Addendum

1.1   Page 2   *Mus Muscularis* should read *Mus Musculus*

1.2.2 Page 5   Cell polarity is first apparent at the 2-cell stage, with the microvilli at the side of the cell not in contact with the second blastomere. This polarity is reversible until compaction when tight junctions form among the cells on the outside of the morula and existing gap junctions between blastomeres persist between cells on the inside of the compacted morula, as well as between these cells and the outside cells.

The trophectoderm gives rise only to extraembryonic structures while the ICM cells give rise to the embryo proper and also to extraembryonic tissues.

Figure 1.2   The definitive endoderm arises not from proximal epiblast but from epiblast near the newly forming (posterior) primitive streak.

1.2.4 Page 7   Mice are different to other mammalian species in that the chorion does not envelope the entire conceptus and other three foetal membranes.

1.2.5 Page 10   Presently, evidence for a neural inducing signal from the notochord and head processes underlying medial and anterior ectoderm is lacking. The alternate theory for a default differentiation of primitive ectoderm that has not passed through the streak may not, however, explain neurectoderm formation fully (reviewed by Streit and Stern, 1999)[*].

Tonsils, thymus, thyroid and lungs are foregut not hindgut derivatives.

5.1 Page 140   passage 24 should read passage 14.

Only after subcloning the parental line was 1 of 5 subclones able to form predominantly ES cell-derived live newborn offspring after passage 14.

Tetraploid cells can contribute up to 15% to the chimaeric embryo without causing death.

6.5   The reporter gene activity may not faithfully reflect *Fbx15* activity *in vivo*, as RT-PCR analyses for the *Fbx15* gene show a weak PCR product for some tissues in which β-galactosidase activity is not detected. *In situ* hybridisation analysis using an *Fbx15*-specific probe will therefore be required to confirm the endogenous gene expression profile. The integrity of the reporter gene expression could also be further evaluated by RT-PCR analyses for the *βgeo* sequence.

*Streit, A. and Stern, C.D. (1999). Neural induction: a bird's eye view. [Review]. *Trends. Genetics* 15(1), 20-24.

IDENTIFICATION AND CLONING OF EMBRYONIC STEM CELL-SPECIFIC GENES

Carmel Maureen O'Brien
B.App.Sc. (Biochemistry)
M.Rep.Sc. (Reproductive Sciences)

Thesis presented for the degree of Doctor of Philosophy
Faculty of Medicine
Monash University, Melbourne, Australia

May, 2001

To my parents, Victor and Noreen, with much love and thanks.

Contents

Chapter One

Introduction

Chapter Two

Materials and Methods

Chapter Three

Gene trap integration & in vitro screening for ES cell-specific genes

Chapter Four

Molecular analysis of 28 gene trap integrations in candidate ES cell-specific genes

Chapter Five

Detailed analysis of gene trap insertions within five novel ES cell-specific genes

Chapter Six

A detailed expression profile for the COB54 gene trap insertion in ES cells and transgenic mice

Chapter Seven

Conclusion & Future directions

Summary

Embryonic stem (ES) cells are pluripotential cells derived from the inner cell mass of preimplantation embryos which can spontaneously differentiate *in vitro* to cell types representing all three germ layers of the developing embryo. While ES cells are most widely known for their ability to transmit mutations introduced *in vitro* into the mouse germline, recent reports of human ES cells have lead to widespread speculation that these unique cells may ultimately provide an unlimited source of many differentiated cell types for the development of human cell-based therapies. Presently, however, the realization of such applications is hindered by the inability to duplicate the robust *in vitro* culture systems underpinning the usefulness of mouse ES cells. A greater understanding of the biological processes controlling mouse ES cell self-renewal and commitment to restricted cell lineages may provide new opportunities to address this limitation.

This study reports a successful and powerful new approach for the identification and cloning of genes which are expressed in undifferentiated mouse ES cells and downregulated during *in vitro* differentiation. A strategy employing the introduction of an improved gene trap vector into mouse ES cells, coupled with a stringent three-tiered *in vitro* screen for gene trap clones expressing a reporter gene in an ES cell-restricted fashion, has been used to screen in excess of 4000 randomly generated clones. This strategy identified 28 ES cell clones harboring gene trap integrations in candidate ES cell-specific genes, 10 of which yielded correctly spliced, single 5' entrapped gene sequence reads for mRNA fusion sequence following direct sequencing of 5'-RACE products. Five cDNAs of interest were identified for further analysis, four of which presently stand as uncharacterized genes.

One ES cell gene trap line, COB54, was capable of transmitting the gene trap insertion into the germline, and was of particular interest for its 5' sequence homology with ESTs from mouse preimplantation cDNA libraries and tightly restricted *in vitro* expression profile. The COB54 cell line carries a mutation in a gene only recently identified as a member of a new family of approximately 50 mammalian proteins which incorporate a conserved 38 amino acid F-box motif (Cenciarelli *et al.*, 1999; Winston *et al.*, 1999a). F-box proteins are substrate-specific adaptor subunits involved in recruiting target proteins for rapid intracellular degradation via the ubiquitin-dependent proteolytic

pathway, and are linked by their common motif to a core ubiquitination complex. Emerging evidence implies an important role for this expanding family of proteins in the regulation of many cellular processes including G1-S phase transition, transcription, signal transduction and development.

COB54 ES cells and transgenic mice show dramatic downregulation of the nuclear localized reporter gene expression upon ES cell differentiation and blastocyst implantation. Interestingly, a striking upregulation of expression is seen in the adult testis, which appears to be associated at a cellular level with the meiotic phase of spermatogenesis. While mice homozygote for the F-box protein gene trap insertion do not show any obvious morphological or behavioural phenotype, the carboxy end integration site, as well as possible low level expression of the wild type transcript, does not exclude the possibility of an important developmental function for this gene.

This study reports the identification of a new ES cell marker gene using a random gene trap and *in vitro* screening strategy. The exquisitely restricted expression pattern and the emerging regulatory role of F-box gene family members makes COB54 a particularly interesting candidate gene in ES cell regulation. Further studies including a functional deletion and elucidation of the substrate(s) targeted by this F-box protein may shed light on a new ES cell regulatory pathway and in doing so, provide new opportunities to enhance ES cell growth for a range of applications including cell-based therapies.

Declaration

The material contained in this thesis is original work, which has not been accepted for the award of any other degree or diploma at any university, and to the best of my knowledge does not contain material previously published or written by any other person, except where referenced within the text.
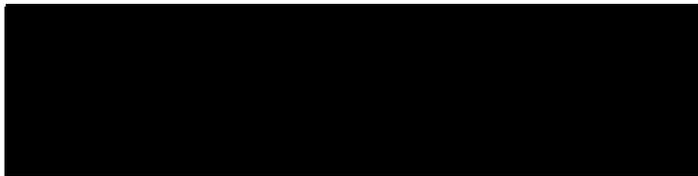
Carmel O'Brien

Acknowledgements

Publications

Journal publications:

O'Brien, C.M. and Mountford, P.S.   Identification of an embryonic stem cell-restricted expression profile for a novel F-box protein. (*In preparation*).

Munsie, M.J., **O'Brien, C.M.,** and Mountford, P.S.   Transgenic strategies for investigation and development of nuclear reprogramming techniques. (*In preparation*).

Munsie, M.J., Michalska, A.E., **O'Brien, C.M.,** Trounson, A.O. and Mountford, P.S. (2000).   Isolation of pluripotent embryonic stem cells from reprogrammed adult somatic cell nuclei. *Curr. Biol.*, 10, 989-992.

Mountford, P., Nichols, J., Zevnik, B., **O'Brien, C.,** Smith, A. (1998).   Maintenance of pluripotential embryonic stem cells by stem cell selection. *Reprod. Fertil. Dev.*, 10, 527-533.

Conference abstracts:

O'Brien, C.M. and Mountford, P.S. (2000).   Gene trap cloning of embryonic stem cell-restricted genes. *Gordon Research Conference on Mammalian Gametogenesis and Embryogenesis,* New London, CT, USA.   *Invited speaker* – Received junior speaker award.

O'Brien, C.M., Trounson, A.O., Mountford, P.S. (1999).   Identification of a novel gene with embryonic stem cell-specific expression. *Combio 99 conference* - joint annual meetings for *the Australian & New Zealand Society for Cell & Developmental Biology Incorporated (ANZSCDBI), Australian ·Society for Biochemistry & Molecular Biology (ASBMB), and Australian Society of Plant Physiologists (ASPP),* Gold Coast, Australia.   Received ANZSCDBI Keith Dixon award in Developmental Biology.

