*TP-2* riboprobe.

arrow, cells undergoing apoptosis; st, seminiferous tubules; sg, spermatogonia; sp, mature tailed spermatocytes; le, leydig cells; scale bar, 100μm.

Figure 5.4A

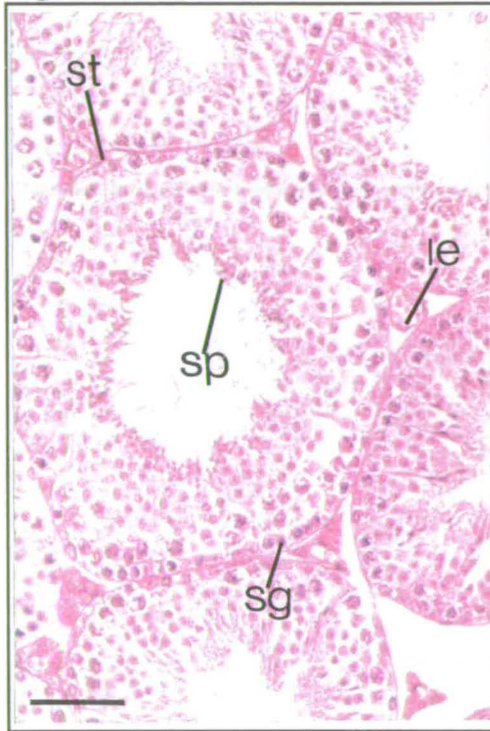


Figure 5.4B

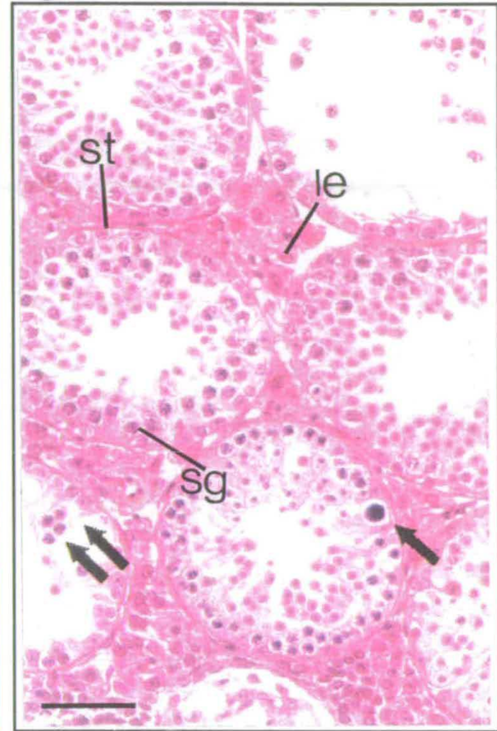


Figure 5.4C

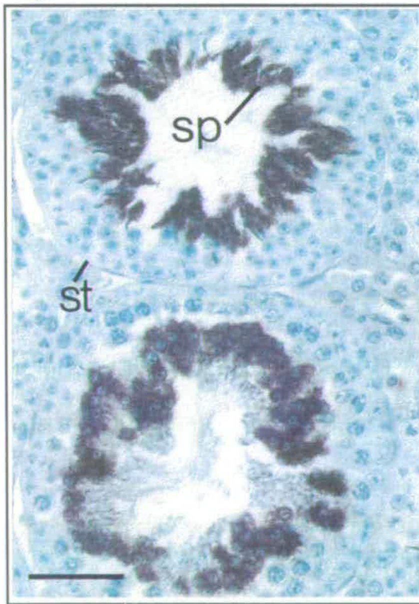
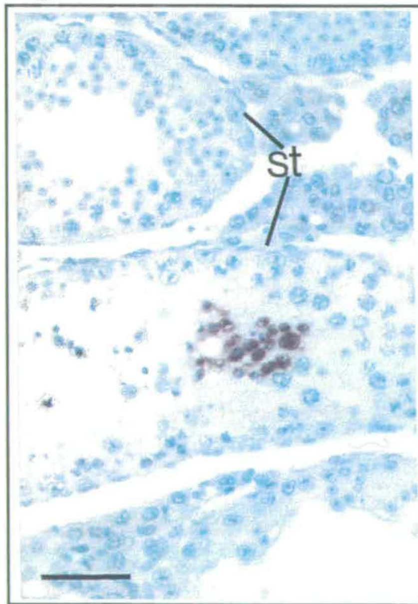


Figure 5.4D



5.5 Summary

57% of neonates homozygous for the R124 integration, on inbred backgrounds, die between the first 12-24 hours after birth. This incomplete penetrance lethality was associated with a localised right ventricle dilatation and congestion of blood in the renal veins. The proportion of homozygotes surviving to adulthood had equal numbers of males and females. None of these displayed the right ventricle dilatation but ~30% displayed dysmorphogenesis of the right ventricle wall. Hearts and kidneys collected from all inbred homozygous males were enlarged. The increase in hearts was characterised by generalised hypertrophy. The increase in kidneys was characterised by hypertrophy, hyperplasia and the presence of aberrant membrane-bound vesicles in the proximal tubules. These mutant kidneys also displayed defective glomeruli. Osmolarity measurements indicated that the inbred male homozygotes were secreting abnormally high levels of salt in their urine. The inbred males were also found to be infertile due to a defect in sperm maturation. None of the adult phenotypes have been shown to affect inbred females, outbred males and females homozygous for the R124 integration.

5.6 Discussion

Neonatal Phenotype

The right ventricle heart defect observed in neonates homozygous for the R124 integration occurs later than woul be predicted from the early expression of the reporter gene in the primitive heart tube and is more restricted in comparison to the homogenous expression of the reporter gene throughout the developing heart. This suggests that genes with a redundant function may be expressed in an overlapping pattern with the trapped gene or that the function of the gene is not essential in all sites of expression. In this respect it resembles null mouse mutants generated for the early cardiac markers *MEF2c* and *Nkx2.5* both of which display phenotypes later than anticipated and more restricted in distribution than their expression patterns during heart development (Lin *et al.*, 1997; Biben *et al.*, 1997). 57% of the neonates homozygous for the R124 integration die between 12-24hours after birth. This lethality is associated with a localised right ventricle dilatation and congestion of blood in renal veins indicating that the neonates die from right ventricle dysfunction/failure. Accordant with this phenotype the R124 gene-trap integration has been named *d*efective *h*eart *i*nduced *l*ethality 1 (*dhil1*), [dhil (punjabi) translates to heart].

Why the right ventricle dilatation is localised in a particular region cannot be explained from the reporter gene expression and published data. In normal development after the specification of the primitive ventricle the ventricular epicardial surface of the wall (known as the compact zone) undergoes expansion to form the thickened ventricle wall, the trabeculation within the chamber and contributes to the formation of the interventricular septum between the future right and left ventricles. In addition no cardiac markers have been described which are expressed in the right ventricle in such a restricted fashion. Thus, no difference is apparent from the morphogenesis of the heart or from cardiac markers to indicate that the region defective in the R124 homozygotes is distinct from any other region of the ventricle wall. A number of

mouse mutants with right ventricle dilatations have been documented but none display localised dilatations. For example mutations in RXRα show hypoplasia of the ventricular compact zone and ventricular septal defects which in combination compromise the efficiency of the heart to function causing the right ventricle wall to dilate and ultimately lead to the death of mutants on the 15th day of gestation due to a form of congestive heart failure (Sucov *et al.*, 1994; Dyson *et al.*, 1995). General ventricular dilatations are often associated with changes in hemodynamic forces affecting the heart which arise due to septation defects and lead to hypoplastic growth of the chamber walls. It is difficult to explain the localised dilatation in R124 neonatal homozygotes by altered hemodynamic forces because no associated septal defect was observed. Proposing a developmental defect in the right ventricle wall would involve predicting that the gene trapped by the R124 integration interacts with regionally specific gene products that are part of a pathway which specify (morphogenetically or functionally) the region of the ventricle involved in the dilatation. Thus removal of the trapped gene product would result in a deficiency. The dilatation was not observed in live homozygotes before the lethal period suggesting that although the primary defect may be a structural deficiency in the right ventricle wall (will be investigated at the ultrastructural level), the dilatation was induced/aggravated by some "crisis" event after birth. A "crisis" event may arise because of the environmental and biological changes that take place in the first 24 hours of postnatal life eg. competition between siblings for suckling, inflation of the lungs and morphogenetic changes within the heart to partition oxygenated/deoxygenated blood causing changes in hemodynamic forces affecting the heart. Crisis events revealing phenotypes affecting muscle function have been documented for *MHCα* (Jones *et al.*, 1996), *MLC2V* (Fewell *et al.*, 1997) and *MyoD* (Rudnicki *et al.*, 1992) null mutants. The functional capacity of the mutant hearts (homozygote and heterozygote) could be determined using a working heart model system (described in Grupp *et al.*, 1993). This preparation allows physiological

measurements, such as ventricular pressure and volume ejections, to be made during the cardiac cycle (systolic and diastolic) and application of external forces to assess heart function. This approach will also be useful to analyse the functional heart capacity of homozygous animals that survive to adulthood.

The neonatal lethality was not fully penetrant, 40% of homozygotes from inbred backgrounds surivive. In addition no significant loss of homozygotes on the outbred MF1 background was observed. Survival of homozygotes is likely to be due to heterogenity in the genetic background of the animals. Inheritance of modifiers, unlinked to the trapped gene, maybe compensating for the mutation generated by the R124 integration. Many examples of the effects of strain background on mutant phenotypes have been documented e.g. *Egfr* (Threadgill *et al.*, 1995), *TGFβ1* (Bonyadi *et al.*, 1997); *CRABPII* (Fawcett *et al.*, 1995). For example, null mutations in the epidermal growth factor receptor (*Egfr*) gene (Threadgill *et al.*, 1995): on an outbred CF-1 background mutant embryos have defects in the inner cell mass that cause peri-implanation lethality; on the 129/SvJ inbred background mutants have placental defects that cause mid-gestation lethality; and on a congenic C57BL/6J inbred background defects in various organs cause juvenile lethality. Further backcrossing of the R124 integration onto pure 129/CGR and C57BL/6 is ongoing to generate congenic strains. Comparison of simple sequence length, restriction length and chromosomal marker polymorphisms associated with the penetrance of the phenotype on 129/CGR, C57BL/6 and MFI background strains could be used to identifiy genetic modifiers that affect the phenotype.

Adult Phenotype

Homozygote neonates surviving the lethal period after birth show no dilatation but hearts from some homozygote adults (~30%) have displayed dysmorphogenesis of the right ventricle wall in the same region. In these cases a localised disarray of myofibrils

and congestion of blood in renal veins has been observed suggesting that the right ventricle is dysfunctional in these animals. However the lack of severity observed in adult homozygotes maybe because they inherited a genetic modifier(s) which compensates for the mutation caused by the R124 integration allowing these animals to survive the "crisis" event in the critical period after birth. Recovery from ventricular failure due to ischemia and infarction is often associated with compensatory hypertrophy (Beltrami *et al.*, 1994). Significant levels of hypertrophy of the ventricle walls was observed in adult males but not in adult females. The levels of hypertrophy associated with right ventricle dysmorphogenesis may be masked in females by the majority of females not suffering a "crisis" event (2 hearts from 10 show dysmorphogenesis). However the males not only suffer from enlarged hypertrophic hearts but also from enlarged kidneys. The morphology of the mutant kidneys suggests renal dysfunction which was not observed in females. This is supported by preliminary physiological experiments which showed abnormally high levels of salt secretion in urine from homozygote males. Implications of these observations are that blood pressure may be significantly and detrimentally altered in the homozygotes. Thus studies are underway [in collaboration with Dr. John Mullins] to investigate blood pressure in homozygotes. Blood pressure, renal hemodynamics, fluid and electrolyte homeostasis are interlinked processes regulated by the renin-angiotensin system (Lee *et al.*, 1993). Characterisation of expression of the genetic components of the renin-angiotensin in R124 mutants will determine if this pathway is altered. The morphology of glomeruli in mutants shows the relationship of cells surrounding the juxtaglomerular apparatus (JGA) was abnormal. The JGA regulates the production of renin and angiotensin which has been shown to be expressed in juxtaglomerular (JG) cells and in proximal tubules (Gomez *et al.*, 1991; Burns *et al.*, 1993). The expression of these and other components of the renin-angiotensin pathway are influenced by the circulating steroid hormone and salt levels (Taugner *et al.*, 1989). Thus the prediction

would be that the morphological changes will be associated with altered levels of the renin-angiotensin pathway components resulting in abnormal levels of blood pressure. The alteration in blood pressure has significant effects on the cardiovascular system and is associated with hypertrophy of the heart (Johnson *et al.*, 1996). Thus the significant cardiac enlargement in males may result from a combination of hypertrophy to compensate for ventricular dysfunction and alteration in blood pressure due to renal dysfunction.

Sexual Dimorphism

The phenotype observed in adults homozygous for the R124 integration suggests a sexually dimorphic role of the trapped gene in adulthood. Evidence of differences in blood pressure between males and females have been documented and pathways regulating blood pressure are known to be influenced by steroid hormones (Chen, 1996). The sexual dimorphism of the adult phenotype has important implications in the practicalities of studying the neonatal and adult phenotypes. It will potentially allow the separation of the cardiac phenotype which stems from a "developmental defect" of the right ventricle, from the renal dysfunction-associated hypertrophy in adult animals. Thus the dysfunctional right ventricle can be studied in adult female mutants without the complication of secondary effects due to renal dysfunction. Future studies will include morphometric experiments, to measure size and characterise pathology of mutant organs, and physiological experiments, to measure urine osmolarity and blood pressure, over time in male versuses female homozygotes this will show when the male-specific phenotype arises and whether this has any correlation with the expression of male-specific steroids. An alternative approach would be to castrate homozygote males or conversely treat females with male-specific steroids e.g. testosterone, and determine whether male-specific steroids play a role in inducing the adult phenotypes.

Another possibility is that sex-specific genetic modifiers compensate for the mutation generated by the R124 integration in adult females. The modifiers modulating the penetrance of the neonatal lethality may be influenced by sex-specific changes after birth or the expression of other genes involved in the transitition from juvenile to adult may directly modify the mutation generated by the R124 integration. Investigation of this will also require the onset of the male-specific adult phenotype to be determined.

Male Infertility

The sexually dimorphic function of the trapped gene was further emphasised by the infertility of homozygous males but not of homozygous females. The expression of the reporter gene (Chapter 3) was not detected in homologous structures in the testis (Sertoli cells) and ovaries (primitive oocytes and mesenchymal cells) suggesting that it would play different roles in the males and females.

Sertoli cells are involved in the initiation, maintenance and regulation of spermatogenesis. These processes are regulated through cell-cell communication between Sertoli and germ cells via interactions involving the complex cytoarchitecture of the Sertoli cells and secretion of nutritional metabolites, growth factors (e.g. *IGF*s, *TGFβ3*, *Sl*, *PDGF*) and transport-binding proteins (e.g. transferrin, androgen-binding protein, cellular retinoic acid-binding protein) (Skinner, 1991). Physical interaction between the Sertoli cell cytoarchitecture anchors and compartmentalizes the different stages of germ cells as they mature and travel towards the lumen of the seminiferous tubules to be released as mature spermatozoa. Sertoli cell product secretions are synchronous with the spermatogenic cycle to acheive normal maturation. Loss of signals from Sertoli cells or premature release of germ cells could result in the arrest of germ cell maturation. Alternatively, if the germ cells cannot respond to maturation signals from the Sertoli cells a similar phenotype would arise. The detection of Sertoli cell markers (e.g. *RXRβ*, Kastner *et al.*, 1996;*WT1*, Pelletier *et al.*, 1991) in R124

mutants will allow the evaluation of Sertoli cell function and the effect of the R124 gene-trap integration. Targeting by homologous recombination of the transcription factor *CREM* (cyclic AMP-responsive element modulator) gene, expressed in postmieotic germ cells, to generate null homozygous mutant males has shown a phenocopy of the infertility observed in R124 homozygous males (Nantel *et al.*, 1996; Blendy *et al.*, 1996). Furthermore females homozygous for this mutation are fertile. This could indicate that the R124 gene-trap product may interact with the CREM pathway to regulate spermatogenesis. Experiments to address this interaction of the R124 trapped gene and CREM could include characterising the expression of each gene on the other mutant background to determine which gene is upstream or downstream. If it was found both were in the same pathway, and were not acting in parallel pathways, to cause the same defect it would be interesting to cross the R124 and CREM mutants to determine whether they complement and rescue the fertility defect in males.