O'Brien, C.M., Trounson, A.O., Mountford, P.S. (1997). Identification and cloning of embryonic stem cell-specific genes. *16th annual meeting of the ANZSCDBI,* Melbourne, Australia.

Mountford, P.S. and **O'Brien, C.M.** (1997). IRES vectors: new transgenic tools for investigating early mammalian development. *41st annual meeting of the ASBMB,* Melbourne, Australia. *Invited speaker.*

Abbreviations

A-P axis	anterioposterior axis
APC	anaphase-promoting complex
ATG	initiation of translation codon
BAC	bacterial artificial chromosome
βgal	β-galactosidase
βgeo	*lacZ* and *neo* fusion
βNGF	β nerve growth factor
bp	base pairs
BMP	bone morphogenic protein
BSA	bovine serum albumin
Cdk	cyclin-dependent kinase
cDNA	complementary DNA
Cer-1	*Cerberus-related 1* gene
Ci	Curies
CNTF	ciliary neurotrophic factor
CsCl	caesium chloride
CSF-1	colony-stimulating factor-1
CSL	Commonwealth Serum Laboratories
CT-1	cardiotrophin-1
ddH$_2$O	deionized distilled water
DEPC	diethyl-pyrocarbonate
DIA	differentiation inhibiting factor (see LIF)
dNTPs	deoxynucleotide triphosphates (dATP, dTTP, dGTP, dCTP)
DMEM	Dulbecco's modified Eagle's medium
DMSO	dimethylsulphoxide
DNA	deoxyribonucleic acid
DNase	deoxyribonuclease
dpc	days post coitum
dsDNA	double stranded DNA
DTT	dithiothreitol
EC cells	embryonal carcinoma cells
EDTA	ethylenediaminetetra-acetic acid
EG cells	embryonic germ cells
EGF	epidermal growth factor
EMCV	encephalomyocarditis virus
En-2	*Engrailed-2* gene
ENU	*N*-ethyl-*N*-nitrosourea
EPL cell	primitive ectoderm-like cell
ERK	extracellular regulated kinase
ES cells	embryonic stem cells
ESRF	embryonic stem cell renewal factor
EST	expressed sequence tag
FCS	fetal calf serum
FGF	fibroblast growth factor
FGF2	fibroblast growth factor-2 (also known as basic FGF, bFGF)
FGF4	fibroblast growth factor-4
g	acceleration due to gravity (9.8 meters per sec)

gb	Genbank accession number
G-CSF-R	granulocyte colony-stimulating factor receptor
GMEM	Glasgow minimum essential medium
Gsc	*Goosecoid* gene
GT	gene trap
h	hours
HGF	hepatocyte growth factor
HNF-3β	hepatocyte nuclear factor 3β
HSV*tk*	herpes simplex virus *thymidine kinase* gene
ICM	inner cell mass
IGFII	insulin-like growth factor II
IL-6	interleukin-6
IRES	internal ribosome entry site
JAK	Janus kinase
kb	kilobase pairs
kD	kiloDalton
Klenow	large fragment of DNA polymerase 1
lacZ	*E. coli* gene encoding β-galactosidase
L1 elements	LINE-1 elements
LB medium	Luria-Bertani medium
LIF	leukemia inhibitory factor (see DIA)
LIF-R	LIF-receptor
MAb	monoclonal antibody
MAPK	mitogen-activated protein kinase
min	minutes
MIRD	Monash Institute of Reproduction and Development
MOPS	3-(morpholino)propanesulphonic acid
mRNA	messenger ribonucleic acid
Mw	molecular weight
neo	*E. coli neomycin phosphotransferase* gene
NP40	nonidet P-40
nt	nucleotide
OD	optical density
o/n	overnight
OPT	optimized
OSM	oncostatin-M
P1	phage clone
PBS	phosphate-buffered saline
PCR	polymerase chain reaction
PDGF-A	platelet-derived growth factor A
PGCs	primordial germ cells
poly A	polyadenylation
PNK	polynucleotide kinase
RACE	rapid amplification of cDNA ends (5'- and 3'-)
RBP-Jκ	J kappa recombination signal sequence binding protein
rpm	revolutions per minute
RNA	ribonucleic acid
RNase	ribonuclease
RT	reverse transcription
RT^{0}C	room temperature

s	seconds
SA	splice acceptor
SCF	Skp1-Cdc53/cullin-F-box protein
SDS	sodium dodecyl sulphate
SH2	Src-homology-2
Sl	*Steel* gene
SSLP	simple sequence length polymorphism
STAT	signal transducer and activator of transcription
Su(H)	*Drosophila suppressor of hairless* gene (homolog to *RBP-J$_k$*)
SV40	simian virus 40
T	*Brachyury* gene
Taq	*Thermus aquaticus* bacterium
Tbn	*Taube nuss* gene
TE	trophectoderm
TE buffer	10mM Tris, 1mM EDTA (pH 8.0)
TEMED	N,N,N',N'-tetraethylmethylenediamine
TGF-α	transforming growth factor-α
TGF-β	transforming growth factor-β
TMAC	tetramethylammonium chloride
Tris	tris(hydroxymethyl)aminomethane
TS cells	trophoblastic stem cells
TSS	transformation and storage solution
UTR	untranslated region
UV	ultraviolet light
VCB	VHL/Elongins C/B complex
VE	visceral endoderm
VHL	von Hippel-Lindau (tumor suppressor)
v/v	volume for volume
W	*Dominant white spotting* gene
WEHI	Walter and Eliza Hall Institute
Wnt	Wingless vertebrate protein
w/v	weight for volume
X-gal	5-bromo-4 chloro-3-indolyl-β-D-galactoside
YAC	yeast artificial chromosome

Nomenclature

Gene symbols are always italicized with all letters in lower case, excepting the first letter for mammalian genes. Protein products of genes are in Roman type, and abbreviated protein families are in upper case.

List of Figures

List of Tables

Chapter One

INTRODUCTION

1.1 GENERAL INTRODUCTION

Research efforts in developmental biology are ultimately directed at understanding the precise and highly orchestrated mechanisms that regulate the biological transition from a single cell, the fertilized egg, to a complex and highly organized multicellular adult organism. This increase in tissue complexity and cellular specialization is accompanied at the cellular level by a general loss in cellular potency. That is, cells become committed and then restricted to increasingly specialized cell types.

Pluripotential cells have the capacity to give rise to the three embryonic germ layers of the embryo proper and the germ cell lineage, as well as the extraembryonic tissues that support embryonic development. Pluripotentiality in mammals is restricted to the oocyte, the zygote, early embryonic cells and primordial germ cells. Self-renewal, differentiation and commitment of cells to a specific lineage during early embryogenesis are likely to be under the control of a network of transcription factors, which regulate gene expression in a cell-type, stage-specific manner. While the molecular mechanisms underlying such a network are only beginning to be understood, recent technical advances in the study of mammalian development are greatly facilitating the elucidation of these highly complex events.

1

Compared with other vertebrate animal systems such as *Xenopus laevis*, chick or the zebrafish *Danio rerio*, the study of mammalian development has been made more difficult by the larger size of the genome, slower development of the embryo, and increased difficulty in observation and manipulation of the embryos. The house mouse, *Mus muscularis*, has become the representative mammal for developmental research for reasons including well characterized genetics, the availability of many inbred lines and mutant strains, and the relatively short gestation period of approximately 20 days, (Hogan *et al.*, 1994). Since large-scale screens for mutants are less feasible in the mouse, genetic analysis of development has also been on a much slower scale than that for the invertebrates *Drosophila melanogaster* and *Caenorhabditis elegans* (reviewed by Rossant and Hopkins, 1992). However, the ability to introduce new genetic information into the mouse embryonic genome using a number of strategies has given rise to transgenic mouse models, a valuable tool for the identification and functional analysis of developmentally important genes (reviewed by Jaenisch, 1988).

The *in vitro* isolation and culture of embryonic stem (ES) cells from mice (Evans and Kaufman, 1981; Martin, 1981), primates (Thomson *et al.*, 1995, 1996), and more recently from humans (Thomson *et al.*, 1998) has heralded a new age in mammalian developmental research. ES cells provide a source of pluripotential cells which can, under defined conditions, be propagated indefinitely *in vitro* whilst retaining the ability to give rise to potentially any cell type of the adult organism.

Mouse ES cells are amenable to genetic manipulation and can be reintroduced to a host blastocyst where they are able to contribute to all somatic lineages and the germ cells of the resulting embryo (Gossler *et al.*, 1986; Robertson *et al.*, 1986). A wide range of predetermined modifications can now be introduced into mice by homologous recombination in ES cells (Thomas and Capecchi, 1987). More recently, entrapment vectors have come into use allowing the random introduction of a reporter gene tag to identify endogenous genes of interest on the basis of reporter gene expression profiles (Gossler *et al.*, 1989). ES cells provide an *in vitro* system for screening targeted and random mutation events prior to generating mouse lines. The opportunity to identify new genes and investigate gene function in an *in vitro* system provides a particularly powerful vehicle for the investigation of early developmental regulation.

To place the experimental objective of this study in context, this review will begin with an overview of early mouse embryogenesis and some of the molecular mechanisms

2

implicated in regulating early development. The derivation of mammalian pluripotential cell lines, and their considerable potential in both basic research and cell-based human therapies will be discussed. Techniques for modification of the mouse genome will be outlined with an emphasis on the use of entrapment vectors in ES cells for the identification and mutational analysis of developmentally regulated genes.

1.2 EARLY EMBRYOGENESIS OF THE MOUSE

Mouse embryonic development takes approximately 20 days and occurs in three stages of cleavage, gastrulation and organogenesis. Unlike embryonic development in organisms such as the sea urchin, *Drosophila* and *Xenopus*, mouse embryonic development is characterized by a slow growth phase during the first 4.5 days following fertilization, as the embryo travels down the oviduct to the uterus. Prior to intrauterine development, the embryo undergoes a series of cell divisions and generates the first two differentiated lineages (the trophectoderm and the inner cell mass) without an increase in overall mass. The rate of cell division and differentiation increases dramatically following implantation and the establishment of intrauterine connections to the mother's nutrient supply, particularly in the group of pluripotential cells (the primitive ectoderm or epiblast) that will form the embryo proper. By the tenth day following fertilization, the epiblast has expanded and given rise to the three primary germ layers - endoderm, ectoderm and mesoderm - in a process termed gastrulation. This process lays down the basic body plan of the future organism, while the final phases of development involve continued differentiation of organs and fetal growth. A time course for mouse development following fertilization is shown in Figure 1.1.

As this study is primarily concerned with the mechanisms controlling the pluripotentiality of preimplantation stage embryonic cells and their commitment following implantation to the somatic and germ cell lineages, the following discussion is an overview of these developmental events up to the end of gastrulation and early organogenesis.

1.2.1 Cleavage of the fertilized egg

The mature ovulated egg is arrested at metaphase II of the second meiotic division. It is about 85 µm in diameter and is surrounded by an acellular matrix, the zona pellucida, which mediates the events leading to fusion of the sperm and egg plasma membranes.
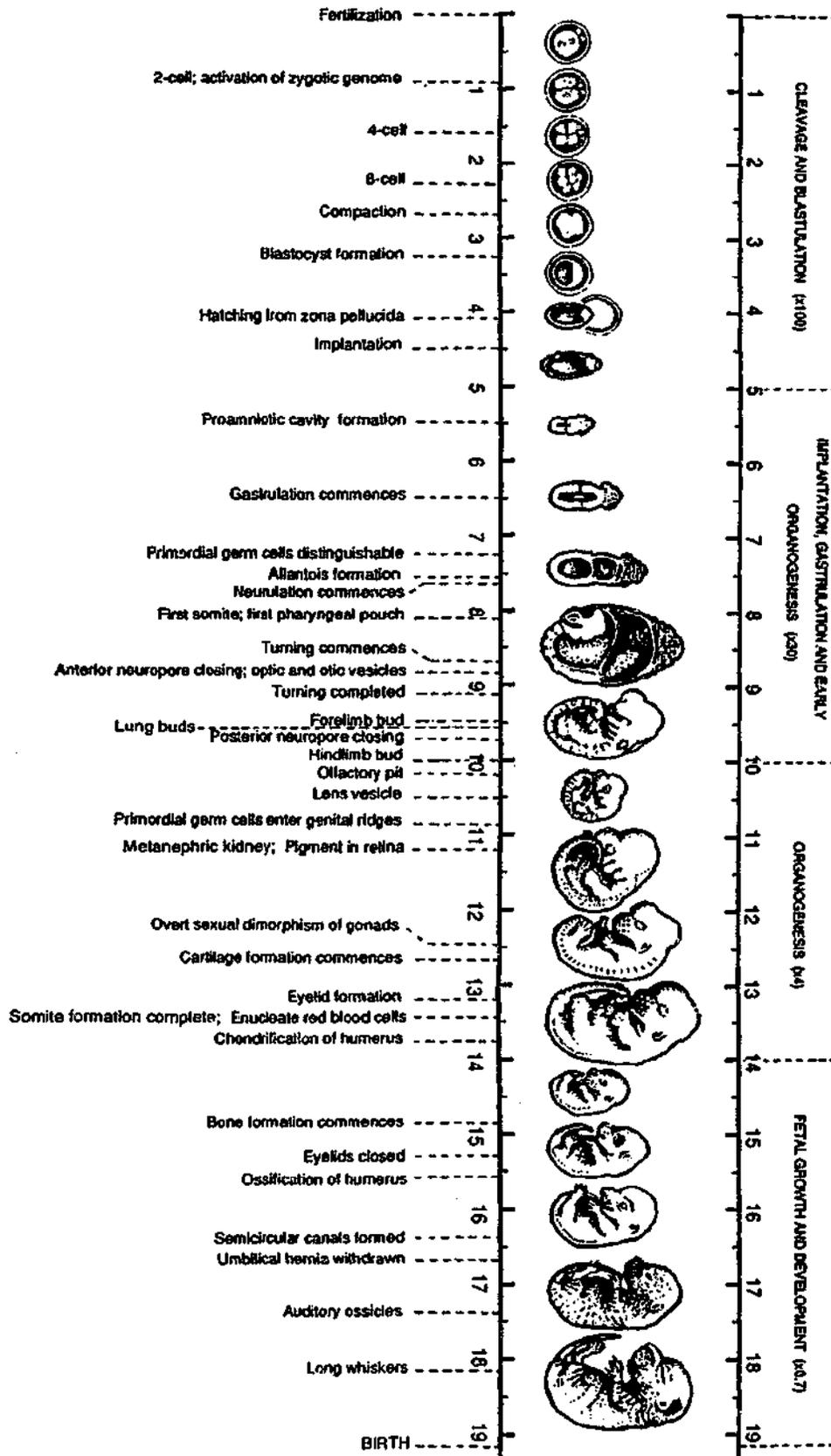
Figure 1.1: A time course for mouse development. From 0-5 days the fertilized egg undergoes cleavage and blastulation; between days 5-10 the embryo implants then undergoes gastrulation and early organogenesis; further organogenesis takes place from days 10-14, followed by a final period of fetal growth and development until birth at 19-20 days (taken from Hogan *et al.*, 1994).

ZP3, one of three sulfated glycoproteins that make up the zona pellucida, has been demonstrated in the mouse to be the primary receptor for sperm binding and is responsible for induction of the acrosome reaction (Wassarman *et al*, 1986). Fertilization takes place in the oviduct, at which time the second meiotic division of the egg is completed.

Embryonic development starts at fertilization, except in the case of spontaneous parthenogenetic activation of eggs, which occurs in about 10% of the oocytes of the LT/Sv strain of mice (Hogan *et al.*, 1994). Since both maternal and paternal genomes are required for normal development, parthenogenetic embryos that implant do not develop much beyond the egg cylinder stage at 7 days post coitum (dpc), (McGrath and Solter, 1984; Surani *et al.*, 1984). Cleavage (Figure 1.2A) takes place in the oviduct, the first division occurring about 24 hours (h) after fertilization, with subsequent blastomere cleavages at about 10-12 h intervals resulting in a solid ball of cells, the morula. This is a much slower cleavage rate than for lower vertebrates or invertebrates which begin rapid cycles of division soon after formation of the diploid zygote nucleus (Bowman and McLaren, 1970). Up to the mid 2-cell stage (27 h postfertilization) the embryo appears to rely on maternal mRNA and proteins synthesized during oogenesis (reviewed by Schultz, 1986; Kidder, 1992), whereas *de novo* transcription of mRNAs from the embryonic genome is required for development beyond the 2-cell stage (Flach *et al.*, 1982), at which time the inherited mRNA is rapidly degraded (Brinster *et al.*, 1980).

Cleavages in the early embryo are not preceded by cellular growth, such that there is a successive decrease in the cytoplasmic to nuclear ratio of resulting blastomeres. Up to the 8-cell (morula) stage, the spherical blastomeres are totipotent and equipotent. It has been demonstrated that single blastomeres dissociated from a 4-cell embryo can give rise to a new individual. While early 8-cell stage blastomeres will not generate a new mouse by themselves, they are able to give rise to a wide range of differentiated tissues in chimaeric offspring (Kelly, 1977; reviewed by Pederson, 1986). As the embryo cleaves onwards from the 8-cell stage, there is a gradual restriction in the developmental potency of the cells.