No effect on fertility of homozygous females suggests the function of the trapped gene is not essential or is substituted by a functionally redundant gene product in female ovaries, which may include the involvement of genetic modifiers. This may also apply to the brain where reporter gene expression was detected but no gross defects (data not shown) or behavioural changes were observed in homozygous animals. However the studies on the brain were not exhaustive and further detailed experiments will be necessary to rule out any subtle defects.

Phenotypes versus Sites of Reporter Gene Activity

The phenotypes detected (hypertrophy of the heart; hypertrophy, hyperplasia, aberrant membrane-bound vesicles in proximal tubules and glomerular defects of the kidney, and arrest of sperm maturation) correlate with the sites of reporter gene activity detected in inbred adult males (Chapter 3). The exception was the brain, where

reporter gene activity was detected but no defect was discovered but this could be because more detailed studies are required to reveal subtle anatomical and/or behavioural defects. No gross defects were detected in tissues where no reporter gene acitvity was detected but one cannot rule out subtle defects. Complete expression analysis of the endogenous trapped gene will allow this question to be readdressed and detailed phenotype analysis will be performed on tissues which showed no abnormalities. No defects were observed in inbred females, outbred females or outbred male homozygotes although reporter gene activity was detected in the same tissues as the inbred male suggesting that these may be protected by inheriting genetic modifiers which compensate for the defect generated by the R124 integration.

Phenotype versus Endogenous Trapped Gene Sequences

Homozygotes were selected from intercross litters using a RFLP which allowed distinction between wild type and trapped alleles (Chapter 4). No segregation of the trapped allele and, neonatal and adult, phenotypes has been observed over five generations of breeding inbred strain animals indicating that the disruption of the trapped gene due to integration was the likely cause of the phenotypes and not due to breeding strain related recessive mutations to homozygosity. This could be confirmed by targeting the trapped gene by homologous recombination.

Northern blot analysis of wild type RNA using endogenous trapped gene specific sequence (Chapter 4) showed that the expression of the trapped gene was more complex than that inferred from the expression of the reporter gene. Interestingly phenotype analysis revealed the tissues expressing reporter gene are defective in animals homozygous for the R124 integration. This supports the postulation that the gene-trap vector has trapped a specific splice isoform of an endogenous gene. The expression of the isoform as indicated by reporter gene activity is restricted and it would seem it plays some unique function which cannot be compensated by the

alternative splice isoforms arising from the endogenous gene which are also expressed in reporter activity positive tissues bar the brain where no abnormality was detected. *In situ* hybridisation studies will complement northern blot analysis and may reveal that the different splice isoforms although expressed in the same organs, because total RNA from whole organs was used, they may be expressed in different cell types and thus the phenotype is restricted.

It will be important to isolate the complete sequence of the transcript trapped by the integration and to show by using sequences 3' to the integration as probes that this transcript is disrupted in homozygotes. Transcript specific probes will also allow the exclusive expression of this transcript to be confirmed. Another approach to confirm the multiple splice isoforms were transcribed from the trapped gene and not from a related gene would be to target the trapped gene by homologous recombination and characterise the distribution of transcripts. If transcripts corresponding to the alternative isoforms were expressed it would suggest other gene transcripts were cross-reacting with the probes used for northern blot analysis, alternatively if all transcripts were lost and the phenotype was restricted as with the gene-trap it would suggest the loss of the alternative isoforms may be compensated for in the other tissues. However, one would predict that the complete knock-out of the gene would result in a more severe phenotype affecting many tissues.

5.7 Conclusion

The gene trapped by the R124 integration is important in the development of the heart. In the adult it is involved in heart and kidney function, and also in male infertility. The defects observed affect tissues in which reporter gene activity was detected suggesting that the reporter was mimicking the endogenous gene or more accurately the transcript that it had trapped.

Lethality of neonates homozygous for the R124 integration was associated with a localised right ventricle dilatation and congestion of blood in the renal veins indicating these neonates were dying from right ventricle dysfunction/failure. This phenotype is not fully penetrant on the inbred background (57% loss) and no neonates are lost on the outbred background suggesting that the phenotype may be influenced by a genetic modifier(s). In surviving homozygote adults two phenotypes affecting the heart are observed: 30% suffer from right ventricle dysmorphogenesis which may be associated with the neonatal phenotype and is not fully penetrant; and the second, which is fully penetrant, causes enlargement of the heart in inbred homozygous males only. Homozygous males, on the inbred background, also suffer from abnormal kidneys associated with a change in urine osmolarity suggesting that the blood pressure of these animals may be abnormal. Inbred homozygote males are infertile due to arrest of spermatogenesis. The latter aspects of the adult phenotype suggests a secondary function of the trapped gene in adults which may be male specific. Conversely sex-specific genetic modifiers may be operating in the adult mutants which protect the females from the deficiency generated by the R124 integration.

Chapter 6

6.0 General Discussion

The data presented in this thesis demonstrates the successful characterisation of a gene-trap integration, R124, leading to the identification of a novel gene involved in heart development. This novel gene is also important for male fertility as well as heart and kidney function in the adult. The preliminary characterisation of the gene-trap integration R68 highlighted some of the problems encountered when using a gene-trap strategy to identify genes but *in vitro* analysis demonstrated that the reporter gene was expressed in cardiomyocytes.

Recent strategies to identify novel genes involved in heart development and function have included the generation of catalogues of ESTs from heart cDNA libraries (Fung *et al.*, 1996). In attempt to categorise novel ESTs into classes of gene products differential display has been used to determine which ESTs are expressed in the neonatal or adult heart and degenerate primers have been used to identify specific functional motifs, for example genes encoding zinc fingers (Wang *et al.*, 1997). Subtractive hybridisation has been employed by Ruiz-Lazano *et al.*, 1998 to identify target genes downstream of RXRα which may be involved in heart development. Many novel cDNAs (924 cDNAs Fung *et al.*, 1996; 8 cDNAs Wang *et al.*, 1997; 5 cDNAs, Ruiz-Lazano *et al.*, 1998) have been identified by these approaches, however the precise function and specificity to the cardiac lineage of these cannot be predicted until expression and phenotype analysis is performed. One should also be aware that many EST sequences are partial therefore may be represented in a pool multiple times by different segments of sequence, therefore the 924 novel cDNAs isolated by Fung *et al.*, 1996 may be an over estimation. This has been documented by Ruiz-Lazano *et al.*, 1998 where one novel sequence was represented 12 times.

Two large mutagenesis screens, using ENU as a point mutagen, have been performed in zebrafish (*D.rerio*) identifying an estimated 500 different mutants involved in heart development (Driever *et al.*, 1996; Haffter *et al.*, 1996). Mutants with defects affecting

many aspects of heart development have been identified but the isolation of genes, which will have to be done by positional cloning, is limited by the rate of segregating and identifying each mutated locus.

All these strategies will greatly benefit from the data accumulating from genome sequencing projects and hence the candidate gene approach to characterise and identify novel genes can be employed. The most informative analysis to dissect heart development has been the targeted disruption of known genes by homologous recombination in the mouse. Some genes shown to be important for heart development could not be predicted by their expression pattern e.g. Hox 1.5 (a3) (Chisaka *et al.*, 1991).

Thus a combination of studies, which include sequence isolation, expression analysis and phenotype analysis linked with a specific mutation event, are required to show the function of a gene within heart development or any other developmental process. These essential components to identifying novel genes are married in the gene-trap approach, as described in the introduction, thus allow the function of a novel gene to be addressed directly. However, there remain complications that arise when using this technology some of which are highlighted in this thesis. How some of these complications can be overcome are discussed and alternative strategies are proposed.

Molecular analysis of gene-trap integration R124 revealed the intron gene-trap vector, PT1.ATG, used to generate these gene-trap cell lines was functioning as an efficient exon gene-trap vector via the activation of cryptic SDs within the vector. This resulted in the incorporation of vector-derived *en-2* intron sequences in the fusion transcript between the reporter gene and endogenous trapped sequences. Integration of the vector into an endogenous exon was anticipated to generate a mutation and as shown animals homozygous for the R124 integration suffer from a number of defects concomitant with reporter gene expression. Thus it may prove fruitful to study unconventional exon gene-trap integrations but only if the trapped sequences are clonable, for example

in the case of R124. As demonstrated by Townely *et al.*, 1996, RACE products generated from fusion transcript containing vector-derived intron sequence can be identified and the ability to clone sequences can be determined by using the direct sequencing approach. This screen showed that 20% (23/115 lines) of the gene-trap cell line RACE products contained intron sequences and in a smaller screen from our lab we found 23% (3/13 lines) of lines contained intron sequences. Gene-trap integrations classified as unclonable (30%) by direct sequencing showed superimposed sequences suggesting more than one splicing event was occurring in these lines. This could be due to a number of scenarios e.g. mixed ES cell clones; more than one integration; because the endogenous gene undergoes alternative splicing or as observed in the R124 gene-trap cell line there is competition between cryptic splice sites that have been activated due to the integration. Another problem encountered when RACE cloning R68 and R124 gene-trap integrations was products arising from unspliced primary transcripts which contain intron sequences and thus complicate RACE procedures by giving false impressions of more than one integration event. This could be eliminated by using polyA RNA for RACE procedures. This proportion of gene-trap cell lines containing intron sequences could be classified by using a range of probes specific to the 5' end of the vector-derived intron sequence and correspondingly primers could be designed to identify endogenous trapped sequences by RACE or direct sequencing. To pursue characterisation of lines containing multiple RACE products one would have to justify the merit of each particular integration and select lines by some additional parameter for example lineage specific reporter gene expression. This could be done efficiently using *in vitro* differentiation protocols or by the generation of chimeras.

Many screens have been conducted using reporter gene expression, *in vitro* (Baker *et al.*, 1997; Johnson *et al.*, 1993) and *in vivo* (Wurst *et al.*, 1995; Hill *et al.*, 1993), to determine the lineage specificity of the endogenous trapped gene and thus to select gene-trap events on the basis of an interesting expression pattern. These screens rely

on the gene-trap vector behaving as predicted but as discussed above that is not always the case. In addition to problems of cloning trapped sequences, one should be aware of the possibility that the reporter could be silenced and not reflect the endogenous trapped gene expression. This is dependent upon the genomic context of the integration and may be complicated by alternative or cryptic splicing which could result in splicing around the vector and thus no expression of the reporter. Silencing of reporter gene expression can also result from epigenetic modifications such as methylation or imprinting.

In the R124 gene-trap integration the reporter gene expression in the adult was restricted to the heart, kidney, brain and gonads, and a single fusion transcript was detected. However multiple endogenous transcripts are detected in a wider range of tissues with trapped gene-specific sequences suggesting in the first instance that the reporter is not mimicking the expression of the endogenous gene. However there is one alternatively spliced transcript that is expressed in a pattern similar to the reporter gene activity. Thus it is possible that a single alternative splice variant has been trapped by the gene-trap vector in the R124 integration. This is supported by animals homozygous for the R124 integration displaying defects only in organs where reporter gene activity was detected.

Reporter gene expression based gene-trap screens can be further defined by the addition of preselection steps e.g. induction screen by Forrester *et al.*, 1996 or by the use of modified gene-trap vectors e.g. secretory-trap screen by Skarnes *et al.*, 1995 to detect specific classes of genes. To optimise these more defined screens one would propose the addition of a penultimate selection step of direct sequencing to determine which integrations would be clonable before initiating characterisation of gene-trap cell lines.

Another issue of the gene-trap strategy is whether the integration of a gene-trap vector is mutagenic. Again this will be affected by the genomic context of the integration and

the transcriptional regulation of the trapped gene. Three possibilities that can arise have been documented i.e. the vector interrupts the endogenous gene and causes a mutation as predicted e.g. *jumonji*, Takeuchi *et al.*, 1995; the vector interrupts the endogenous gene but alternative splicing results in the generation of wild type transcripts e.g. R213, McClive *et al.*, 1998 and *cordon-bleu*, Gasca *et al.*, 1995; the vector integrates in an unpredicted manner and causes a mutation e.g. R124. In addition the observation of a phenotype will be dependent on the function of the gene and whether that function is unique or if it can be substituted by a redundant family member. Phenotypes may also be influenced by the action of unlinked genetic modifiers and this would be reflected in the expressivity or the penetrance of a defect. Thus it is important to show that the integration is genetically linked to the phenotype and so is the likely cause of the mutation.

The R124 gene-trap integration has been shown to be linked to the phenotypes observed in animals homozygous for the integration by RFLP analysis, which allowed distinction between wild type and trapped alleles. The neonatal phenotype generated by the R124 integration is not fully penetrant therefore is likely to be under the influence of an unlinked genetic modifier. In addition the adult phenotypes are male specific, as far as has been determined, and thus may be under the influence of sex-specific modifiers. To complete phenotype analysis it will be essential to classify the protein encoded by the trapped gene at the sequence level and determine its relationship with known genetic markers. Sequence data of the trapped gene can also be used to target the endogenous gene by homologous recombination to generate mutations if it has not been disrupted by the integration of the gene-trap vector.

A major disadvantage of gene-trap approaches is that in random only one ES cell clone for any gene-trap vector integration event is isolated in any screen. If this ES clone, as highlighted by the analysis of the R68 gene-trap cell line, has lost its pluripotency or is suboptimal the ability to transmit an integration event is compromised thus *in vivo*

analysis cannot be performed. Other consequence of working with a single clone that need to be considered include the presence of mutations, unlinked or linked to the gene-trap integration, within the gene-trap cell line that may influence phenotype analysis. An unlinked mutation can be segregated from the integration event through successive backcross generations to generate congenic strains. Analysis to show the integration is associated with the phenotype through successive generations benefits from a reliable genotyping protocol. This was done by identifying a RFLP to distinguish wild type and trapped alleles for the R124 gene-trap integration. Closely linked mutations are difficult to segregate and could include subtle rearrangements of the genome around the site of integration e.g. deletions or invertions. Another possibility is the integration has disrupted regulation elements that do not affect the expression of the trapped gene but of closely linked genes. To eliminate these possibilities one has to target the endogenous trapped gene using a conventional homologous recombination strategy to show the same phenotype. Conversely, one could propose "knock-in" experiments to rescue the phenotype using endogenous trapped gene sequences.

Thus in summary it is important to collect as much sequence data as one can for any gene-trap integration. This information allows one to determine the gene-trap integration event and the approach required to clone the trapped sequence. The function of the gene can be predicted from its sequence and combined with expression analysis the mutant phenotype, as a consequence of disrupting the endogenous trapped gene, can be anticipated and better understood.

The future characterisation of many more gene-trap integrations in detail will provide further information to asses the reliability of the predicted gene-trap technology and problems that can arise due to unconventional gene-trap integrations.