1.2.2 Compaction

At the late 8-cell stage, the embryo changes shape for the first time in a process known as compaction (Figure 1.2A). The morula becomes tightly aggregated as the blastomeres progressively flatten to increase cell surface in contact with each other, and take on a polarity which gives the cells distinct apical and basal membranes as well as cytoplasmic domains. Compaction results in the formation of inner and outer cells connected by tight intercellular gap junctions such that individual blastomeres are no longer discernable. The exterior surfaces of the cells carry microvilli, whereas the inner surfaces are smooth (Reeve and Ziomek, 1981), and changes in cell-adhesion molecules are evident. For example, the transmembrane cadherin uvomorulin is initially synthesized around the 4-cell stage and uniformly distributed on the cell surfaces. With morula compaction the protein becomes distributed in clusters at basolateral cell contact sites (Vestweber *et al.*, 1987; reviewed by Kemler, 1993; Peifer, 1993). Changes associated with compaction are both radial and tangential, and do not occur synchronously in all cells of the morula with further cleavages being somewhat variable. At the 32-cell stage the morula contains about 10 internal cells and more than 20 outer cells (Wolpert *et al.*, 1998).

The changes occurring with compaction lead to the first differentiation event in the preimplantation embryo, resulting in the formation of two distinct cell groups. From the 16-cell stage the outermost cells give rise to the trophectoderm (TE) while the innermost cells form the inner cell mass (ICM), (Johnson and Ziomek, 1981). The TE cells form a discrete outer layer which will eventually give rise to the extraembryonic structures that support development of the embryo proper arising from a population of the ICM cells.

1.2.3 Blastocyst development

At 3.5 days gestation, when the compacted morula contains about 20-30 cells, the TE cells begin to osmotically pump fluid into the center of the embryo via a Na+/K+-ATPase mechanism. This pushes out the cells to form a hollow fluid-filled cavity, the blastocoel. Further cell division generates the fully expanded blastocyst stage embryo (Figure 1.2A) which consists of a single peripheral layer of flattened TE cells surrounding a group of about 20 pluripotential ICM cells, located asymmetrically within the expanded blastocoel cavity (Gardner and Papaioannou, 1975). The TE cells are

5

Figure 1.2: Development of the mouse embryo from cleavage to gastrulation. (A) The fertilized mouse egg cleaves from the 2-cell stage through to formation of a hollow blastocyst. Following the 8-cell stage the developing morula undergoes compaction, whereby cells take on a polarity and become tightly aggregated. In the formation of the blastocyst the internal cells give rise to the inner cell mass and the outer cells to the trophectoderm layer. (B) The embryo implants on the fifth day following fertilization at which time the inner cell mass divides into the primitive ectoderm (or epiblast) that will give rise to all three germ layers of the embryo proper, and the primitive endoderm, which will contribute to extraembryonic (extra-embryonic) structures. The polar trophectoderm in contact with the epiblast forms extra-embryonic tissues, the ectoplacental cone, and extra-embryonic ectoderm which contributes to the placenta. The mural trophectoderm gives rise to the trophoblast giant cells. At 5.5 dpc the epiblast elongates and develops an internal proamniotic cavity giving it a cup-shaped form. At 6.0 dpc the cylindrical structure containing both the epiblast and the extra-embryonic tissue derived from the polar trophectoderm is known as the egg cylinder. The onset of gastrulation takes place at 6.5 dpc with the appearance of the primitive streak at the posterior end of the epiblast and extends anteriorly to the tip of the cylinder. Cells in the distal epiblast expand to occupy most parts of the embryonic ectoderm, while proliferating proximal epiblast cells pass through the streak to become mesoderm and definitive endoderm. As the streak fully extends the node arises at the anterior end, and gives rise to the notochord mesoderm along the midline, while extra-embryonic mesoderm arises at the posterior end, which contributes to the amnion, visceral yolk sac, and the allantois and chorion. P=posterior; A=anterior; (adapted from Wolpert *et al.*, 1998).

A

| Two-cell | Four-cell | Eight-cell | Compacted morula | Blastocyst |

B

Blastocyst at time of implantation

polar trophectoderm

epiblast

primitive endoderm

mural trophectoderm

Implanted embryo 5½ days

Extra-embryonic

epiblast

parietal endoderm

trophoblast giant cells

Embryonic

Egg cylinder 6 days

ectoplacental cone

extra-embryonic ectoderm

proamniotic cavity

visceral endoderm

Onset of gastrulation 6½ days

extra-embryonic mesoderm

primitive streak

A   P

embryonic mesoderm

Full length primitive streak 7½ days

Extra-embryonic

chorion

amnion

gut endoderm

notochord

Embryonic

exocoelom

allantois

primitive streak

node

specialized to form an intact epithelium and are involved in implantation of the embryo. Those TE cells in direct contact with the ICM are termed polar TE, while the remaining TE cells lining the cavity make up the mural TE. Importantly, the TE will only contribute to extraembryonic, but not to embryonic tissues during development.

In the final phase of preimplantation development at 4 days gestation, when there are 20-40 cells in the ICM, the embryo undergoes a second differentiation event (reviewed by Gardner, 1983). The ICM surface cells in contact with the blastocoel differentiate into an epithelial layer of primitive endoderm which will later contribute only to extraembryonic membranes, but performs vital roles for the development of the early embryo (see section 1.2.4). The remaining core of ICM cells is termed the primitive ectoderm or epiblast. These cells remain in an undifferentiated state until after implantation when they will give rise to the ectodermal layer of the amnion and, following gastrulation, to the three germ layers of the embryo proper, to the germ cells and the mesodermal components of the extraembryonic membranes (Gardner and Rossant, 1979; reviewed by Beddington, 1983). The pluripotential properties of the epiblast cells prior to gastrulation will be further discussed in section 1.4.

1.2.4 Implantation

During the fifth day of development, the embryo releases itself from the zona pellucida and implants into the uterine wall, after which it undergoes a massive increase in growth. This hatching process may be effected by the action of a trypsin-like enzyme synthesized in the mural TE cells on the zona pellucida glycoproteins (Wassarman *et al.*, 1984). Rhythmic expansion/contraction of the blastocyst may also play a role, as well as uterine enzymes *in vivo*, although hatching will normally occur *in vitro* (Hogan *et al.*, 1994). At the same time, the maternal uterine tissue is primed to receive the embryo by the high levels of ovarian estrogen circulating during estrous, followed by a few days of post-ovulation progesterone secretion by corpora lutea and a small surge of estrogen on the fourth day of gestation. Concurrent with this final estrogen surge and the onset of blastocyst implantation is a surge in uterine expression of the cytokine leukemia inhibitory factor (LIF) or differentiating inhibiting activity (DIA), (Bhatt *et al.*, 1992; Smith *et al.*, 1992) which appears to be necessary for the maternal initiation of implantation.

6

At the time of implantation (Figure 1.2B), mural TE cells at the abembryonic pole (opposing the ICM) replicate their DNA without cell division, giving rise to the trophoblastic giant cells. These are the first cells to make contact with maternal tissue (Gardner and Johnson, 1972) as the blastocyst first adheres to the anti-mesometrial uterine wall. This attachment induces the formation of a uterine crypt and triggers a decidual reaction, whereby stromal (decidual) cells undergo an epithelial transition and proliferate to form a spongy mass of cells, the "deciduum", around the embryo. In due course, the epithelium between the blastocyst and the stroma is eroded and the trophoblastic giant cells invade the deciduum via the secretion of the protease urokinase-type plasminogen activator, as well as various metalloproteinases and their inhibitors (Strickland and Richards, 1992).

The cells of the polar TE maintain a proliferative capacity following implantation (Rossant et al., 1978) and grow to form two extraembryonic tissues, the extraembryonic ectoderm and the ectoplacental cone, which contribute to placental formation as well as forming secondary giant cells (Rossant, 1986). The extraembryonic ectoderm pushes the epiblast down into the blastocoel cavity in a finger-like projection that develops an internal proamniotic cavity, forming the early egg cylinder, or conceptus (Figure 1.2B), (Copp, 1978, 1979). Later, when the epiblast has given rise to the extraembryonic mesoderm (see section 1.2.5), the extraembryonic ectoderm fuses with the latter forming the chorion which envelopes the entire conceptus and the other three fetal membranes, the amnion, yolk sac and allantois. The ectoplacental cone arises from the growth of polar TE into the uterine crypt, forming the bulk of the placenta, and some of these cells also become secondary giant cells. Interestingly, trophoblastic stem (TS) cell lines have been recently derived from early postimplantation extraembryonic ectoderm, as well as from blastocyst TE, and these can be maintained in vitro on embryonic fibroblast feeder cells in the presence of fibroblast growth factor-4 (FGF4) and heparin. In the absence of any of these requirements, TS cells will terminally differentiate to giant cells, but are otherwise capable of differentiating exclusively to trophoblastic subtypes, both in vitro and in vivo (Tanaka et al., 1998).