The sequence data gathered for the R124 gene-trap integration is not complete and will be necessary to identify the precise function of the endogenous trapped gene. However

the genotype and phenotype analysis suggests the R124 integration has disrupted a gene important for right ventricle heart development and function in the neonate. The incomplete penetrance of this phenotype and the detection of heart defects in the adult may reflect a progressive state of disease. The heart defect resembles right ventricle dysfunction found in humans and may be useful as a model system to investigate the parameters that affect this condition (Rockman *et al.*, 1994). In humans right ventricle dysfunction is a progressive disease that leads to congestive heart failure and is associated with ischemia or infarction of the ventricle wall which weakens the ventricle wall causing it to dilate (Goldstein *et al.*, 1998; Parfrey *et al.*, 1995). This leads to aberrant hemodynamics and electrical activity of the right ventricle (Goldstein *et al.*, 1990). Compensation of this defect results in hypertrophy of the ventricle walls. Re-modeling of the ventricle wall during hypertrophy is associated with changes in the morphology of cardiomyocytes, expression of genetic markers e.g. actin, atrial natriuretic factor, βMHC within cardiomyocytes and in the components of the extracellular matrix surrounding the cardiomyocytes e.g. fibronectin, collagen, integrin (Beltrami *et al.*, 1995; Terracio *et al.*, 1991). It will be important to study changes in parameters mentioned above in the defective hearts observed in animals homozygous for the R124 integration to correlate these to the human condition.

Hypertrophy of the heart is often associated with changes in renin-angiotensin system which regulates blood pressure. Each is required for the function of the other and any alteration results in an unbalance causing damage to the heart and kidney (Johnson *et al.*, 1996). Indications that blood pressure may be altered in R124 mutants has been shown by the dysmorphogenesis of kidneys, including glomerulosclerosis, and the alteration in urine osmolarity in these animals. Both these observations suggest renal dysfunction in salt secretion and fluid filtration which could be the cause or consequence of abnormal blood pressure. Glomerulosclerosis is a degenerative condition affecting the kidney which is characterised by increased levels of

extracellular matrix deposition, as associated with the hypertrophic heart (Peten *et al.*, 1994).

Thus the heart disease syndrome observed in surviving adult R124 mutants may be the combined result of a developmental heart defect, (detected as dilatation of the right ventricle of the heart in the neonate which can be compensated for by hypertrophy and may be influenced by genetic modifiers) and kidney defects which accentuate each other.

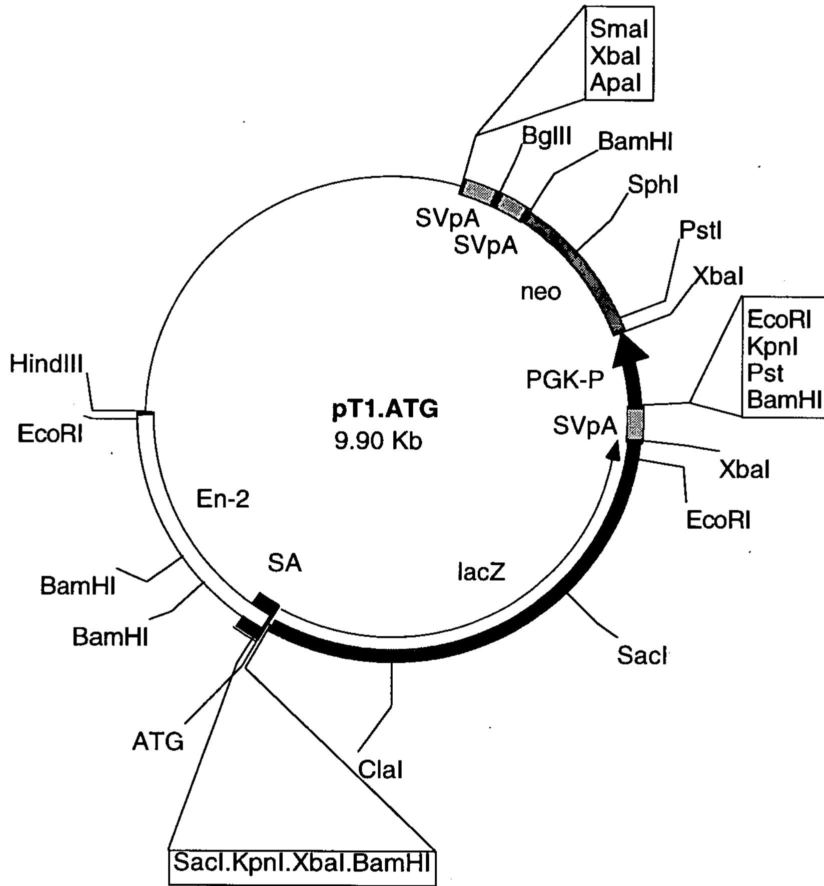
The use of R124 mutants as models to study heart disease will be validated further by showing that the gene-trapped by the R124 integration is conserved, by screening the human and other species genomes by Southern blot analysis, and if the human gene or locus is linked to a known cardiac syndrome. RFLP analysis indicates the endogenous trapped gene is a single copy gene and phenotype analysis indicates it has an important function thus it is likely this gene may be conserved.

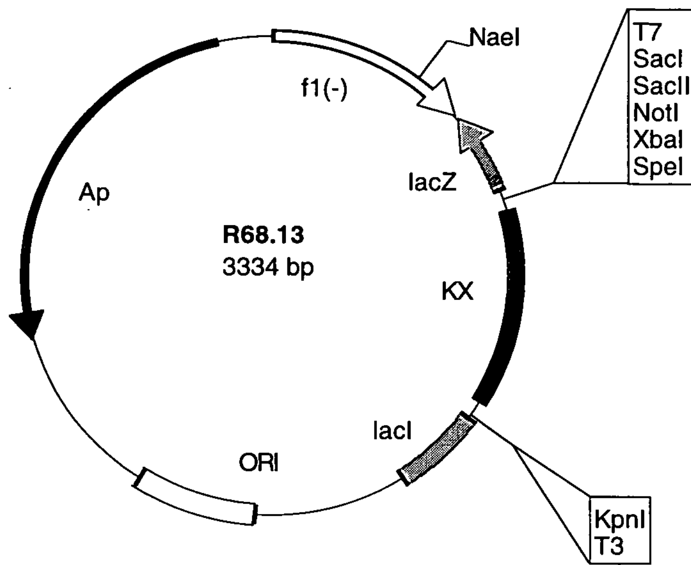
Similar reasoning can be used to determine the use of the male infertility caused by the R124 integration to be considered as a model for human male infertility. Infertility in human males due to arrest of sperm maturation has been documented and can be characterised with a number of genetic markers (Propst *et al.*, 1988).

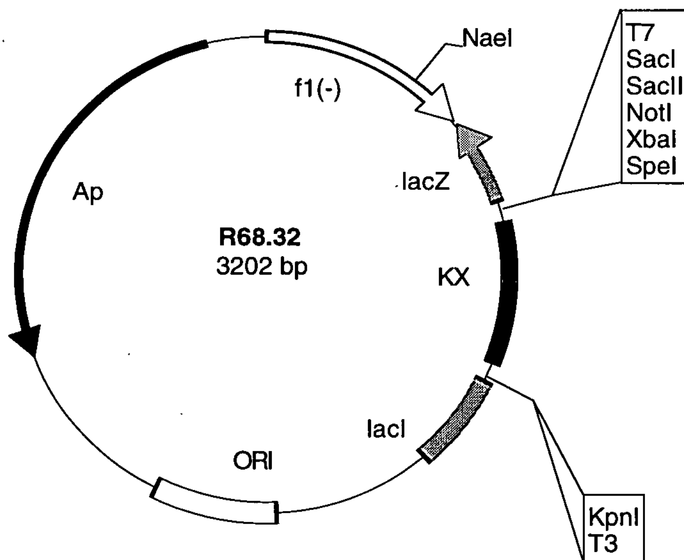
Part of this work has been presented for publication with Peter McClive as a joint author and is in press, citation: McClive, P., Pall, G., Newton, K., Lee, M., Mullins, J., and Forrester, L. (1998) Gene-trap integrations expressed in the developing heart: Insertion affects splicing of the PT1-ATG vector, *Developmental Dynamics*.

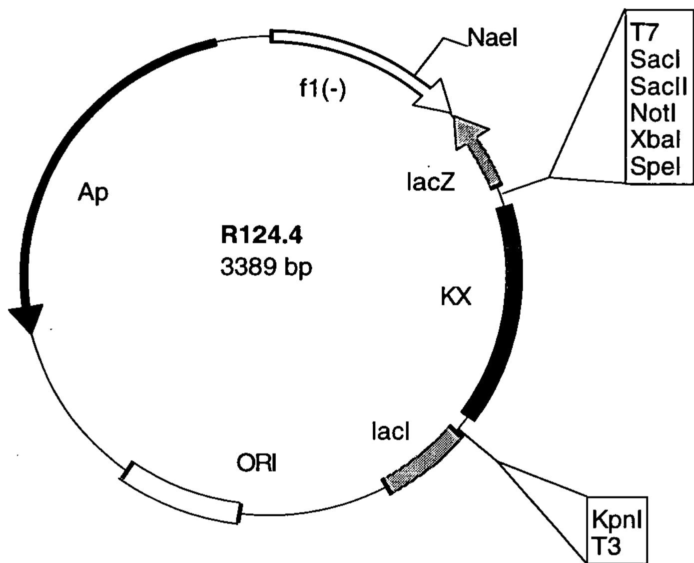
Appendix

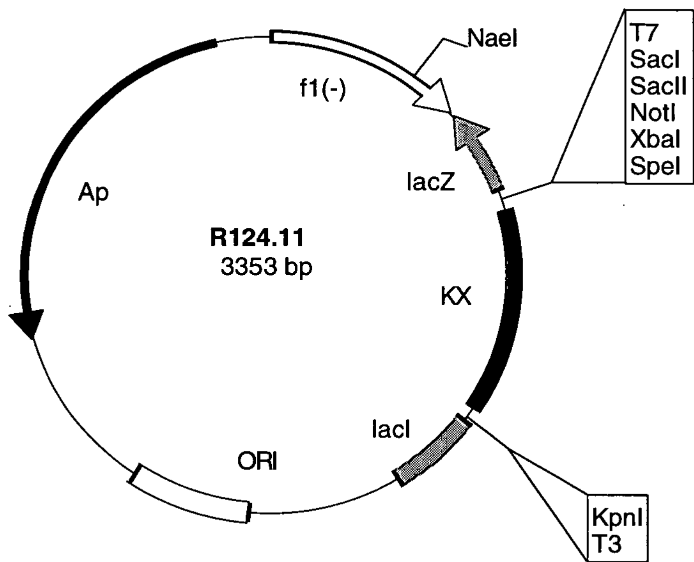
Appendix I: Maps of plasmid constructs used for molecular analysis of gene-trap integrations.

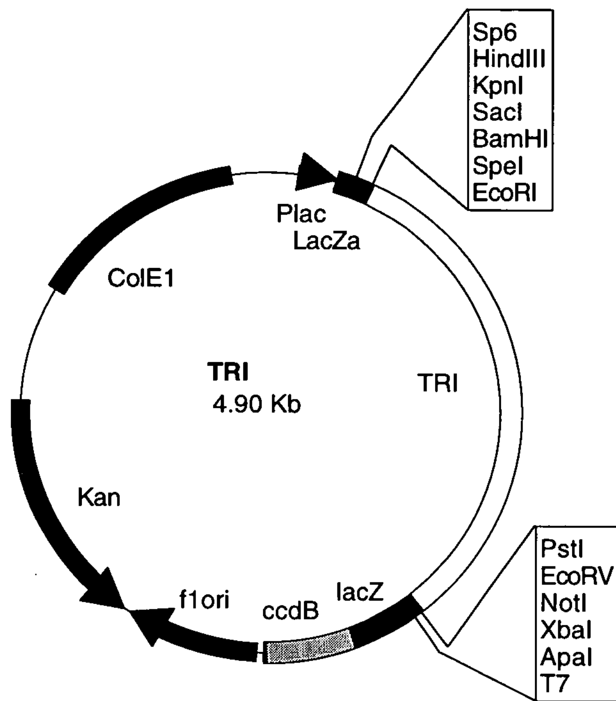


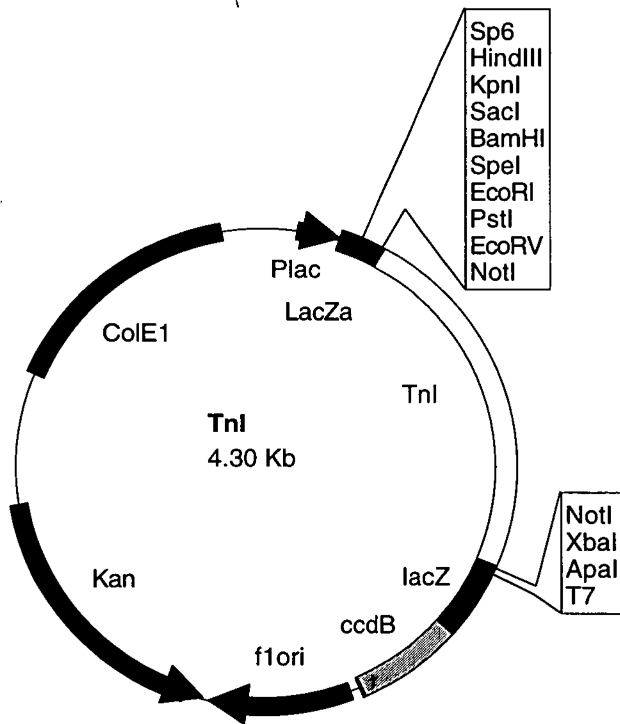

NaeI
T7
SacI
SacII
NotI
XbaI
SpeI
f1(-)
lacZ
Ap
R68.13
3334 bp
KX
lacI
ORI
KpnI
T3

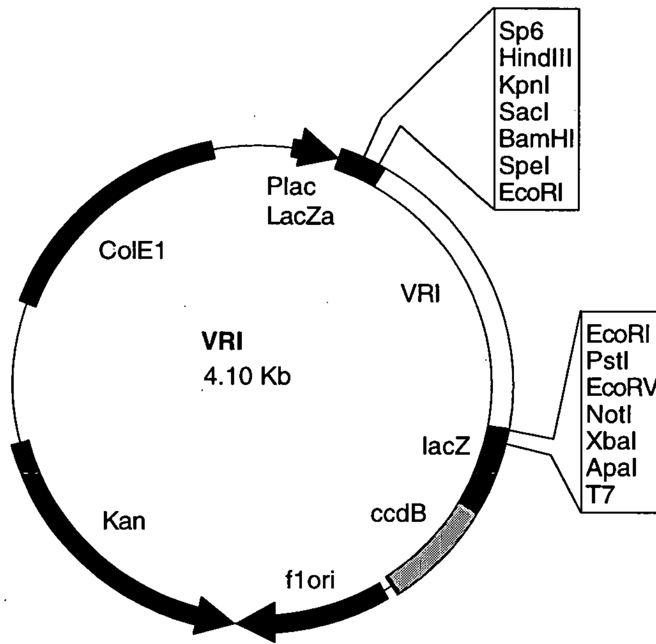

NaeI
T7
SacI
SacII
NotI
XbaI
SpeI
f1(-)
lacZ
Ap
R68.32
3202 bp
KX
lacI
ORI
KpnI
T3


NaeI
f1(-)
T7
SacI
SacII
NotI
XbaI
SpeI
lacZ
Ap
R124.4
3389 bp
KX
lacI
ORI
KpnI
T3


NaeI
f1(-)
T7
SacI
SacII
NotI
XbaI
SpeI
lacZ
Ap
R124.11
3353 bp
KX
lacI
ORI
KpnI
T3


Sp6
HindIII
KpnI
SacI
BamHI
SpeI
EcoRI
Plac
LacZa
ColE1
TRI
4.90 Kb
TRI
Kan
PstI
EcoRV
NotI
XbaI
ApaI
T7
lacZ
f1ori
ccdB


Sp6
HindIII
KpnI
SacI
BamHI
SpeI
EcoRI
PstI
EcoRV
NotI
Plac
LacZa
ColE1
TnI
TnI
4.30 Kb
Kan
NotI
XbaI
ApaI
T7
lacZ
ccdB
f1ori


Sp6
HindIII
KpnI
SacI
BamHI
SpeI
EcoRI
Plac
LacZa
ColE1
VRI
VRI
4.10 Kb
EcoRI
PstI
EcoRV
NotI
XbaI
ApaI
T7
lacZ
ccdB
Kan
f1ori

Appendix II: PT1.ATG gene-trap vector sequence

The PT1.ATG vector is based on the original PT1.4.5a-PGKneo gene-trap vector. Vector is linearised at the HindIII site before electroporation into ES cells. Listed below are sequences of the vector used in the molecular analysis of the gene-trap integrations, essentially 5' sequence to lacZ reporter gene, to design PCR primers, probe fragments and perform Southern blot analysis. The first nine bases of the vector, containing HindIII and EcoRI sites, have been lost, possibly by endogenous endonuclease activity before integration into the ES cell genome in the R124 gene-trap integration. cSD, cryptic splice donor; SJ, splice junction; intron sequences lower case; exon sequences upper case; primer sites underlined.