Concurrent with the growth of the TE-derived tissues, some of the primitive endodermal cells migrate to cover the whole inner surface of the mural TE, forming the parietal endoderm. These cells, together with the trophoblastic giant cells, form the parietal yolk sac and are responsible for laying down the thick Reichert's basement membrane

7

that encapsulates the yolk sac cavity. The remaining primitive endodermal cells form the visceral endoderm (VE) that covers the elongating egg cylinder and later forms the visceral yolk sac with extraembryonic mesoderm (Gardner, 1983). VE cells are polarized and specialized for metabolic exchanges between the maternal circulation and the yolk sac, as well as for synthesis and secretion of proteins needed by the embryo such as α-fetoprotein, transferrin and apolipoproteins. In recent years, however, it has come to light that the VE also plays an important role in the complex myriad of molecular events regulating early embryonic development and is the site where anterioposterior (A-P) gene expression is first conferred, even prior to gastrulation (reviewed by Rossant, 1995, Beddington and Robertson, 1998). The VE has been proposed as the source of a "death signal" for programmed cell death in a model for the cavitation of the epiblast during the formation of the early egg cylinder, and may involve expression of the transforming growth factor *Bmp-2* (Coucouvanis and Martin, 1995, 1999). Some of the genes expressed in the VE and implicated in early embryogenesis will be further discussed in section 1.3.

From implantation to gastrulation the epiblast cells undergo subtle modifications in their surface polysaccharides and protein synthesis. X-inactivation takes place and cell cycle length decreases rapidly between 5.5 and 7.5 dpc (Snow, 1977; Gardner and Beddington, 1988). There is also increased methylation of the genome following the demethylation of CpG islands associated with genomic imprinting in the cleavage stage embryo (Monk *et al.*, 1987). By 6.5 days postfertilization, the cavitated epiblast lies as a cup-shaped epithelium containing about 700 cells (Snow, 1977), from which the entire fetal soma and the germline will arise. Evidence for this is largely derived from studies assessing the fate of 4.5-7.5 dpc epiblast single cells or tissue when injected into host blastocysts of a different genotype (Gardner and Rossant, 1979; Gardner *et al.*, 1985), injected to ectopic sites in adult mice (Stevens, 1970), or injected with lineage tracer to an intact embryo (Lawson *et al.*, 1991).

1.2.5 Gastrulation and early organogenesis

Gastrulation is the process whereby the developing embryo is transformed from a simple sheet of epithelium into a complex multilayered structure (Figure 1.2B). The first visible sign of the embryo's future axis is at about 6.5 dpc, when gastrulation begins with the formation of the primitive streak, an invagination within the cup-shaped

8

primitive ectoderm. The primitive streak originates as a localized thickening adjacent to the junction of the embryonic and extraembryonic ectoderm, as primitive ectoderm cells rapidly proliferate and extensively migrate to what will be the future posterior end of the embryo (Lawson *et al.*, 1991). Cells in the distal epiblast expand to occupy most parts of the embryonic ectoderm which will give rise to the neuroectoderm and surface ectoderm, while proliferating cells of the proximal epiblast delaminate as they ingress through the streak, spreading out laterally and anteriorly between the ectoderm and the VE to form a new mesodermal layer.

Orthotopic and heterotopic transplantation experiments have demonstrated that the proximal and distal epiblast cells of pre- and early-primitive streak stage embryos are not irreversibly allocated to any specific lineages, including the germline, and will adopt the developmental fate that is typical of the cell population at the site of transplantation (Tam and Zhou, 1996). This suggests that allocation of epiblast cells to the specific lineages is subject to certain site-specific influences in the epiblast during gastrulation, and may be governed by the morphogenic options available to specific groups of epiblast cells as ingression through the streak takes place.

As the primitive streak elongates anteriorly a condensation of cells at its anterior end becomes apparent. This structure, known as the node, shows overlapping gene expression profiles and properties of Hensen's node in the chick and the dorsal lip (Spemann's organizer) in *Xenopus* (Blum *et al.*, 1992). Cells moving through the node early in gastrulation move anteriorly along the midline to form the mesoderm of the head process. As the primitive streak regresses, the node extends posteriorly and the cells moving through the node form a cord of mesoderm, the notochord, along the midline. The notochord is a transient organ that not only reflects the A-P axis of the embryo, but also acts as an organizer of adjacent embryonic structures (Nicolet, 1971; Placzek *et al.*, 1990). Those cells migrating laterally from the primitive streak form bands of mesoderm running longitudinally along each side of the notochord, extending posteriorly to displace the extraembryonic ectoderm. Some of the cells moving through the anterior node, as well as cells delaminating directly from the primitive ectoderm through the mesoderm, establish the definitive endoderm as they enter and gradually displace the VE, forming the outermost epithelial layer beneath the regressing primitive streak (Lawson *et al.*, 1991; Tam and Beddington, 1992).

9

With the three germ layers established, a series of morphogenic changes then take place to create specific tissues and organs. In brief, early organogenesis begins with neurectoderm formation in the midline and anterior ectoderm in response to signals from the underlying notochord and head process (Placzek *et al.,* 1990). This neural plate forms folds at around 8.5 dpc and finally closes dorsally to develop into a neural tube, the precursor of the central nervous system. The ectodermal cells covering the embryo after neurulation form the epidermal layer of the skin. As the neural folds begin to form at the center of the embryo and the primitive streak regresses, paraxial mesoderm cells condense into paired cylindrical blocks, somites, which later form vertebrae, muscles and the dermis of the skin (Tam and Meier, 1982). The neural tube and somites are established simultaneously and progressively in an anterior to posterior direction. Intermediate mesoderm cells develop into the kidney and genital ridges, while mesoderm cells lateral to this give rise to a variety of tissues including the lining of the body cavities, the heart, blood vessels and blood cells of the circulatory system, as well as mesodermal components of the limbs, apart from the muscles. The mesoderm cells that have become posteriorly extraembryonic contribute to formation of the visceral yolk sac, amnion, chorion and allantoic extraembryonic membranes and also give rise to blood islands, and eventually the blood vessels of the yolk sac (Haar and Ackerman, 1971; Hogan *et al.,* 1994).

The definitive endoderm forms the primitive gut, which gives rise to the linings of the digestive system and respiratory tract. The primitive gut initially forms in two pockets, the foregut and the hindgut, as it pinches towards the center of the embryo. The hindgut region gives rise to the tonsils, thymus and thyroid glands as well as to the lungs, respiratory tubes and the intestines. The oesophagus, then the stomach form in the foregut region, with buds from the digestive tube forming the lining of the liver, pancreas and gall bladder. The pituitary gland is formed from an interaction between outer epidermal cells at the oral opening and the brain ectoderm.

The germ cell lineage is derived from the primitive ectoderm (Falconer and Avery, 1978; Gardner *et al.,* 1985). Clonal lineage studies have identified the precursors of primordial germ cells (PGCs) within the proximal margin of the 6.5 dpc epiblast, close to the extraembryonic ectoderm (Lawson and Hage, 1994). PGCs are first detected during gastrulation at 7.0 dpc as a cluster of cells expressing tissue non-specific alkaline phosphatase (Hahnel, 1990) within the extraembryonic mesoderm near the posterior

part of the primitive streak (Chiquoine, 1954; Ginsburg *et al.*, 1990). It has been suggested that this localization is strategically important for isolating pluripotential cells destined for the germ lineage from the forces driving somatic cell differentiation and tissue patterning (Dixon, 1994). This proposal is supported by *in vitro* models demonstrating that the removal of differentiated cells by microsurgically isolating the epiblast from mouse blastocysts (Brook and Gardner, 1997), or via mouse ES cell selection strategies (McWhir *et al.*, 1996; Mountford *et al.*, 1998), is advantageous for the *de novo* isolation and maintenance of pluripotential ES cell cultures (see section 1.4). In the previously discussed transplantation experiments for pre- and early-primitive streak stage embryos (Tam and Zhou, 1996), the finding that distal epiblast cells could form PGCs after transplantation to proximal epiblast supports the concept that specification of the germline cells takes place amongst the precursors of the extraembryonic mesoderm.

PGCs become incorporated into the base of the allantois from where they migrate along the hindgut endoderm, arriving at the developing genital ridges by 10.5 dpc (reviewed by Buehr, 1997). The PGC population contains approximately 150 cells at 8.5 dpc and divides about once every 16 h during this migration, such that at 12.5 dpc each gonad primordium is colonized by about 26,000 PGCs (Tam and Snow, 1981). By 13.5 dpc the PGCs within the ridge cease dividing and aggregate with each other and somatic cells of the ridges forming the primitive sex cords. The female germ cells enter meiosis and those in the male undergo mitotic arrest (Ginsburg *et al.*, 1990).

At the early somite stage when gastrulation is in its final stages, the embryo undergoes an episode of complex turning such that the endoderm (gut) is internalized, the ectoderm (neural tube and surface ectoderm) lies on the outside, and the embryo is surrounded by its embryonic and extraembryonic membranes (Kaufman, 1992). Gastrulation is complete by the 10th day of gestation; the primitive streak is replaced by the tail bud and the posterior neuropore has closed. The embryo now has a distinct head and the forelimb buds are starting to develop. This is followed by the final stages of organogenesis, then growth and maturation of the embryo until birth.

1.3   GENETIC CONTROL OF EARLY MAMMALIAN DEVELOPMENT

Large-scale mutational screening in *Drosophila* (see section 1.5) has repeatedly demonstrated that regulatory proteins, mainly DNA-binding transcription factors,

determine the fate of cells during development by modulating the transcription of certain genes at restricted stages during development (Nüsslein-Volhard and Wieschaus, 1980; reviewed by Akam, 1987; Dressler and Gruss, 1988). Homology searches against many genes identified to be important for controlling development of invertebrates and lower vertebrates has led to the identification of highly conserved gene families with related functional motifs being expressed during mammalian embryogenesis (reviewed by Kessel and Gruss, 1990; Wolpert et al., 1998). Examples of this include identification of transcription factors with related DNA-binding domains such as the homeo, paired, zinc finger, winged-helix and helix-loop-helix domains. Growth factor and oncogene related protein signaling molecules such as those related to the fibroblast growth factor (FGF), transforming growth factor-β (TGF-β), epidermal growth factor (EGF), hedgehog and wingless (Wnt) families, as well as transmembrane receptors for different classes of signaling molecules, are further examples of the fruits of homology searching for related mouse genes. The availability of ES cell lines (see section 1.4) has overcome some of the difficulty in obtaining sufficient material from mammalian eggs and embryos for experimentation, and expediated the cloning of many developmentally regulated murine genes. The following discussion presents some of the genes currently implicated to be of importance in early mouse embryogenesis and gastrulation.

1.3.1 Preimplantation development

Most murine developmental control genes identified by sequence similarity to *Drosophila* genes are expressed at the postimplantation stage. While the fate of cells and the morphogenic changes associated with preimplantation development are well established, relatively little is known about the genes controlling these events. Since the mammalian embryo depends largely on its own genomic transcription for mRNA and protein sources during cleavage, it is likely that sets of transcription factors in use are unique from those utilized in lower vertebrates and invertebates.

RNA transcripts, and often corresponding proteins, have been detected in the preimplantation embryo for a number of growth factors which may act to promote cell proliferation and/or survival. Maternal transcripts for platelet-derived growth factor A (PDGF-A) and transforming growth factor-α (TGF-α) are present in the unfertilized, ovulated egg and decline during the 2-cell stage of cleavage. In the late cleavage-early blastocyst stage, however, zygotic transcripts for these factors appear. TGF-β1

12

transcripts are not present in the egg but are increased during cleavage (Rappolee *et al.*, 1988). Transcripts have also been detected during preimplantation development for insulin-like growth factor II (IGFII), IGFII receptor, IGFI receptor, insulin receptor and EGF receptor (Kidder, 1992; Rappolee *et al.*, 1992; Wiley *et al.*, 1992).

Two members of the mammalian POU domain family of transcription factors, *Oct-4* (also called *Oct-3* and *NF-A3*) and *Oct-6* are expressed in the preimplantation embryo and are strongly implicated in the regulation of gene expression during this stage of development (Schöler *et al.*, 1989a, 1989b; Meijer *et al.*, 1990; Okomoto *et al.*, 1990; Rosner *et al.*, 1990; Schöler *et al.*, 1990). These genes encode for transcription factors that share a conserved POU homeodomain and bind the octamer motif ATGCAAAT, although the Oct-4 protein differs from all other members of the POU domain family at several conserved positions within the domain. Several of the developmental control genes that are expressed after *Oct-4* and *Oct-6*, such as *Hox 1.3*, harbor octamer motif(s) in their promoter regions and thus might be regulated by these transcription factors (reviewed by Brehm *et al.*, 1990; Schöler, 1991). *Oct-4* expression is found in the unfertilized egg as well in the pluripotent cells of the pregastrulation embryo. It is downregulated during differentiation of these cells to mesoderm and endoderm, becoming restricted to the PGCs and female germ cells (Rosner *et al.*, 1990; Schöler *et al.*, 1990; Yeom *et al.*, 1996; Pesce *et al.*, 1998). *Oct-6* expression is also found in the undifferentiated cells of the blastocyst, as well as in the testis and in specific neurons of the developing and adult brain (Schöler *et al.*, 1989b; Meijer *et al.*, 1990).

Oct-4 interacts with a series of partners including a stem cell-restricted E1A-like activity (Schöler *et al.*, 1991), the HMG-box protein Sox-2 (Ambrosetti *et al.*, 1997; Nishimoto *et al.*, 1999), and the transcription factor Rox-1 (Ben-Shushan *et al.*, 1998) to activate the transcription of specific genes, with the amount of *Oct-4* relative to these partners appearing to be crucial (Schöler *et al.*, 1991). The gene encoding FGF4 is a candidate target gene of *Oct-4* during early embryogenesis. *Fgf-4* has an octamer-containing enhancer in its 3'-noncoding region and responds to *Oct-4* in synergy with *Sox-2* (Yuan *et al.*, 1996; Ambrosetti *et al.*, 1997). It has been recently demonstrated through the use of targeted gene deletion (see section 1.5.3), that *Oct-4* activity is crucial for the maintenance, and probably establishment, of the pluripotential cell population in the ICM (Nichols *et al.*, 1998). *Oct-4* deficient embryos will develop to blastocyst stage, but the ICM cells are not pluripotential and differentiate only along the

trephoblast lineage. Furthermore, the proliferati⸍e capacity of the trophoblast is not maintained in the absence of a true ICM. In this study, the restoration of trophoblastic precursor cells by culturing *Oct-4* mutant embryos in the presence of recombinant FGF4 also demonstrates that *Oct-4* determines paracrine growth factor signaling from the pluripotent cells to the trophectoderm.

A recent publication now reports the identification of a highly conserved sequence for a novel gene *Taube nuss* (*Tbn*), which is crucial for the survival of the ICM cells in the 3.5 dpc blastocyst stage embryo. *Tbn* deficient embryos generated by a gene trap insertion (see section 1.6.2) fail to develop beyond 4.0 dpc as a result of apoptosis occurring in the ICM cells (Voss *et al.*, 2000). A functional relationship between *Oct-4* and *Tbn* may not exist, however, as *Tbn* 3.5 dpc mutant embryos express Oct-4 protein normally, and Tbn protein is expressed in *Oct-4* mutants. *Tbn* mutant embryos also express FGF4 and, in contrast to *Oct-4* mutants (Nichols *et al*, 1998), do not survive when cultured in media supplemented with recombinant FGF4 (Voss *et al.*, 2000).

1.3.2 Postimplantation and gastrulation

FGF4 has also been demonstrated to be essential for postimplantation mouse development. Embryos homozyogous for a null mutation of the *Fgf-4* gene are able to implant in the uterus and induce a decidual reaction, but are not able to develop thereafter. When cultured *in vitro*, *Fgf-4* null embryo⸍ display severely impaired proliferation of the ICM, which can be rescued by the addition of FGF4 protein (Feldman *et al.*, 1995). That *Tbn* mutant ICMs cannot proliferate, but die under the same conditions, suggests that either FGF4 is essential for the embryo proper about a day later than Tbn, or that maternal FGF4 is available longer than maternal Tbn (Voss *et al*, 2000).