Primers	Sequence(5' - 3')
*a1(*en-2* intron)	ACTTGGCCTCACCAGGC
*a2(*en-2* exon)	TGCTCTGTCAGGTACCTGTTGG
*b2(*en-2* intron)	TGAGCACCAGAGGACATCCG
#1(*lacZ*)	GCAAGGCGATTAAGTTGGGT
#3(*en-2* exon)	CCGTCGACTCTGGCGCCGCT
#4(*en-2* exon)	TGCTCTGTCAGGTACCTGTTG

*primers used for genomic PCRs

#primers used for 5' RACE

```
H
i
n       E
d       c   E
I       o   a
I       R   r                cSD32
I       I   I                  ↓
    aagcttggaattcatgggaagaggaaccgaaagtatgtttttcagatgttctttctcaga
  1 ---------+---------+---------+---------+---------+---------+
    ttcgaaccttaagtacccttctccttggctttcatacaaaaagtctacaagaaagagtct

    aataggagtttgcggaggttggagtgtgtgttgtaggacacgaaccccagggtggaggag
 61 ---------+---------+---------+---------+---------+---------+
    ttatcctcaaacgcctccaacctcacacacaacatcctgtgcttggggtcccacctcctc

    actggaggacagagccctctttcccagggagggaaggaggagagtttgagatccgctccg
121 ---------+---------+---------+---------+---------+---------+
    tgacctcctgtctcgggagaaagggtccctcccttcctcctctcaaactctaggcgaggc

    gaagtcggggttcaggtttgagcaggccaggcctctcccgtggtctcgccctcttgtcct
181 ---------+---------+---------+---------+---------+---------+
    cttcagccccaagtccaaactcgtccggtccggagagggcaccagagcgggagaacagga

    agaagcctcactggccaggtgtaagccaggtcgtgggtgccgagccctgctccctcatcc
241 ---------+---------+---------+---------+---------+---------+
    tcttcggagtgaccggtccacattcggtccagcacccacggctcgggacgagggagtagg

    tcagcatggatgtgaagaggactgtatggcgtgcgggtgtgtgtgaccgtgggtacactt
301 ---------+---------+---------+---------+---------+---------+
    agtcgtacctacacttctcctgacataccgcacgcccacacacactggcacccatgtgaa
                                                 primer b2
    aaaacaccgggttttggatctgcactgtcccggatgtcctctggtgctcaaagaccctttt
361 ---------+---------+---------+&lt;--------+---------+---------+
    ttttgtggcccaaaacctagacgtgacagggcctacaggagaccacgagtttctgggaaa

              cSD435
                ↓
    tgggtttgccctttggtaagagcgccgggatctacttgtctggaggccagggagtcctca
421 ---------+---------+---------+---------+---------+---------+
    acccaaacgggaaaccattctcgcggccctagatgaacagacctccggtccctcaggagt
```

```
                                                               K
                                                               p
                                                               n
                                                               I
    gccgaggcttgccgcccctgactgcactgcactgagtagtggatgggagagtctggtacc
481 ---------+---------+---------+---------+---------+---------+
    cggctccgaacggcggggactgacgtgacgtgactcatcacctaccctctcagaccatgg

    gcactgccggtttcctccaccatccccgcagcgcagggcagtgcattccgtcctggctgc
541 ---------+---------+---------+---------+---------+---------+
    cgtgacggccaaaggaggtggtaggggcgtcgcgtcccgtcacgtaaggcaggaccgacg

    gaaggggatggtcgggccttctccagcctcttccgcttctagcgaaggggccttgatgg
601 ---------+---------+---------+---------+---------+---------+
    cttcccctaccagcccggaagaggtcggagaaggcgaagatcgcttccccggaactacc

    aagggcccgcatgtctccaaagttgattcatgcttcttgcacagagaaagaccagaaaga
661 ---------+---------+---------+---------+---------+---------+
    ttcccgggcgtacagaggtttcaactaagtacgaagaacgtgtctctttctggtctttct

    aggtctcaagttttagccggtagcccggatggccttttcctgcacggcaccatatgaacc
721 ---------+---------+---------+---------+---------+---------+
    tccagagttcaaaatcggccatcgggcctaccggaaaaggacgtgccgtggtatacttgg

    ttgtgaccctgactttgagacccctctaacccaaggcccctaccactttaccctttccct
781 ---------+---------+---------+---------+---------+---------+
    aacactgggactgaaactctggggagattgggttccggggatggtgaaatgggaaaggga

    ttgaaggctttcccacaccaccctccacacttnccccaaacactgccaactatgtaggag
841 ---------+---------+---------+---------+---------+---------+
    aacttccgaaagggtgtggtgggaggtgtgaanggggtttgtgacggttgatacatcctc

    gaaggggttgggactaacagaagaacccgttgtggggaagctgttgggagggtcacttta
901 ---------+---------+---------+---------+---------+---------+
    cttccccaaccctgattgtcttcttgggcaacaccccttcgacaaccctcccagtgaaat

    tgttcttgcccaaggtcagttgggtggcctgcttctgatgaggtggtcccaaggtctggg
961 ---------+---------+---------+---------+---------+---------+
    acaagaacgggttccagtcaacccaccggacgaagactactccaccagggttccagaccc
```

```
     gtagaaggtgagagggacaggccaccaaggtcagcccccccccctatcccataggagcc
1021 ---------+---------+---------+---------+---------+---------+
     catcttccactctccctgtccggtggttccagtcggggggggggatagggtatcctcgg

     aggtccctctcctggacaggaagactgaaggggagatgccagagactcagtgaagcctgg
1081 ---------+---------+---------+---------+---------+---------+
     tccagggagaggacctgtccttctgacttcccctctacggtctctgagtcacttcggacc

         K
         p
         n
         I                                    primer a1
     ggtaccctattggagtccttcaaggaaacaaacttggcctcaccaggcctcagccttggc
1141 ---------+---------+---------+---------+---------+---------+
     ccatgggataacctcaggaagttcctttgtttgaaccggagtggtccggagtcggaaccg

                                  B
                                  a
                                  m
                                  H
                                  I
     tcctcctgggaactctactgcccttgggatcccttgtagttgtgggttacataggaaggg
1201 ---------+---------+---------+---------+---------+---------+
     aggaggacccttgagatgacgggaaccctagggaacatcaacacccaatgtatccttccc

     gacggattccccttgactggctagcctactcttttcttcagtcttctccatctcctctca
1261 ---------+---------+---------+---------+---------+---------+
     ctgcctaaggggaactgaccgatcggatgagaaaagaagtcagaagaggtagaggagagt

             T
             a
             q
             I
     ccgttctctcgaccctttccctaggatagacttggaaaaagataaggggagaaaaacaaa
1321 ---------+---------+---------+---------+---------+---------+
     ggcaagagagctgggaaagggatcctatctgaacctttttctattcccctctttttgttt

                                                            B
                                                            a
                                                            m
                                                            H
                                                            I
     tgcaaacgaggccagaaagatttggctgggcattccttccgctagcttttattgggatc
1381 ---------+---------+---------+---------+---------+---------+
     acgtttgctccggtctttctaaaccgacccgtaaggaaggcgatcgaaaataaccctag
```

```
     ccctagtttgtgataggccttttagctacatctgccaatccatctcattttcacacacac
1441 ---------+---------+---------+---------+---------+---------+
     gggatcaaacactatccggaaaatcgatgtagacggttaggtagagtaaaagtgtgtgtg

     acacaccactttccttctggtcagtgggcacatgtccagcccccaacacttgtatggcct
1501 ---------+---------+---------+---------+---------+---------+
     tgtgtggtgaaaggaagaccagtcacccgtgtacaggtcgggggttgtgaacataccgga

     tggcggggtcatcccccccccacccccagtatctgcaacctcaagctagcttgggtgcgt
1561 ---------+---------+---------+---------+---------+---------+
     accgccccagtagggggggggtgggggtcatagacgttggagttcgatcgaacccacgca

     tggttgtggataagtagctagactccagcaaccagtaacctctgccctttctcctccatg
1621 ---------+---------+---------+---------+---------+---------+
     accaacacctattcatcgatctgaggtcgttggtcattggagacgggaaagaggaggtac
```

en-2 intron ← SJ → *en-2* exon

```
     acaaccagGTCCCAGGTCCCGAAAACCAAAGAAGAAGAACCCTAACAAAGAGGACAAGCG
1681 ---------+---------+---------+---------+---------+---------+
     tgttggtcCAGGGTCCAGGGCTTTTGGTTTCTTCTTCTTGGGATTGTTTCTCCTGTTCGC

     GCCTCGCACAGCCTTCACTGCTGAGCAGCTCCAGAGGCTCAAGGCTGAGTTTCAGACCAA
1741 ---------+---------+---------+---------+---------+---------+
     CGGAGCGTGTCGGAAGTGACGACTCGTCGAGGTCTCCGAGTTCCGACTCAAAGTCTGGTT

           K                                            K
           p                                            p
           n                                            n
           I  primer 4 = a2     primer 3                I
     CAGGTACCTGACAGAGCAGCGGCGCCAGAGTCTGGCACAGGAGCTCGGTACCCGGATGAT
1801 ---------+---------+---------+---------+---------+---------+
     GTCCATGGACTGTCTCGTCGCCGCGGTCTCAGACCGTGTCCTCGAGCCATGGGCCTACTA

                  B
           X      a
           b      m
           a      H
           I      I
     CTGGACTCTAGAGGATCCCGTCGTTTTACAACGTCGTGACTGGGAAAACCCTGGCGTTAC
1861 ---------+---------+---------+---------+---------+---------+
     GACCTGAGATCTCCTAGGGCAGCAAAATGTTGCAGCACTGACCCTTTTGGGACCGCAATG
```

en-2 exon ← → *lacZ*

```
                                                      P
                                                      v
                                                      u
                                                      I
     primer 1                                         I
     CCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATAGCGAAGAGGC
1921 ---------+---------+---------+---------+---------+---------+
     GGTTGAATTAGCGGAACGTCGTGTAGGGGGAAAGCGGTCGACCGCATTATCGCTTCTCCG

     CCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTGAATGGCGAATGGCGCTTTGCCTG
1981 ---------+---------+---------+---------+---------+---------+
     GGCGTGGCTAGCGGGAAGGGTTGTCAACGCGTCGGACTTACCGCTTACCGCGAAACGGAC

     GTTTCCGGCACCAGAAGCGGTGCCGGAAAGCTGGCTGGAGTGCGATCTTCCTGAGGCCGA
2041 ---------+---------+---------+---------+---------+---------+
     CAAAGGCCGTGGTCTTCGCCACGGCCTTTCGACCGACCTCACGCTAGAAGGACTCCGGCT

     TACTGTCGTCGTCCCCTCAAACTGGCAGATGCACGGTTACGATGCGCCCATCTACACCAA
2101 ---------+---------+---------+---------+---------+---------+
     ATGACAGCAGCAGGGGAGTTTGACCGTCTACGTGCCAATGCTACGCGGGTAGATGTGGTT

     CGTAACCTATCCCATTACGGTCAATCCGCCGTTTGTTCCCACG
2161 ---------+---------+---------+---------+---> remainder
     GCATTGGATAGGGTAATGCCAGTTAGGCGGCAAACAAGGGTGC     of gene-trap
                                                     vector
```

Abbreviations
List of Figures
List of Tables

Abbreviations

ATG	translational start codon
β–gal	β–galactosidase
β–geo	β–galactosidase and neomycin fusion
bp	basepairs
BSA	bovine serum albumin
cDNA	complementary deoxyribonucleic acid
CHL	chlorambucil
cpm	counts per minute
cSD	cryptic splice donor
dATP	deoxyadenosine triphosphate
dCTP	deoxycytidine triphosphate
dGTP	deoxyguanosine triphosphate
dNTP	deoxynucleoside triphosphate
dTTP	deoxythymidine triphosphate
DEPC	diethylpyrocarbonate
DIA	differentiation inhibiting activity
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic Acid
d.p.c.	days*post coitum*
d.p.p.	days *post partum*
DTT	dithiothreitol
EB	embryoid body
EDTA	ethylenediaminetetraacetic acid
en-2	mouse engrailed-2
ENU	ethylitrosourea
ES cells	embryonic stem cells
EST	expressed sequence tag
FISH	fluorescent *in situ* hybridisation
IRES	internal ribosomal entry site
kan	kanamycin
kb	kilobases
LIF	leukemia inhibiting factor
MHC	myosin heavy chain
MLC	myosin light chain
mRNA	messenger ribonucleic acid

OMIM	online Mendelian inheritance in man
ORF	open reading frame
p.c.	*post coitum*
PCR	polymerase chain reaction
PGK	phophoglycerate kinase
p.p.	*post partium*
RA	retinoic acid
RAR	retinoic acid receptor
RARE	retinoic acid receptor element
RACE	rapid amplification of cDNA ends
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
RT-PCR	reverse transcription polymerase chain reaction
RXR	retinoid X receptor
SA	splice acceptor
SD	splice donor
TESPA	3-aminopropyltriethoxysilane
UTR	untranslated region
VAD	vitamin A deficiency

List of Figures

Figure 1.1: Mammalian Heart Development

Figure 1.2: A Genetic Pathway Determining Asymmetry of the Heart

Figure 1.3: Characteristic Normal Cardiac Cycle

Figure 3.1: *In vitro* Differentiation of ES cell Derived EBs

Figure 3.2: R68 ES cell derived Beating Cardiomyocytes

Figure 3.3: R124 ES cell derived Beating Cardiomyocytesg

Figure 3.4: Reporter Gene Expression during Embryogenesis

Figure 3.5: Reporter Gene Expression in Adult Tissues

Figure 3.6: Chromosomal Mapping

Figure 3.7: Detection of Fusion Transcripts

Figure 3.8: Predicted Splice Events from RACE Products

Figure 3.9: RNase Protection with RACE Clones

Figure 4.1: R124 RACE clone 11

Figure 4.2: PCR Analysis of the R124 Genomic Integration Site

Figure 4.3: cDNA Clones Isolated from a Mouse Heart cDNA Library Screen

Figure 4.4: Contig of RACE and cDNA Sequences

Figure 4.5: Expression of Endogenous Transcripts

Figure 4.6: Southern Blot Analysis of Genomic DNA

Figure 4.7: RFLP Analysis of the R124 Integration

Figure 4.8: Linkage of Endogenous Sequences to the Vector at the Genomic Level

Figure 4.9: Predicted Structure of the R124 Gene-Trap Integration

Figure 5.1: Neonatal Phenotype

Figure 5.2: Enlarged Hearts

Figure 5.3: Kidney Dysmorphogenesis

Figure 5.4: Male Infertility

List of Tables

Table 1.1: Genes Implicated in Mouse Heart Development

Table 1.2A: Characterised Gene-Trap Integrations

Table 1.2B: Large Scale Gene-Trap Screens

Table 2.1: Plasmid Constructs used for Molecular Analysis

Table 3.1: β-gal Response to RA-exposure

Table 3.2: Quantitation of Fusion Transcripts

Table 3.3: RACE Products from R68 and R124 Gene-Trap cell lines

Table 4.1: Database Search Results

Table 5.1: Genotyping of R124 Intercross Litters

Table 5.2: Organ and Body Weights

Table 5.3: Metabolic Measurements

References

References

Adams, M. D., Kerlavage, A. R., Fleischmann, R. D., Fuldner, R. A., Bult, C. J., Lee, N. H., Kirkness, E. F., Weinstock, K. G., Gocayne, J. D., White, O., and al. (1995). Initial assessment of human gene diversity and expression patterns based upon 83million nucleotides of cDNA sequence. *Nature* **377,** 3-174.