Proliferation of the early TE appears to be controlled by its close proximity to the ICM derivatives and, as demonstrated in the study of *Oct-4* mutant embryos (Nichols *et al.*, 1998), TE cells do not proliferate in the absence of a normal ICM but instead become giant cells. The *Mash-2* gene which encodes a basic helix-loop-helix transcription factor of the *achaete-scute* family is a specific marker of diploid TE cells. Expression of this imprinted gene is required for maintenance of the giant cell precursor population following implantation (Scott *et al.*, 2000), and in null mutants leads to death from placental failure around 10 dpc (Guillemot *et al.*, 1994). *Cdx-2* is also a specific marker

of TE cells and is required for normal trophoblastic proliferation in the extraembryonic ectoderm during implantation, probably via the induction of *Fgf-4* (Beck *et al.,* 1995). The proto-oncogene *c-fms* is expressed at high levels in the trophoblast from about 9.5 dpc and encodes the receptor for the colony-stimulating factor-1 (CSF-1) cytokine (Regenstreif and Rossant, 1989). The CSF-1 ligand is expressed in uterine epithelium at the same time, suggesting a paracrine role for TE *c-fms* expression. The transcription factor *Hand-1* (Cserjesi *et al.,* 1995) is expressed in the ectoplacental cone, but not in TS cells (section 1.2.4). Expression of *Hand-1* is required to promote normal giant cell transformation of the mural TE, and overlaps with the expression of *Mash-2* in the ectoplacental cone and placental spongiotrophoblast to regulate the maintenance and proliferation of giant cell precursors (Riley *et al.,* 1998; Scott *et al.,* 2000; reviewed by Cross, 1998).

Following implantation, the signals that lead to the initiation of gastrulation at a particular point along the embryonic/extraembryonic junction remain unknown. Several genes, for example, *Goosecoid* (*Gsc*), (Blum *et al.,* 1992), *Brachyury (T),* (Wilkinson *et al.,* 1990), *Evx-1* (Dush and Martin, 1992), *Nodal* (Zhou *et al.,* 1993), *Fgf-8* (Crossley and Martin, 1995), *Lim-1* (Tam and Behringer, 1997), and the signaling molecules of the Wnt and Activin families (Albano *et al.,* 1994), begin expression as the primitive streak appears, although their precise roles in gastrulation are not yet elucidated (reviewed by Faust and Magnuson, 1993). Embryos lacking normal expression for *Nodal,* a TGF-β related gene, cannot initiate primitive streak formation and are arrested at the gastrulation stage of development (Conlon *et al.,* 1991; Varlet *et al.,* 1997). *T* deficient mouse embryos, however, do not generate enough mesoderm and die at mid-gestation with prominent defects in the notochord, allantois and primitive streak (Wilkinson *et al.,* 1990; Beddington *et al.,* 1992).

Gene expression associated with the derivatives arising from the streak suggests that these may be already committed to specific fates, rather than arising as progenitor tissues. Genes such as *Msx-1, Msx-2* and *Lim-1* are expressed in lateral mesoderm but not more medially (Davidson and Hill, 1991; Barnes *et al.,* 1994). *Lim-1* null mutants, for example, show mislocalized proximal expression of *Gsc* and lack anterior head structures, but the remaining body axis develops normally (Blum *et al.,* 1992; Shawlot and Behringer, 1995). Genes that appear to be associated with paraxial mesoderm include the winged-helix domain gene *Mf-1, Mox-1* and *Mox-2, Sek,* and *Follistatin*

15

(Candia *et al.*, 1992; Nieto *et al.*, 1992; Sasaki and Hogan, 1993; Albano *et al.*, 1994). The winged-helix transcription factor, hepatocyte nuclear factor 3β (HNF-3β), marks axial mesoderm and definitive endoderm during gastrulation and is required for normal formation of the node and the notochord, the floor plate and motor neurons in the neural tube, as well as tissues that arise from the gut endoderm (Ang and Rossant, 1994; Sasaki and Hogan, 1994; Weinstein *et al.*, 1994). The homeobox gene *Otx-2* is initially expressed throughout the undifferentiated epiblast before gastrulation, and then marks the prospective forebrain and midbrain at 7.5 dpc (Ang *et al.*, 1994). *Otx-2* deficient mice show severe defects in gastrulation and in formation of axial mesoderm, resulting in the loss of anterior neural tissues, and display mislocalized expression distally for *Cerberus-related 1 (Cer-1)*, (Biben *et al.*, 1998), and proximally for *Gsc* (Ang *et al.*, 1996).

There is now a considerable amount of evidence to support an important role for the extraembryonic lineages in molecular signaling during early embryonic development. Mutations in genes expressed in the VE, for example *Smad* (Sirard *et al.*, 1998), *Evx-1* (Spyropoulos and Capecchi, 1994) and *HNF-4* (Chen W. *et al.*, 1994), lead to rapid and extensive ectodermal death early in gastrulation. Similarly, removal of the anterior VE has been shown to inhibit differentiation of neural tissue (Thomas and Beddington, 1996). Transgenic explant culture studies have demonstrated that signals expressed in the posteriorizing VE during gastrulation are crucial for specifying hematopoietic and endothelial cell fates in the extraembryonic mesoderm, and play a role in patterning the A-P axis (Belaoussoff *et al.*, 1998).

Recent mutation studies for an increasing number of genes that are expressed in the anterior VE before, or concurrent with gastrulation, implicate this tissue as the origin of molecular anterior patterning prior to primitive streak formation (Beddington and Robertson, 1998). Normal patterning of the A-P axis requires expression in the anterior VE of the VE-1 antigen (Rosenquist and Martin, 1995) and the genes *Otx-2* (Ang and Rossant, 1994; Acampora *et al.*, 1995), *Lim-1* and *Gsc* (Belo *et al.*, 1997), *Cer-1* (Belo *et al.*, 1997; Biben *et al.*, 1998), and *Hex* (Thomas *et al.*, 1997; 1998), and precedes formation of the primitive streak by half a day or more (reviewed by Beddington and Robertson, 1998). At the time of nascent streak formation at the opposite side of the egg cylinder, this anterior strip of VE is further patterned by the onset of expression for the genes *Mrg-1*, corresponding to the site of future heart development (Dunwoodie *et*

16

al., 1998), and *Hesx-1 (or Rpx)* in the domain overlying epiblast destined for anterior CNS formation (Hermesz *et al.,* 1996; Thomas and Beddington, 1996).

A recent model has been proposed for the determination of the A-P axis by two distinct organizing centers in the anterior VE and the distal primitive streak, which induce and pattern the head and trunk, respectively. Precursors of these centers originate on the distal and proximal ends of the pregastrulation embryo, with coordinate cell movements rotating this pre-existing distal-proximal axis into an A-P axis prior to gastrulation (Beddington and Robertson, 1998). The earliest A-P asymmetry in gene expression is seen for the homeobox gene *Hex,* which delineates a small cluster of VE cells at the distal end of the egg cylinder immediately following implantation. Expression of *Hex* then switches to one side of the embryo, as these VE cells only give rise to anterior progeny (Thomas *et al.,* 1997, 1998). Cells expressing genes associated with the primitive streak rapidly congregate posteriorly, reflecting this axis (Beddington, and Robertson, 1998). Supporting this model is the recent finding that extraembryonic expression of *Smad-2,* a component of the TGF-β signaling pathway, determines A-P identity both within the epiblast and in the overlying VE (Waldrip *et al.,* 1998). In *Smad-2* mutants the epiblast exclusively forms extraembryonic mesoderm, and fails to give rise to the three primary germ layers. Further, it has been demonstrated that expression of the EGF-like gene *Cripto* is required in the epiblast for correct orientation of the A-P axis (Ding *et al.,* 1998). *Cripto* mutants display a "head-without-trunk" phenotype, with anterior tissue remaining in a distal location and the most posterior epiblast derivative, extraembryonic mesoderm, emerging proximally. They have neither a streak nor embryonic mesoderm, although markers of the streak are expressed throughout the epiblast (reviewed by Beddington, 1998).

Initiation of the germline in the mouse is now known to depend on extraembryonic expression of the TGF-β intercellular signaling protein *Bmp-4,* before gastrulation begins. *Bmp-4* is expressed in the TE-derived extraembryonic ectoderm just prior to gastrulation, and in the extraembryonic mesoderm but not the PGCs in the mid- to late-primitive streak stage. The size but not the expansion of the founding population of PGCs is significantly reduced in heterozygote null mutants for *Bmp-4,* while homozygote embryos completely lack both PGCs and an allantois, and do not survive beyond the early somite stage (Winnier *et al.,* 1995; Lawson *et al.,* 1999).