Alam J. and Cook JL. (1990) Reporter genes: application to the study of mammalian gene transcription. *Anal Biochem* **188(2),** 245-254.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215,** 403-410.

Baker, R. K., Haendel, M. A., Swanson, B. J., Shambaugh, J. C., Micales, B. K., and Lyons, G. E. (1997). *In vitro* preselection of gene-trapped embryonic stem cell clones for characterizing novel developmentally regulated genes in the mouse. *Dev Biol* **185,** 201-214.

Beeson, W. T. I., Kirby, M. L., van Mierop, L. H. S., and Teabeaut, J. R. I. (1986). Effects of the size of lesions of the cardiac neural crest at various embryonic ages on incidence and types of cardiac defects. *Circ.* **73,** 360-364.

Bellen, H. J., O'Kane, C. J., Wilson, C., Grossniklaus, U., Pearson, R. K., and Gehring, W. J. (1989). P-element-mediated enhancer detection : a versatile method to study development in Drosophila. *Genes & Dev* **3,** 1288-1300.

Bellofatto, V., Shapiro, L., and Hodgson, D. A. (1984). Generation of a Tn5 promoter probe and its use in the study of gene expression in *Caulobacter crescentus. Proc Natl Acad Sci* **81,** 1035-1039.

Beltrami, C. A., Finato, N., Rocco, M., Feruglio, G. A., Puricelli, C., Cigola, E., Sonnenblick, E. H., Olivetti, G., and Anversa, P. (1995). The cellular basis of dilated cardiomyopathy in humans. *J. Mol. Cell Cardiol.* **27,** 291-305.

Biben, C., and Harvey, R. P. (1997). Homeodomain factor Nkx2.5 controls left/right asymmetric expression of bHLH geen *eHAND* during murine heart development. *Genes and Development* **11,** 1357-1369.

Bier, E., Jan, L. Y., and Jan, Y. N. (1990). *rhomboid,* a gene required for dorsoventral axis establishment and peripheral nervous system development in *Drosophila melanogaster. Genes and Dev.* **4,** 190-203.

Bishop, S. P. (1995). Regulation of cardiac myocyte differentiation, hyperplasia, and hypertrophy during normal and stressed fetal and neonatal growth. *In* “Heart Hypertrophy and Failure” (N. S. Dhalla, G. N. Pierce, V. Panagia, and R. E. Beamish, Eds.), pp. 183-196. Kluwer Academic Publishers, Boston/Dordrecht/London.

Blendy, J. A., Kaestner, K.H., Weinbauer, G.F., Nieschlag, E. and Schutz, G. (1996). Severe impairment of spermatogenesis in mice lacking the CREM gene. *Nature* **380,** 162-165.

Bodmer, R. (1993). The gene *tinman* is required for specification of the heart and visceral muscles in *Drosophila. Development* **118,** 719-729.

Boguski, M. S., Lowe, T. M., and C.M., T. (1993). dbEST--database for "expressed sequence tags". *Nature Genetics* **4,** 332-333.

Bonyadi, M., Rusholme, S. A. B., Cousins, F. M., Su, H. C., Biron, C. A., Farrall, M., and Akhurst, R. J. (1997). Mapping of a major genetic modifier of embryonic lethality in TGFβ1 knockout mice. *Nature Genetics* **15,** 207-211.

Bradley, A., Evans, M., Kaufmann, M. H., and Robertson, E. J. (1984). Formation of germ-line chimeras from embryo derived teratocarcinoma cell lines. *Nature* **309,** 255-256.

Brannan, C. I., Perkins, A. S., Vogel, K. S., Ratner, N., Nordlund, M. L., Reid, S. W., Buchberg, A. M., Jenkins, N. A., Parada, L. F., and Copeland, N. G. (1994). Targeted disruption of the neurofibromatosis type-1 geneleads to developmental abnormalities in heart and various neural crest-derived tissues. *Genes and Development* **8,** 1019-1029.

Brodsky, W. Y., and Arefyeva, A. M. (1980). Mitotic polyploidization of mouse heart myocytes during the first postnatal week. *Cell Tissue Research* **210,** 133-144.

Brook, F. A., and Gardner, R. L. (1997). The origin and efficient derivation of embryonic stem cells in the mouse. *Proc.Natl.Acad.Sci. USA* **94,** 5709-5712.

Burns, K. D., Homma, T. and Harris, R.C. (1993). The intrarenal renin-angiotensin system. *Semin. Nephrol.* **13,** 13-30.

Capecchi, M. R. (1989). Altering the genome by homologous recombination. *Science* **244,** 1288-1292.

Casabadan, M. J., Chou, J., and Cohen, S. N. (1980). *In vitro* gene fusion that joins an enzymatically active bgal segment to amino terminal fragments of exogenous proteins: *Escherichia coli* plasmid vectorss for the detection and cloning of translational initiation signals. *J Bacteriol* **143,** 971-980.

Chamberlain, J. S., Grant, S. G., Reeves, A. A., Mullins, L. J., Stephenson, D. A., Hoffman, E. P., Monaco, A. P., Kunkel, L. M., Caskey, C. T., and Chapman, V. M. (1987). Regional localization of the murine Duchennemuscular dystrophy gene on the mouse X chromosome. *Somat. Cell Molec. Genet.* **13,** 671-678.

Chambon, P. (1994). The retinoid signaling pathway: molecular and genetic analyses. *Sem. in Cell Biol.* **5,** 115-125.

Chen, Z., Friedrich, G. A., and Soriano, P. (1994). Transcriptional enhancer factor -1 disruption by a retroviral gene trap leads to heart- defects and embryonic lethality in mice. *Gen Dev* **8,** 2293-2301.

Chen, J. N., Haffter, P., Odenthal, J., Vogelsang, E., Brand, M., vanEeden, F. J. M., FurutaniSeiki, M., Granato, M., Hammerschmidt, M., Heisenberg, C. P., Jiang, Y. J., Kane, D. A., Kelsh, R. N., Mullins, M. C., and NussleinVolhard, C. (1996). Mutations affecting the cardiovascular system and other internal organs in zebrafish. *Development* **123,** 293-302.

Chen, Y.-F. (1996). Sexual dimorphism of hypertension. *Current Opinion in Nephrology and Hypertension* **6,** 181-185.

Chen, J., Nachabah, A., Scherer, C., Ganju, P., Reith, A., Bronson, R., and Ruley, H. E. (1996). Germ-line inactivation of the murine Eck receptor kinase by gene-trap retroviral insertion. *Oncogene* **12,** 979-988.

Chien, K. R. (1993). Molecular advances in cardiovascular biology. *Science* **260,** 916-917.

Chisaka, O., and Capecchi, M. R. (1991). Regionally restricted developmental defects resulting from targted disruption of the mouse homeobox gene *hox-1.5*. *Nature* **350,** 473-479.

Chomczynski, P., and Sacchi, N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162,** 156-159.

Chowdhury, K., Bonaldo, P., Torres, M., Stoykova, A., and Gruss, P. (1997). Evidence for the stochastic integration of gene trap vectors into the mouse germline. *Nuc Acids Res* **25,** 1531-1536.

Coceani, F., and Olley, P. M. (1988). The control of cardiovascular shunts in the fetal and perinatal period. *Can. J. Physiol. Pharmacol.* **66,** 1129-1134.

Collignon, J., Varlet, I., and Robertson, E. J. (1996). Relationship between asymmetric nodal expression and the direction of embryonic turning. *Nature* **381,** 155-158.

Collins, F. S. (1995). Positional cloning moves from perditional to traditional. *Nature Genetics* **9,** 347-350.

Colucci, W. S. (1994). "Atlas of Heart Diseases, Heart Failure: Cardiac Function and Dysfunction." Mosby,

Conlon, R. A. (1995). Retinoic acid and pattern formation in vertebrates. *TIG* **11,** 314-319.

Courtney, K. D., and Ebron, M. T. (1978). Isozyme profiles of lactic dehydrogenase and creatine phosphokinase in neonatal mouse hearts. *Biol. Neonate* **34,** 203-208.

Cross, J. C. (1995). Hxt encodes a basic helix-loop-helix transcription factor that regulates trophoblast cell development. *Development* **121,** 2513-2523.

Cserjesi, P., Brown, D., Lyons, G.E. and Olson, E.N. (1995). Expression of the novel basic helix-loop-helix gene eHAND in neural crest derivatives and extraembryonic membranes during mouse development. *Developmental Biology* **170,** 664-678.

Cui, C., Wani, M. A., Wight, D., Kopchick, J., and Stambrook, P. J. (1994). Reporter genes in transgenic mice. *Transgenic Research* **3,** 182-194.

Danos, M. C., and Yost, H. J. (1995). Linkage of cardiac left-right asymmetry and dorsal-anterior development in *Xenopus. Development* **121,** 1467-1474.

DeHaan, R. L. (1965). Morphogenesis of the vertebrate heart. *In* "Organogenesis" (R. L. DeHaan and H. Ursprung, Eds.), pp. 377-419. Holt, Rinehart and Winston Inc., New York.

DeGregori, T., Russ, A., Vonmelchner, H., Rayburn, H., Priyaranjan, P., Jenkins, N. A., Copeland, N. G., Ruley, H. E. (1994) A murine homolog of the yeast rna1 gene is required for postimplantation development. *Genes & Development* **8**, 265-276.

Deng, J. M., and Behringer, R. R. (1995). An insertional mutation in the BTF3 transcription factor gene leads to an early postimplantation lethality in mice. *Trans Res* **4,** 264-269.

Dickman, E. D., Thaller, C. and Smith, S.M. (1997). Temporally-regulated retinoic acid depletion produces specific neural crest, ocular and nervous system defects. *Development* **124,** 3111-3121.

Dietrich, W. F., Copeland, N. G., Gilbert, D. J., Miller, J. C., Jenkins, N. A., and Lander, E. S. (1995). Mapping the Mouse Genome: Current status and future prospects. *Proc Natl Acad Sci* **92,** 10849-10853.

Dix, D. J., Allen, J.W., Collins, B.W., Poorman-Allen, P., Mori, C., Blizard, D.R., Brown, P.R., Goulding, E.H., Strong, B.D. and Eddy, E.M. (1997). HSP70-2 is required for desynapsis of synaptonemal complexes during meiotic prophase in juvenile and adult mouse spermatocytes. *Development* **124,** 4595-4603.

Doe, C. Q., Chu-Lagraff, Q., Wright, D. M., and Scott, M. P. (1991). The *prospero* gene specifies cell fates in the *Drosophila* central nervous system. *Cell* **65,** 451-464.

Doetschman, T. C., Eistetter, H., Katz, M., Schmidt, W., and Keller, R. (1985). The in vitro development of blastocyst-derived embryonic stem cell lines: Formation of visceral yolk sac, blood islands and myocardium. *J.Embryol.Exp.Morph.* **87,** 27-45.

Doetschman, T., Shull, M., Kier, A., and Coffin, J. D. (1993). Embryonic stem cell model systems for vascular morphogenesis and cardiac disorders. *Hypertension* **22,** 618-629.

Driever, W., SolnicaKrezel, L., Schier, A. F., Neuhauss, S., Malicki, J., Stemple, D. L., Stainier, D., Zwartkruis, F., Abdelilah, S., Rangini, Z., Belak, J., and Boggs, C. (1996). A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* **123,** 37-46.

Dyson, E., Sucov, H. M., Kubalak, S. W., Schmid-Schonbein, G. W., DeLano, F. A., Evans, R. M., Ross Jr., J., and Chein, K. R. (1995). Atrial-like phenotype is associated with embryonic ventricular failure in retinoid X receptor α -/- mice. *Proceedings of the National Academy of Science, USA* **92,** 7386-7390.

Echelard, Y., Epstein, D. J., St-Jacques, B., Shen, L., Mohler, J., McMahon, J. A., and McMahon, A. P. (1993). Sonc hedgehog, a member of a family of putative signaling molecules, is implicated in the regulation of the CNS polarity. *Cell* **75,** 1417-1430.

Edmondson, D. G., Lyons, G. E., Martin, J. F., and Olson, E. N. (1994). Mef2 gene expression marks the cardiac and skeletal muscle lineages during mouse embryogenesis. *Development* **120,** 1251-63.

Engels, W. R. (1989). P elements in Drosophila melanogaster. *In* "Mobile DNA" (E. M. Berg and M. Howe, M, Eds.), pp. 437-484. American Society for Microbiology, Washington D.C.

Evans, M. J., and Kaufman, M. H. (1981). Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292,** 154-156.

Evans, R. (1988). The steroid and thyroid hormone receptor superfamily. *Science* **13,** 889-895.

Fasano, L., Roder, L., Core, N., Alexandre, E., Vola, C., Jacq, B., and Kerridge, S. (1991). The gene *teashirt* is required for the development of *Drosophila* embryonic trunk segments and encodes a protein with widely spaced zinc finger motifs. *Cell* **64,** 63-79.

Fassler, R., and Meyer, M. (1995). Consequences of lack of β1 integrin gene expression in mice. *Genes and Development* **9,** 1896-1908.

Fassler, R., Rohwedel, J., Maltsev, V., Bloch, W., Lentini, S., Guan, K., Gulberg, D., Hescheler, J., Addicks, K., and Wobus, A. M. (1996). Differentiation and integrity of cardiac muscle cells are impaired in the absence of β1 integrin. *Journal of Cell Science* **109,** 2989-2999.

Fawcett, D., Pasceri, P., Fraser, R., Colbert, M., Rossant, J., and Giguere, V. (1995). Postaxial polydactyly in forelimbs of CRABP-II mutant mice. *Development* **121,** 671-679.

Fewell, J. G., Osinska, H., Klevitsky, R., NG, W., Sfyris, G., Bahrehmand, F., and Robbins, J. (1997). A treadmill exercise regimen for identifying cardiovascular phenotypes in transgenic mice. *Am. J. Physiology* **273,** 1595-1605.

Field, L. J. (1993). Transgenic mice in cardiovascular research. *Annu. Rev. Physiol.* **55,** 97-114.

Fishman, M. C., and Chein, K. R. (1997). Fashioning the vertebrate heart: earliest embryonic decisions. *Development* **124,** 2099-2177.

Flucher, B. E., and Franzini-Armstrong, C. (1996). Formation of junctions involved in excitation-contraction coupling in skeletal and cardiac muscle. *Proceedings of the National Academy of Science, USA* **93,** 8101-8106.

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

Franco, D., Kelly, R., Lamers, W. H., Buckingham, M., and Moorman, A. F. M. (1997). Regionalized transcriptional domains of myosin light chain 3f transgenes in the embryonic mouse heart: Morphogenetic implications. *Developmental Biology* **188,** 17-33.

Franz T. (1989) Persistent truncus arteriosus in the Splotch mutant mouse. *Anat Embryol (Berl)* **180(5)**, 457-464.

Friedrich, G., and Soriano, P. (1991). Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes & Dev* **5,** 1513-1523.

Friedrich, G. and Soriano, P. (1993). Insertional Mutagenesis by Retroviruses and Promoter Traps in Embryonic Stem Cells. *In* “Methods in Enzymology”, Vol. 225. Academic Press.

Friedrich, G. A., Hildebrand, J. D., and Soriano, P. (1997). The secretory protein Sec8 is required for paraxial mesoderm formation in the mouse. *Dev. Biol.* **192,** 364-374.

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

Fukiishi, Y., and Morriss-Kay, G. M. (1992). Migration of cranial neural crest cells to the pharyngeal arches and heart in rat embryos. *Cell.Tiss.Res.* **268,** 1-8.

Fung, Y.-W., and Liew, C.-C. (1996). Identification of genes associated with myocardial development. *J. Mol. Cell Cardiol.* **28,** 1241-1249.

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

George, E. L., Georges-Labouesse, E. N., Patel-King, R. S., Rayburn, H., and Hynes, R. O. (1993). Defects in mesoderm, neural tube and vascular development in mouse embryos lacking fibronectin. *Development* **119,** 1079-1091.

Gerhart, J., and Keller, R. (1986). Region-specific cell activities in amphibian gastrulation. *Ann.Rev.Cell Biol.* **2,** 201-229.

Gogos, J. A., Thompson, R., Lowry, W., Sloane, B. F., Weintraub, H. and Horwitz, M. Gene trapping in differentiating cell-lines - regulation of the lysosomal protease cathepsin-b in skeletal myoblast growth and fusion. *Journal Of Cell Biology* **134,** 837-847.

Goldstein, J. A., Barzilai, B., Rosamond, T. L., Eisenberg, P. R., and Jaffe, A. S. (1990). Determinants of hemodynamic compromise with severe right ventricular infarction. *Circulation* **82,** 359-368.

Goldstein, J. A. (1998). Right heart ischemia: pathophysiology, natural history, and clinical management. *Prog. Cardiovasc. Dis.* **40,** 325-341.

Gomez, R. A., Pupilli, C. and Everett, A.D. (1991). Molecular and cellular aspects of renin during kidney ontogeny. *Pediatr. Nephrol.* **5,** 80-87.

Gomez, R. A. a. N., V.F. (1995). Developmental Consequences of the Renin-Angiotensin System. *American Journal of Kidney Diseases* **26,** 409-431.

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

Gossler, A. a. Z., J. (1993). Gene and enhancer trap screens in ES cell chimeras. *In* "Gene Targeting" (A. Joyner, Ed.), pp. 181-227. IRL Press, Oxford.

Gould, S. J., Subramani, S., and Scheffler, I. E. (1989). Use of the DNA polymerase chain reaction for homology probing: isolation of partial cDNA or genomic clones encoding the iron-sulfur protein of succinate dehydrogenase from several species. *Proc. Nat. Acad. Sci. USA* **86,** 1934-1938.

Green, S., and Chambon, P. (1988). Nuclear receptors enhance our understanding of transcription regulation. *TiGs* **4,** 309-314.

Grupp, I. L., Subramaniam, A., Hewett, T. E., Robbins, J., and Grupp, G. (1993). Comparison of normal, hypodynamic, and hyperdynamic mouse hearts using isolated work-performing heart preparations. *Am. J. Physiol.* **265,** H1401-1411.

Haddox, M. K., Roeske, W. R., and Russell, D. H. (1979). Independent expression of cardiac type I and II cyclic AMP-dependent protein kinase during murine embryogenesis and postmatal development. *Biochim. Biophys. Acta.* **585,** 527-534.

Haffter, P., Granato, M., Brand, M., Mullins, M. C., Hammerschmidt, M., Kane, D. A., Odenthal, J., vanEeden, F., Jiang, Y. J., Heisenberg, C. P., Kelsh, R. N., FurutaniSeiki, M., Vogelsang, E., Beuchle, D., Schach, U., Fabian, C., and NussleinVolhard, C. (1996). The identification of genes with unique and essential functions in the development of the zebrafish, Danio rerio. *Development* **123,** 1-36.

Haldane, J. B. S., Sprunt, A. D., and Haldane, N. W. (1915). *J. Genetics* **5,** 133-135.

Han, Y., Dennis, J. E., Cohen-Gould, L., Bader, D. M., and Fischman, D. A. (1992). Expression of sarcomeric myosin in the presumptive myocardium of chicken embryos occurs within six hours of myocyte commitment. *Dev.Dynamics* **193,** 257-265.

Harrison, S. M., Dunwoodie, S.L., Arkell, R.M., Lehrach, H. and Beddington, R.S.P. (1995). Isolation of novel tissue-specific genes from cDNA libraries representing the individual tissue constituents of the gastrulating mouse embryo. *Development* **121,** 2479-2489.

Harvey, R. P. (1996). Nk-2 homeobox genes and heart development. *Dev Biol* **178,** 203-216.

Hicks, G. G., Shi, E.-g., Li, X.-M., Li, C.-H., Pawlak, M., and Ruley, H. E. (1997). Functional genomics in mice by tagged sequence mutagenesis. *Nature Gen* **16,** 338-344.

Hill, D., and Wurst, W. (1993). Screening for novel pattern formation genes using gene trap approaches. *Meth Enzymol* **225,** 664-81.

Hogan, B., Beddington, R., Costanini, F., and Lacy, E. (1994). *"Manipulating the Mouse Embryo. A Laboratory Manual."* Cold Spring Harbor Laboratory Press, Hogan, B. L. M. (1996). Bone morphogenetic proteins: multifunctional regulators of vertebrate development. *Genes & Dev* **10,** 1580-1594.

Hollenberg, S. M., Sternglanz, R., Cheng, P. F., and Weintraub, H. (1995). Identification of a new family of tissue-specific basic helix-loop-helix proteins with a two-hybrid system. *Mol.Cell.Biol* **15,** 3813-3822.

Horvitz, H. R., and Sternberg, P. W. (1991). Multiple intercellular signalling systems control the development of the *Caenorhabditis elegans* vulva. *Nature* **351,** 535-541.

Huang, J. X., Potts, J. D., Vincent, E. B., Weeks, D. L., and Runyan, R. B. (1995). Mechanisms of cell transformation in the embryonic heart. *Ann.N.Y.Acad.Sci.* **752,** 317-330.

Icardo, J. M., and Manasek, F. J. (1983). Fibronectin distribution during early chick embryo heart development. *Dev.Bol* **95,** 19-30.

Icardo, J. M., and Sanchez de Vega, M. J. (1991). Spectrum of heart malformations in mice with situs solitus, situs inversus, and associated visceral heterotaxy. *Circulation* **84,** 2547-2558.

Isaac, A., Sargent, M. G., and Cooke, J. (1997). Control of vertebrate left-right asymmetry by a *Snail*-related zinc finger gene. *Science* **275,** 1301-1304.

Ishikawa, H., and Yamada, E. (1975). Differentiation of the sarcoplasmic reticulum and T-system in developing mouse cardiac muscle. *In* "Developmental and Physiological Correlates of Cardiac Muscle" (M. Lieberman and T. Sano, Eds.). Raven Press, New York.

Itasaki, N., Nakamura, H., Sumida, H., and Yasuda, M. (1991). Actin bundles on the right side of the caudal part of the heart tube play a role in dextro-looping in the embryonic heart. *Anat.Embryol.* **183,** 29-39.

Jacobson, A. G., and Sater, A. K. (1988). Features of embryonic induction. *Development* **104,** 341-359.

Jaenisch, R. (1981). Chromosomal position and activation of retroviral genomes inserted into the germ line of mice. *Cell* **24,** 510-529.

Jegou, B. (1993). The Sertoli-germ cell communication network in mammals. *Int. Rev. Cytol.* **147,** 25-96.

Jin, Y., Pasamarthi, K. B., Bock, M. E., Lytras, A., Kardami, E., and Cattini, P. A. (1994). Cloning and expression of fibroblast growth factor receptor-1 isoforms in the mouse heart: evidence for isoform switching during heart development. *J. Mol. Cell Cardiol.* **26,** 1449-1459.

Johnson, D. B. a. D. I., L.J. (1996). Cardiac hypertrophy and failure in hypertension. *Current Opinion in Nephrology and Hypertension* **5,** 186-191.

Johnson M. D. and Mahon, K. A. (1993). Detection of Genes that are differentially expressed during mouse embryogenesis by genetic trapping strategies. *Am J Obstet Gynecol* **169,** 683-689.

Jones, W. E., Grupp, I. L., Doetschman, T., Grupp, G., Osinska, H., Hewett, T. E., Boivin, G., Gulick, J., Ng, W. A., and Robbins, J. (1996). Ablation of the murine α myosn heavy chain gene leads to dosage effects and functional deficits in the heart. *J.Clin.Invest.* **98,** 1906-1917.

Joyner, A. L., and Martin, G. R. (1987). *En-1* and *En-2*, two mouse genes with sequence homology to the *Drosophila engrailed* gene: expression during embryogenesis. *Genes Dev* **1,** 29-38.

Kamino, K., Hirota, A., and Fuji, S. (1981). Localisation of pacemaking activity in the early embryo heart monitored using voltage-sensitive dye. *Nature* **260,** 595-597.

Kastner, P., Grondona, J. M., Mark, M., Gansmuller, A., LeMeur, D., D., Vonesch, J.-L., Dolle, P., and Chambon, P. (1994). Genetic analysis of RXRα developmental function: Convergence of RXR and RAR signaling pathways in herart and eye morphogenesis. *Cell* **78,** 987-1003.

Kastner, P., Mark, M., Leid, M., Gansmuller, A., Chin, W., Grondona, J.M., Decimo, D., Kerzel, W., Dierich, A. and Chambon, P. (1996). Abnormal spermatogenesis in RXR beta mutant mice. *Gene Dev* **10,** 80-92.

Kaufman, M. H., and Navaratnam, V. (1981). Early differentiation of the heart in mouse embryos. *Journal of Anatomy* **133,** 235-246.

Kaufman, M. H. (1992). "The atlas of mouse development." Academic Press Limited, London.

Keller, G. M. (1995). *In vitro* differentiation of embryonic stem cells. *Curr Op Cell Biol* **7,** 862-869.

Kirby, M. L., and Stewart, D. E. (1983a). Neural crest origin of cardiac ganglion cells in the chick embryo: identification and extirpation. *Dev.Biol* **97,** 433-443.

Kirby, M. L., Gale, T. F., and Stewart, D. E. (1983b). Neural crest contribution to normal aorticopulmonary septation. *Science* **220,** 1059-1061.

Kirby, M. L., and Bockman, D. E. (1984). Neural crest and normal development: a new perspective. *Anat.Rec.* **209,** 1-6.

Kirby, M. L., and Stewart, D. E. (1984). Adrenergic innervation of the developing chick heart: neural crest ablations to produce sympathetically aneural hearts. *Am.J.Anat.* **171,** 295-305.

Komuro, I., and Izumo, S. (1993). A murine homeobox-containing gene specifically expressed in the developing heart. *Proc.Natl.Acad.Sci. USA* **90,** 8145-8149.

Koncz, C., Martini, N., Mayerhofer, R., Knocz-kalman, Z., Korber, H., Redei, G. P., and Schell, J. (1989). High frequency T-mediated gene tagging in plants. *Proc Natl Acad Sci* **86,** 8467-8471.

Korn, R., Schoor, M., Neuhaus, H., Henseling, U., Soininen, R., Zachgo, J., and Gossler, A. (1992). Enhancer trap integrations in mouse embryonic stem cells give rise to staining patterns in chimeric embryos with a high frequency and detect endogenous genes. *Mech Dev* **39,** 95-109.

Kozak, M. (1987). At least six nucleotides preceding he AUG initiator codon enhance translation in mammalian cells. *J Mol Biol* **196,** 947-950.

Kreidberg JA., Sariola H., Loring JM., Maeda M., Pelletier J., Housman D. and Jaenisch R. (1993) WT-1 is required for early kidney development. *Cell* **74(4)**, 679-691.

Kubalak, S. W., Miller-Hance, W. C., O'Brien, T. X., Dyson, E., and Chien, K. R. (1994). Chamber specification of atrial myosin light chain-2 expression precedes septation during murine cardiogenesis. *Journal of Biological Chemistry* **269,** 16961-16970.

Kuo, C. T., Morrisey, E. E., Anandappa, R., Sigrist, K., Lu, M. M., Parmacek, M. S., Soudais, C., and Leiden, J. M. (1997). GATA4 transcription factor is required for ventral morphogenesis and heart tube formation. *Genes Dev* **11,** 1048-1060.

Kwee, L., Baldwin, H. S., Shen, H. M., Stewart, C. L., Buck, C. A., and Labow, M. A. (1995). Defective development of the embryonic and extraembryonic circulatory systems in vascular cell adhesion molecule (VCAM-1) deficient mice. *Development* **121,** 489-503.

Lammer, E. J., Chen, D. T., and Hoar, R. M. (1985). Retinoic acid embryopathy. *New Eng.J.Med.* **313,** 837-841.

Lau, M. M. H., Stewart, C. E. H., Liu, Z., Bhatt, H., Rotwein, P., and Stewart, C. L. (1994). Loss of the imprinted IGF2/cation-independent mannose 6-phosphate receptor results in fetal overgrowth and perinatal lethality. *Genes and Development* **8,** 2953-2963.

Layton Jr, W. M. (1976). Random determination of a developmental process. *The Journal of Heredity* **67,** 336-338.

Lee, M. A., Bohm, M., Paul, M. and Ganten, D. (1993). Tissue renin-angiotensin systems. Their role in cardiovascular disease. *Circulation* **87,** iv7-iv13.

Lee, K. J., Hickey, R., Zhu, H., and Chien, K. R. (1994). Positive regulatory elements (HF-1a and HF-1b) and a novel negative regulatory element (HF-3) mediate ventricular muscle-specific expression of myosin light-chain 2-luciferase fusion genes in transgenic mice. *Mol. Cell Biol.* **14,** 1220-1229.

Leveen P., Pekny M., Gebre-Medhin S., Swolin B., Larsson E. and Betsholtz C. (1994) Mice deficient for PDGF B show renal, cardiovascular, and hematological abnormalities. *Genes Dev* **8(16)**,1875-1887.

Levi, G., Crossin, K. L., and Edelman, G. M. (1987). Expression, sequences and distribution of two primary cell adhesion molecules during embryonic development of *Xenopus laevis. J. Cell Biol.* **105,** 2359-2372.

Levin, M., Pagan, S., Roberts, D. J., Cooke, J., Keuhn, M. R., and Tabin, C. J. (1997). left/right patterning signals and the independent regulation of different aspects of *Situs* in the chick embryo. *Developmental Biology* **189,** 57-67.

Liang, P., Averboukh, L., and Pardee, A. B. (1993). Distribution and cloning of eukaryotic mRNAs by means of differential display: refinements and optimization. *Nucl. Acid Res.* **21,** 3269-3275.

Lih, C.-J., Wu, S-M. and Lin-Chao, S. (1995). Rapid identification and isolation of transcriptionally active regions from mouse genomes. *Gene* **164,** 289-294.

Lin, Q., Schwarz, J., Bucana, C., and Olson, E. N. (1997). Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C. *Science* **276,** 1404-1407.

Linask, K. K., and Lash, J. W. (1986). Precardiac cell migration: Fibronectin localization at mesoderm-endoderm interface during directional movement. *Dev.Biol.* **114,** 87-101.

Lindenbaum, M. H., and Grosveld, F. (1990). An *in vitro* globin gene switching model based on differentiated embryonic stem cells. *Genes Dev.* **4,** 2075-2085.

Lints, T. J., Parsons, L. M., Hartley, L., Lyons, I., and Harvey, R. P. (1993). Nkx-2.5: a novel murine homeobox gene expressed in early heart progenitor cells and their myogenic descendants. *Development* **119,** 419-31.

Lohnes, D., Mark, M., Mendelsohn, C., Dolle, P., Dierich, A., Gorry, P., Gansmuller, A., and Chambon, P. (1994). Function of the retinoic acid receptors (RARs) during development (1) Craniofacial and skeletal abnormalities in RAR double mutants. *Developement* **120,** 2723-2748.

Lovell-Badge, R. H. (1987). Introduction of DNA into embryonic stem cells. *In* "In Teratocarcinomas and Embryonic Stem Cells: A Practical Approach" (E. J. Robertson, Ed.), pp. 113-151. IRL Press, Oxford.

Lowe, L. A., Supp, D. M., Sampath, K., Yokoyama, T., Wright, C. V. E., Potter, S., Overbeek, P., and Kuehn, M. R. (1996). Conserved left-right asymmetry of *nodal* expression and alterations in murine *situs inversus. Nature* **381,** 158-161.

Lyons, G. E., Schiaffino, S., Sassoon, D., Barton, P. and Buckingham, M. (1990). Developmental regulation of myosin gene expression in mouse cardiac muscle. *Journal of Cell Biology* **111,** 2427-2436.

Lyons, I., Parsons, L.M., Hartley, L., Li, R., Andrews, J.E., Robb, L. and Harvey, R.P. (1995). Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeobox gene Nkx2.5. *Genes and Development* **9,** 1654-1666.

Maltsev, V. A., Wobus, A. M., Rohwedel, J., Bader, M., and Hescheler, J. (1994). Cardiomyocytes differentiated *in vitro* from embryonic stem cells developmentally express cardiac-specific genes and ionic currents. *Circ.Res.* **75,** 233-244.

Manasek, F. J. (1969). Myocardial cell death in the embryonic chick ventricle. *J. Embryol.Exp.Morphol.* **21,** 271-284.

Manasek, F. J., Burnside, M.B. and Waterman, R.E. (1972). Myocardial cell shape change as a mechanism of embryonic heart looping. *Develomental Biology* **29,** 349-371.

Manasek, F. J. (1975). The extracellular matrix of the early embryonic heart. *In* "Developmental and Physiological Correlates of Cardiac Muscle" (M. Lieberman and T. Sano, Eds.). Raven Press, New York.

Manasek, F. J. (1977). Structural glycoproteins of the embryonic cardiac extracellular matrix. *J. Mol. Cell. Cardiol.* **9,** 425-439.
Mangelsdorf, D. J. (1994). The retinoid receptors. *In* "The Retinoids" (M. B. Sporn, A. B. Roberts, and D. S. Goodman, Eds.), pp. 319-350. Raven Press, New York.

Mannan, M. A., and O'Shaughnessy, P. J. (1991). Steroidogenesis during postnatal development in the mouse ovary. *J. Endocrinology* **130,** 101-106.

Martin, G. R. (1981). Establishment of pluripotential cell lines from embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Nat. Acad. Sci. USA* **78,** 7634-7638.

McClive, P., Pall, G., Newton, K., Lee, M., Mullins, J., and Forrester, L. (1998). Gene-trap integrations expressed in the developing heart: Insertion affects splicing of the PT1-ATG vector. *Dev. Dynamics* **In Press.**

McGinnis, W., and Krumlauf, R. (1992). Homeobox genes and axial patterning. *Cell* **68,** 283-302.

Mendelsohn, C., Lohnes, D., Decimo, D., Lufkin, T., LeMeur, M., Chambon, P., and Mark, M. (1994). Function of the retinoic acid receptors (RARs) during development (II). *Development* **120,** 2749-2771.

Menichini P., Viaggi S., Gallerani E., Fronza G., Ottaggio L., Comes A., Ellwart JW. and Abbondandolo A. (1997) A gene trap approach to isolate mammalian genes involved in the cellular response to genotoxic stress. *Nucleic Acids Res.* **25(23),** 4803-4807.

Meno, C., Siajoh, Y., Fujii, H., Ikeda, M., Yokoyama, T., Yokoyama, Y., Toyoda, Y. and Hamada, H. (1996). Left-right asymmetric expression of the TGFβ-family member *lefty* in mouse embryos. *Nature* **381,** 151-155.

Meyer, D., and Birchmeier, C. (1995). Multiple essential functions of neuregulin in development. *Nature* **378,** 386-390.

Meyer, D., Yamaai, T., Garratt, A., Riethmacher-Sonnenberg, E., Kane, D., Theill, L. E., and Birchmeier, C. (1997). Isoform-specific expression and function of neuregulin. *Development* **124,** 3575-3586.

Miklos, G., and Rubin, G. M. (1996). The role of the genome project in determining gene-function - insights from model organisms. *Cell* **86,** 521-529.

Miller-Hance, W. C., LaCorbiere, M., Fuller, S. J., evans, S. M., Lyons, G., Schmidt, C., Robbins, J., and Chien, K. R. (1993). *In vitro* chamber specification during embryonic stem cell cardiogenesis. *Journal of Biological Chemistry* **268,** 25244-25252.

Mlodzik, M., Hiromi, Y., Weber, U., Goodman, C. S., and Rubin, G. M. (1990). The *Drosophila seven-up* gene, a member of the steroid receptor gene superfamily controls photoreceptor cell fates. *Cell* **60,** 211-224.

Moens, C. B., Stanton, B. R., Parada, L. F., and Rossant, J. (1993). Defects in heart and lung development in compound heterozygotes for two different targeted mutations at the N-myc locus. *Development* **119,** 485-499.

Molkentin, J. D., Black, B. L., Martin, J. F., and Olson, E. N. (1995). Cooperative activation of muscle gene expression by MEF2 and Myogenic bHLH proteins. *Cell* **83,** 1125-1136.

Molkentin, J. D., Lin, Q., Duncan, S. A., and Olson, E. N. (1997). Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. *Genes Dev* **11,** 1061-1072.

Moorman, A. F. M., de Jong, F., and Lamers, W. H. (1997). Development of the conduction system of the heart. *PACE* **20,** 2087-2092.

Morriss-Kay, G. (1991). Retinoic acid, neural crest and craniofacial development. *In* "Seminars in Developmental Biology" (M. Maden and C. Tickle, Eds.), Vol. 2, pp. 211-218.

Morriss-Kay, G. (1993). Retinoic acid and craniofacial development: molecules and morphogenesis. *Bioessays* **15,** 9-15.

Mortensen, R. M., Conner, D.A., Chao, S., Geisterfer-Lowrance, A.A.T and Seidman, J.G. (1992). Production of homozygous mutant ES cells with a single targeting construct. *Molecular and Cellular Biology* **12,** 2391-2395.

Mountford, P. S., and Smith, A. G. (1995). Internal ribosome entry sites and dicistronic RNAs in mammalian transgenesis. *Trends Genetics* **11,** 179-184.

Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. and Roder, J.C. (1993). Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proceedings of the National Academy of Science, USA* **90,** 8424-8428.

Nantel, F., Monaco, L., Foulkes, N.S., Masquilier, D., LeMeur, M., Henriksen, K., Dierich, A., Parvinen, M. and Sassone-Corsi, P. (1996). Spermiogenesis deficiency and germ-cell apoptosis in CREM-mutant mice. *Nature* **380,** 159162.

Nascone, N., and Mercola, M. (1995). An inductive role for the endoderm in *Xenopus* cardiogenesis. *Development* **121,** 515-523.

Neuhaus, H., Bettenhausen, B., Bilinski, P., Simon-Chazottes, D., Guenet, J.-L., and Gossler, A. (1994). Etl2, A novel putative type-I cytokine receptor expressed during mouse embryogenesis at high levels in skin and cells with skeletogenic potential. *Dev. Biol.* **166,** 531-542.

Ng, W. A., Grupp, I. L., Subramaniam, A., and Robbins, J. (1991). Cardiac myosin heavy chain mRNA expression and myocardial function in the mouse heart. *Circ. Res.* **68,** 1742-1750.

Nguyen, M. T., Camenisch, T., Snouwaert, J. N., Hicks, E., Coffman, T. M., Anderson, P. A. W., Malouf, N. N., and Koller, B. H. (1997). The prostaglandin receptorEP_4 triggers remodelling of the cardiovascular system at birth. *Nature* **390,** 78-81.

Nichols, D. H. (1987). Ultrastructure of neural crest cell formation in the midbrain/rostral hindbrain and preotic hindbrain regions of the mouse embryo. *Am. J. Anat.* **179,** 143-154.

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

Noden, D. (1983). The role of neural crest in patterning of avian cranial skeletal, connective and muscle tissues. *Dev. Biol.* **96,** 144-165.

Nusslein-Volhard, C., and Wieschaus, E. (1980). Mutations affecting segment number and polarity in Drosophila. *Nature* **287,** 795-801.

Nusslein-Volhard, C. (1994). Of flies and fishes. *Science* **266,** 572-574.

O'Brien, T. X., Lee, K. J., and Chien, K. R. (1993). Positional specification of ventricular myosin light chain 2 expression in the primitive murine heart tube. *Proceedings of the National Academy of Science, USA* **90,** 5157-5161.

O'Kane, C., and Gehring, W. (1987). Detection in situ of genomic regulatory elements in *Drosophila. Proc Natl Acad Sci* **84,** 9123-9127.

Okamata, N., and Satow, Y. (1975). Cell death in bulbar cushion of normal and abnormal developing heart. *In* "Developmental and Physiological Correlates of Cardiac Muscle" (M. Lieberman and T. Sano, Eds.). Raven Press, New York.

Olson, M., Hood, L., Cantor, C., and Botstein, D. (1989). A common language for physical mapping of the human genome. *Science* **245,** 1434-1435.

Orkin, S., and Kazazian, H. (1984). The mutation and polymorphism of the human β-globin gene and its surrounding DNA. *Ann. Rev. Genet* **18,** 131-171.

Papalopulu, N., Clarke, D. W., Bradley, L., Wilkinson, D., Krumlauf, R., and Holder, N. (1991). Retinoic acid causes abnormal development and segmental patterning of the anterior hindbrain in *Xenopus* embryos. *Development* **113,** 1145-1158.

Parfrey, P. S., Harnett, J. D., and Foley, R. N. (1995). Heart failure and ischemic heart disease in chronic uremia. *Curr. Op. Nephro. Hyperten.* **4,** 105-110.

Pelletier, J., Schalling, M., Buckler, A. J., Rogers, A., Haber, D. A., and Housman, D. (1991). Expression of the Wilms' tumor gene WT1 in the murine urogenital system. *Genes Dev.* **5,** 1345-1356.

Peten, E. P. a. S., L.J. (1994). Progression of glomerular diseases. *Journal of Internal Medicine* **236,** 241-249.

Pollock, R., and Treisman, R. (1991). Human SRF-related proteins: DNA-binding properties and potential regulatory targets. *Genes Dev.* , 2327-2341.

Propst, F., Rosenberg, M. P., Oskarsson, M. K., Russell, L. B., Nguyen-Huu, M. C., Nadeau, J., Jenkins, N. A., Copeland, N. G., and Vande Woude, G. F. (1988). Genetic analysis and developmental regulation of testis-specific RNA expression of Mos, Abl, actin and Hox-1.4. *Oncogene* **2,** 227-233.

Pruitt, K. D. (1997). WebWise: navigating the Human Genome Project. *Genome Research* **7,** 1038-1039.

Rabbits, P., Impey, H., Heppell-Parton, A., Langford, C., Tease, C., Lowe, N., Bailey, D., Ferguson-Smith, M., and Carter, N. (1995). Chromosome specific paints from a high resolution flow karyotype of the mouse. *Nat Genet* **9,** 369-375.

Reaume, A. G., de Sousa, P. A., Kulkarni, S., Langille, B. L., Zhu, D., Davies, T. C., Juneja, S. C., Kidder, G. M., and Rossant, J. (1995). Cardiac malformation in neonatal mice lacking connexin43. *Science* **267,** 1831-1834.

Reddy, S., Rayburn, H., von Melchner, H., and Ruley, H. E. (1992). Fluorescence-activated sorting of totipotent embryonic stem cells expressing developmentally regulated *lacZ* fusion genes. *Proceedings of the National Academy of Science, USA* **89,** 6721-6725.

Rinchik, E. M. (1991). Chemical mutagenesis and fine-structure functional analysis of the mouse genome. *Trends Genetics* **7,** 15-21.

Risau, W., Sariola, H., Zerwes, H. G., Sasse, J., Ekblom, P., Kemler, R., and Doetschman, T. (1988). Vasculogenesis and angiogenesis in embryonic-stem-cell-derived embryoid bodies. *Development* **102,** 471-478.

Robbins, J., Gulick, J., Sanchez, A., Howles, P., and Doetschman, T. (1990). Mouse embryonic stem cells express the cardiac myosin heavy chain genes during development *in vitro. Journal of Biological Chemistry* **265,** 11905-11909.

Robertson, E., Bradley, A., Kuehn, M., and Evans, M. (1986). Germ-line transmission of genes introduced into cultured pluripotential cells by a retroviral vector. *Nature* **309,** 255-256.

Rockman, H. A., Ross, R. S., and Harris, A. N. (1991). Segregation atrial specific and inducible expression of an ANF transgene in an in vivo murine model of cardiac hypertrophy. *Proc. Nat. Acad, Sci. USA* **88,** 8277-8281.

Rockman, H. A., Knowlton, K. U., Ross, J. R. j., and Chien, K. R. (1993). In vivo murine cardiac hypertrophy: A novel model to identify genetic signaling mechanisms which activate an adaptive physiologic response. *Circulation* **87,** 14-21.

Rockman, H. A., Ono, S., Ross, R. S., Jones, L. R., Karimi, M., Bhargava, V., Ross Jr., J., and Chien, K. R. (1994). Molecular and physiological alterations in murine ventricular dysfunction. *Proceedings of the National Academy of Science, USA* **91,** 2694-2698.

Ross, R. S., Navankasattusas, S., Harvey, R. P., and Chein, K. R. (1996). An HF-1a/HF-1b/MEF-2 combinatorial element confers cardiac ventricular specificity and establishes an anterior-posterior gradient of expression. *Development* **122,** 1799-1809.

Rudnicki, M. A., Braun, T., Hinuma, S., and and Jaenisch, R. (1992). Inactivation of MyoD in Mice Leads to Up-Regulation of the Myogenic HLH Gene Myf-5 and Results in Apparently Normal Muscle Development. *Cell* **71,** 383-390.

Ruiz-Lazano, P., Smith, S. M., Perkins, G., Kubalak, S. W., Boss, G. R., Sucov, H. M., Evans, R. M., and Chien, K. R. (1998). Energy deprivation and a deficiency in downstream metabolic target genes during the onset of embryonic heart failure in $RXR\alpha^{-/-}$ embryos. *Development* **125,** 533-544.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). "Molecular Cloning. A Laboratory Manual." Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Sanchez, A., Jones, W.K., Gulick, J., Doetschman, T. and Robbins, J. (1991). Myosin heavy chain gene expression in mouse embryoid bodies. An in vitro developmental study. *Journal of Biological Chemistry* **226,** 22419-22426.

Sanger, F., Niklen, S. and Coulson, A.R. (1977). *Proc. Nat. Acad. Sci. USA* **74,** 5463-5467.

Sawai S., Shimono A., Wakamatsu Y., Palmes C., Hanaoka K. and Kondoh H. (1993) Defects of embryonic organogenesis resulting from targeted disruption of the N-myc gene in the mouse. *Development* **117(4)**,1445-1455.

Scherer, C. A., Chen, J., Nachabeh, A., Hopkins, N., and Ruley, H. E. (1996). Transcriptional specificity of the pluripotent embryonic stem cell. *Cell Growth Diff* **7,** 1393-1401.

Schier, A. F., Joyner, A. L., Lehmann, R., and Talbot, W. S. (1997). From screens to genes: prospects for insertional mutagenesis in zebrafish. *Genes and Development* **10,** 3077-3080.

Schultheiss, T. M., Xydas, S., and Lassar, A. B. (1995). Induction of avian cardiacmyogenesis by anterior endoderm. *Development* **121,** 4203-4214.

Schultheiss, T. M., Burch, J. B. E., and Lassar, A. B. (1997). A role for bone morphogenetic proteins in the induction of cardiac myogenesis. *Genes and Development* **11,** 451-462.

Schuster-Gossler, K., Zachago, J., Soininen, R., Schoor, M., Korn, R., and Gossler, A. (1994). Gene trap integrations in genes active in mouse embryonic stem cells efficiently detect developmentally regulated gene expression. *Transgenics* **1,** 281-291.

Schuster-Gossler, K., Simonchazottes, D., Guenet, J. L., Zachgo, J., and Gossler, A. (1996). Gtl2(lacz) an insertional mutation on mouse chromosome-12 with parental origin-dependent phenotype. *Mammalian Genome* **7** , 20-24.

Scrocchi, L. A., Hearn, S. A., Han, V. K., and Hammond, G. L. (1993). Corticosteroid-binding globulin biosynthesis in the mouse liver and kidney during postnatal development. *Endocrinology* **132,** 910-916.

Shedlovsky, A., King, T. R., and Dove, W. F. (1988). Saturation germ line mutagenesis of the murine *t* region including a lethal allele at the *quaking* locus. *Proc Nat Acad Sci* **85,** 180-184.

Sheffield, J. W., and O'Shaughnessy, P. J. (1988). Testicular steroid metabolism during development in the normal and hypogonadal mouse. *J. Endocrinology* **119,** 257-264.

Sheng, Z., Pennica, D., Wood, W.I. and Chien, K.R. (1996). Cardiotrophin-1 displays early expression in the murine heart tube and promotes cardiac myocyte survival. *Development* **122,** 419-428.

Simeone, A., Acampora, D., Nigro, V., Faiella, A., D'Esposito, M., Stornaiuolo, A., Mavilio, F., and Bonicinelli, E. (1991). Differential regulation by retinoic acid of the homeobox genes of four Hox loci in human embryonal carcinoma cells. *Mech.Dev.* **33,** 215-228.

Simpson, E. M., Linder, C. C., Sargent, E. E., Davisson, M. T., Mobraaten, L. E., and Sharp, J. J. (1997). Genetic variation among 129 substrains and its importance for targeted mutagenesis in mice. *Nature Genetics* **16,** 19-27.

Skarnes, W. C. (1990). Entrapment vectors : A new tool for mammalian genetics. *Biotech* **8,** 827-831.

Skarnes, W. C., Auerbach, B. A., and Joyner, A. L. (1992). A gene trap approach in mouse embryonic stem cells: the LacZ reporter is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice. *Genes Dev* **6,** 903-918.

Skarnes, W. C. (1993). The identification of new genes: gene trapping in transgenic mice. *Curr Biol* **4,** 684-689.

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

Skinner, M. K. (1991). Cell-cell interactions in the testis. *Endocrinology Reveiws* **12.**

Smith, A. G., Heath, J. K., Donaldson, D. D., Wong, G. G., Moreau, J., Stahl, M., and Rogers, D. (1988). Inhibition of pluripotential embryonic stem cell differntiation by purified polypeptides. *Nature* **336,** 688-690.

Smith, A. G. (1991). Culture and Differentiation of Embryonic Stem Cells. *Journal of Tissue Culture Methods* **13,** 89-94.

Smolich, J. J. (1995). Ultrastructural and functional features of the developing mammalian heart: a brief overview. *Reprod. Fertil. Dev.* **7,** 451-461.

Soininen, R., Schoor, M., Henseling, U., Tepe, C., Kisters-Woike, B., Rossant, J., and Gossler, A. (1992). The Mouse Enhancer Trap Locus 1 (Etl-1): A novel mammalian gene related to *Drosophila* and Yeast transcriptional regulator genes. *Mech. Dev.* **39,** 111-123.

Spradling, A. C., Stern, D. M., Kiss, I., Roote, J., Laverty, T., and Rubin, G. M. (1995). gene disruptions using P transposable elements: An integral component of the *Drosophila* genome project. *Proc Natl Acad Sci USA* **92,** 10824-10830.

Srivastava, D., Cserjesi, P., and Olson, E. N. (1995). A subclass of bHLH proteins required for cardiac morphogenesis. *Science* **270,** 1995-9.

Srivastava, D., Thomas, T., Lin, Q., Kirby, M. L., Brown, D., and Olson, E. N. (1997). Regulation of cardiac mesodermal and neural crest development by the bHLH transcription factor, dHAND. *Nature Genetics* **16,** 154-160.

Stainier, D. Y., and Fishman, M. C. (1992). Patterning the zebrafish heart tube: acquisition of anteroposterior polarity. *Dev.Biol.* **153,** 91-101.

Stewart, D. E., Kirby, M. L., and Sulik, K. K. (1986). Hemodynamic changes in chick embryos preceeds heart defects after neural crest ablation. *Circ.Res.* **59,** 545-550.

Strubing, C., Ahnert-Hilger, G., Shan, J., Wiedemann, B., Hescheler, J., and Wobus, A. M. (1995). Differentiation of pluripotent embryonic stem cells into the neuronal lineage in vitro gives rise to mature and excitatory neurons. *Mech. Dev.* **53,** 275-287.

Sucov, H. M., Dyson, E., Gumeringer, C.L., Price, J., Chien, K.R. and Evans, R.M. (1994). RXRα mutant mice establish a genetic basis for vitamin A signaling in heart morphogenesis. *Genes and Development* **8,** 1007-1018.

Tabin, C. J. (1991). Retinoids, homeoboxes and growth factors: Towards the molecular models for limb development. *Cell* **66,** 199-217.

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

Taugner, R. a. H., E. (1989). "The Juxtaglomerular Apparatus: Structure and Function." Springer Verlag, Heidelberg, Germany.

Teitel, D. F. (1988). Circulatory adjustments to postnatal life. *Semin. Perinatol.* **12,** 96-103.

Terracio, L., Rubin, K., Gullberg, D., Balog, E., Carver, W., Jyring, R., and Borg, T. K. (1991). Expression of collagen binding integrins during cardiac development and hypertrophy. *Circ. Res.* **68,** 734-744.

Threadgill, D. W. (1995). Targeted disruption of mouse EGF receptor: Effect of Genetic Background on Mutant Phenotype. *Science* **269,** 230-234.

Tokunaga K., Taniguchi H., Yoda K., Shimizu M. and Sakiyama S. (1986) Nucleotide sequence of a full-length cDNA for mouse cytoskeletal beta-actin mRNA. *Nucleic Acids Res* **14(6),** 2829.

Torres, M., Stoykova, A., Huber, O., Chowdhury, K., Bonaldo, P., Mansouri, A., Butz, S., Kemler, R. and Gruss, P. (1997) An a-E-catenin gene-trap mutation defines its function in preimplantation development. *P.N.A.S. USA* **94**, 901-906.

Townley, D. J., Avery, B. J., Rosen, B., and Skarnes, W. C. (1997). Rapid sequence analysis of gene trap integrations to generate a resource of insertional mutations in mice. *Genome Research* **7,** 293-298.

Tsuda, T., Philp, N., Zile, M. H., and Linask, K. K. (1996). Left-right asymmetric localization of flectin in extracellular matrix during heart looping. *Developmental Biology* **173,** 39-50.

Tsutsui, H., Ishihara, K., and Cooper IV, G. (1993). Cytoskeletal role in the contractile dysfunction of hypertrophied myocardium. *Science* **260,** 682-687.

Viragh, S., and Challice, C. E. (1977a). The development of the conduction system in the mouse embryo. I. The first embryonic A-V conduction pathway. *Developmental Biology* **56,** 382-398.

Viragh, S., and Challice, C. E. (1977b). The development of the conduction system in the mouse embryo heart. II. Histogenesis of the atrioventricular node and bundle. *Developmental Biology* **56,** 397-411.

Viragh, S., and Challice, C. E. (1980). The development of the conduction system in the mouse embryo II. *Developmental Biology* .

Viragh, S., and Challice, C. E. (1982). The development of the conduction system in mouse embryo heart IV. *Developmental Biology* **89,** 25-40.

von Melchner, H. a. R., H. E. (1989). Identification of Cellular Promoters by Using a Retrovirus Promoter Trap. *J Virol* , 3227-3233.

von Melchner, H., Reddy, S. and Ruley, H. E. (1990). Isolation of Cellular Promoters by using a Retrovirus Promoter Trap. *Proc.Natl.Acad.Sci.USA* **87,** 3733-3737.

von Melchner, H., deGregori, J. V., Rayburn, H., Reddy, S., Friedel, C. and Ruley, H. E. (1992). Selective Disruption of Genes Expressed in Totipotent Embryonal Stem Cells. *Gen Dev* **6,** 919-927.

Vuillemin, M., and Pexieder, T. (1989). Normal stages of cardiac organogenesis in the mouse: II. Development of the internal relief of the heart. *Am. J. Anat.* **184,** 114-128.

Vuillemin, M., and Pexieder, T. (1989a). Normal stages of cardiac organogenesis in the mouse: I. Development of the external shape of the heart. *Am. J. Anat.* **184,** 101-113.

Wang, L., Feng, Z. P., Kondo, C. S., Sheldon, R. S., and Duff, H. J. (1996). Developmental changes in the delayed rectifier K+ channels in mouse heart. *Circ. Res.* **79,** 79-85.

Wang, R., Hwang, D. M., Cukerman, E., and Liew, C.-C. (1997). Identification of genes encoding zinc finger motifs in the cardiovascular system. *j. Mol.Cell Cardiol.* **29,** 281-287.

Wang, L., and Duff, H. J. (1997). Developmental changes in transient outward current in mouse ventricle. *Circ. Res.* **81,** 120-127.

Webster, W. S., Johnston, M. C., Lammer, E. J., and Sulik, K. K. (1986). Isotretinoin embryopathy and the cranial neural crest: an in vivo and in vitro study. *J.Cranio.Genetics and Dev.Biol.* **6,** 211-222.

Wilks, A. F., Kurban, R. R., Hovens, C. M., and Ralph, S. J. (1989). The application of the polymerase chain reaction to cloning members of the protein tyrosine kinase family. *Gene* **85,** 76-74.

Williams, R. L., Hilton, D. J., Peases, S., Wilson, T. A., Stewart, C. L., Gearing, D. P., Wagner, E. F., Metcalf, D., Nicola, N. A., and Gough, N. M. (1988). Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336,** 684-687.

Wilson, J. G., and Warkany, J. (1949). Fetal death and maldevelopment resulting from maternal vitamin A deficiency in the rat. *Proc.Soc.Exp.Biol. and Med.* **72,** 687-693.

Wilson, J. G., Roth, C. B., and Warkany, J. (1953). An analysis of the syndrome of malformations induced by maternal vitamin A deficiency. *Am.J.Anat.* **92,** 189-217.

Wilson, C., Pearson, R. K., Bellen, H. J., O'Kane, C., Grossniklaus, U., and Gehring, W. J. (1989). P-element-mediated enhancer detection: an efficient method for isolating and characterising developmentally regulated genes in Drosophila. *Gene Dev* **3,** 1301-1313.

Wobus, A. M., Wallukat, G., and Hescheler, J. (1991). Pluripotent mouse embryonic stem cells are able to differentiate into cardiomyocytes expressing chronotropic responses to adrenergic and cholinergic agents and Ca2+ channel blockers. *Differentiation* **48,** 173-182.

Wobus, A. M., Rohwedel, J., Maltsev, V. and Hescheler, J. (1994). In vitro differentiation of embryonic stem cells into cardiomyocytes or skeletal muscle is specifically modulated by retinoic acid. *Roux's Archives of Developmental Biology* **204,** 36-45.

Wurst, W., Rossant, J., Prideaux, V., Kownacka, M., Joyner, A., Hill, D. P., Guillemot, F., Gasca, S., Cado, D., Auerbach, A., and Ang, S.-L. (1995). A large-scale gene trap screen for insertional mutations in developmentally-regulated genes in mice. *Genetics* **139,** 889-899.

Xu, W., Baribault, H., and Adamson, E. D. (1998). Vinculin knockout results in heart and brain defects during embryonic development. *Development* **125,** 327-337.

Yamamura, H., Zhang, M., Markwald, R. R., and Mjaatvedt, C. H. (1997). A heart segmental defect in the anterior-posterior axis of a transgenic mutant mouse. *Dev.Biol.* **186,** 58-72.

Yang, J. T., Rayburn, H., and Hynes, R. O. (1995). Cell adhesion events mediated by α4 integrins are essential in placental and cardiac development. *Development* **121,** 549-560.

Yokoyama, T., Copeland, N. G., Jenkins, N. A., Montgomery, C. A., Elder, F. F. B., and Overbeek, P. A. (1993). Reversal of Left-right asymmetry: a *situs inversus* mutation. *Science* **260,** 279-682.

Yoshida, M., Yagi, T., Furuta, Y., Takayanagi, K., Kominami, R., Takeda, N., Tokunaga, T., Chiba, J., Ikawa, Y., and Aizawa, S. (1995). A new strategy of gene trapping in ES cells using 3' RACE. *Trans Res* **4,** 277-287.

Yoshida, K., Taga, T., Saito, M., Suematsu, S., Kumanogoh, A., Tanaka, T., Fujiwara, H., Hiarata, M., Yamagami, T., Nakahata, T., Hirabatashi, T., Yoneda, Y., Tanaka, K., Wang, W.-Z., Mori, C., Shiota, K., Yoshida, N., and Kishimoto, T. (1996). Targeted disruption of gp130, a common signal transducer for the interleukin 6 family of cytokines, leads to myocardial and hematological disorders. *Proceedings of the National Academy of Science, USA* **93,** 407-411.

Yost, H. J. (1992). Regulation of vertebrate left-right asymmetries by extracellular matrix. *Nature* **357,** 158-161.

Zambrowicz, B. P., Imamoto, A., Fiering, S., Herzenberg, L. A., Kerr, W. G., and Soriano, P. (1997). Disruption of overlapping transcripts in the ROSA beta geo 26 gene trap strain leads to widespread expression of beta-galactosidase in mouse embryos and hematopoietic cells. *Proceedings Of The National Academy Of Sciences Of The United States Of America* **94**, 3789-3794.

Zhou, M. D., Sucov, H. M., Evans, R. M., and Chien, K. R. (1995). Retinoid-dependent pathways suppress myocardial cell hypertrophy. *Proceedings of the National Academy of Science, USA* **92,** 7391-7395.

Zhu, H., Garcia, A. V., R.S., R., Evans, S. M., and Chien, K. R. (1991). A conserved 28-base-pair element (HF-1) in the rat cardiac myosin light-chain-2 gene confers cardiac-specific and α-adrenergic -inducible expression in cultured neonatal rat myocardial cells. *Molecular and Cellular Biology* **11,** 2273-2281.