The proliferation and migration of PGCs to the genital ridges depends on the expression and interaction of the *Dominant white spotting (W)* gene encoding the c-Kit tyrosine kinase transmembrane receptor (Chabot *et al.,* 1988; Geissler *et al.,* 1988) with its ligand, Steel factor (also known as stem cell factor, mast cell growth factor and c-Kit ligand), encoded by the *Steel (Sl)* gene (Huang *et al.,* 1990; Williams *et al.,* 1990; Zsebo *et al.,* 1990). Studies of mouse mutants for these two genes demonstrate the failure of PGCs to follow their tightly regulated pattern of proliferation and migration, and in homozygotes results in severe fertility defects (Mintz and Russell, 1957; McCoshen and McCallion, 1975). Both mutations also affect the neural crest cells and hematopoietic precursor cells, populations of cells that are required to move over long distances in the embryo (Hogan *et al.,* 1994). In assessing a role for the heterodimeric integrins in PGC migration (reviewed by Hynes, 1992), a recent study demonstrated that homozygote null $\beta 1$ embryonic cells are able to enter the germline but there is no transmission, as PGCs lacking $\beta 1$ integrins are impaired in their colonization of the gonads (Anderson *et al.,* 1999). Following normal establishment of the PGC population in the gonadal ridges, sexual differentiation of the somatic cells is initiated at 11.5-13.5 dpc by the expression of the Y-linked gene *Sry* leading to the development of a testis, and a lack of *Sry* expression in females giving rise to an ovary (Koopman *et al.,* 1990).

While the above overview gives an insight into just some of the molecular mechanisms governing specification of the somatic and germline derivatives that arise from the epiblast of the preimplantation embryo, it is clear that some mechanisms are unique from those regulating lower organisms (Wolpert *et al,* 1998). Not only is the mammalian embryo largely responsible for its own genomic transcription during development, but also the future embryonic axis appears to be determined prior to formation of a primitive streak, with many of the inducing signals intimately involved in development before and during gastrulation being emitted from extraembryonic lineages.

1.4   PLURIPOTENTIAL EMBRYONIC STEM CELLS

The great advances in generating molecular information regarding early mammalian development can be largely attributed to the advent of embryonic cell lines that are amenable to genetic modification, while retaining the capacity to contribute to all tissues of a host embryo, including the germline. These cells are typically derived from

18

the pluripotent stem cell population that resides in the ICM of the preimplantation embryo, and in the subset of cells destined for the germ cell lineage following gastrulation. The following discussion encompasses an overview of the derivation of pluripotential cell lines from the mouse embryo, and addresses more recent advances in stem cell biology including the derivation and differentiation *in vitro* of human ES cell lines.

1.4.1 Pluripotency in the developing embryo

As previously discussed in section 1.2, several experiments have demonstrated the potential of single primitive ectoderm cells to contribute to all of the somatic lineages as well as the gametes, and that prior to gastrulation individual cells in the epiblast are not pre-destined to a particular fate, including the germ lineage (Stevens, 1970; Gardner and Rossant, 1979; Gardner *et al.*, 1985; Lawson *et al.*, 1991; Tam and Zhou, 1996). The inherent ability of early embryonic cells for reprogramming and regeneration has been demonstrated by experiments such as those in which a normal embryo can still develop following removal of blastomeres (Kelly, 1977), or removal of trophectoderm at the blastocyst stage (Handyside, 1978; Hogan and Tilly, 1978), and even following the ablation of more than 80% of embryonic cells at the egg cylinder stage (Snow and Tam, 1979). *In vitro*, ICM of early blastocysts from which already differentiated TE has been removed is able, under certain conditions, to regenerate this tissue as well as delaminate TE and endodermal cells (Nichols and Gardner, 1984).

These properties of the ICM/epiblast cells are characteristic for stem cells which have been described as "undifferentiated cells that are capable of (a) proliferation, (b) self maintenace, (c) the production of a large number of differentiated functional progeny, (d) regenerating the tissue after injury, and (e) a flexibility in the use of these options" (Potten and Loeffler, 1990). While the true definition of a stem cell remains contentious, the essential characteristics of all stem cells are a capacity for unlimited or prolonged self-renewal and the long-term potential to form one or more highly differentiated descendent cell type(s). Although the pluripotent cells of the epiblast contribute to all adult tissues, *in vivo*, they are not destined to undergo self-renewal indefinitely, but rather act as precursor cells proliferating and replacing themselves for a limited period of time before committing to specific lineages. However, when removed from their normal embryonic environment and cultured under appropriate conditions,

19

cells of the epiblast as well as the PGCs that colonize the developing gonadal ridges can proliferate and replace themselves indefinitely, and yet maintain the developmental potential to form advanced derivatives of all three embryonic germ layers, thus satisfying the criteria for pluripotent stem cells (reviewed by Morrison et al., 1997; Watt and Hogan, 2000).

1.4.2 Derivation of pluripotential stem cells from embryonal carcinomas

The potential for a self-renewing stem cell within the epiblast was first noted in a series of experiments demonstrating that grafting of early pre- and postimplantation embryos to vascularised ectopic sites in adult mice gave rise to teratocarcinomas at a high frequency (Solter et al., 1970; Stevens, 1970). Teratocarcinomas are disorganized multi-differentiated tumors containing derivatives of all germ layers which also arise spontaneously from male germ cells in the 129 inbred strain of mice (Stevens and Little, 1954), and from female germ cells that undergo parthenogenetic development in the ovaries (Stevens and Varnum, 1974). The ability of ectopic grafts to form malignant rather than benign tumors is restricted to embryos up to 7.5 dpc, and dissection of embryos prior to grafting attributes this ability to the epiblast (Damjanov et al., 1971; Diwan and Stevens, 1976; Beddington, 1983). Teratocarcinomas appear to arise directly from disruption of the normal growth and differentiation program of embryo stem cells, and contain nests of undifferentiated proliferating stem cells known as embryonal carcinoma (EC) cells amongst the mixture of differentiated derivatives (Stevens, 1970; reviewed by Damjanov and Solter, 1974; Stevens, 1983). These EC cells can be isolated and maintained as pluripotential cell lines in vitro (Evans 1972; Bernstine et al., 1973; Martin and Evans, 1975; Martin, 1980).

EC cells show a remarkable similarity to normal ICM/epiblast cells in their morphology, ultrastructure, protein synthesis profiles and expression of cell surface antigens, and as such provided the first in vitro model for the study of early mammalian development (Martin, 1975, 1980; Diwan and Stevens, 1976). The in vitro differentiation of EC cell lines is rarely spontaneous and usually requires culture in the presence of chemical inducers, or under appropriate density conditions (reviewed by Graham, 1977; Martin, 1980). By culturing EC cells as aggregates, some lines are able to form complex differentiated cystic structures called embryoid bodies, which resemble isolated ICMs cultured in suspension (Martin and Evans, 1975; Martin et al., 1977;

Wiley *et al.*, 1978). Most remarkably, these tumor stem cells are able to resume normal embryonic development following introduction to a host blastocyst or aggregration with morulae, but rarely give rise to normal gametes and offspring (Brinster, 1974; Mintz and Illmensee, 1975; Papaioannou *et al.*, 1975; Stewart C., 1982; Stewart T. and Mintz, 1982; Fujii and Martin, 1983). Most EC cell lines are aneuploid, which may explain their low frequency of germline contribution, and by nature of being malignant are often detrimental to the health of host animals. Thus, the EC cell system has its limitations as a model for the study of normal embryogenesis and as an effective route for transgenic modification of the genome.

Human EC cell lines have been developed as models for cell differentiation in human development and have proven to be quite different in some respects from their mouse counterparts (Andrews *et al.*, 1984; Andrews, 1988; Pera *et al.*, 1989, 1990; Roach S. *et al.*, 1993). Both mouse and human EC cells express alkaline phosphatase, and the transcription factors *Oct-4, Genesis* and *Germ cell derived nuclear factor* (Sutton *et al.*, 1996; Lei *et al.*, 1997; Pera and Herszfeld, 1998), but differ in expression for several cell surface antigens and other markers. Mouse EC cell surface carbohydrate epitopes are recognized by monoclonal antibody (MAb) SSEA-1, and following differentiation by MAbs SSEA-3 and SSEA-4. In contrast, undifferentiated human EC cells express epitopes for MAbs TRA 1-60, TRA-1-81, SSEA-3 and SSEA-4, but not for SSEA-1 (Wenk *et al.*, 1994). Human EC cells also express epithelial markers, carbohydrate epitopes associated with a keratan sulphate/chondroitin sulphate proteoglycan, the tumor necrosis factor receptor CD30, and cyclin D2, all of which are absent on undifferentiated mouse EC cells (reviewed by Pera, 1999).

1.4.3 Derivation of pluripotential stem cells from early embryos

The isolation of pluripotential stem cell lines directly from the embryo without the need for a tumor phase now provides immortal cell lines that have a normal karyotype and can efficiently colonize the germline.

(a)   Mouse ES cells

Soon after the advent of EC cell lines, it was demonstrated in two independent studies that pluripotent, karyotypically normal, mouse embryonic stem (ES) cell lines could be derived directly from either immunosurgically isolated epiblasts or whole implantation-delayed blastocysts, and maintained indefinitely *in vitro* (Evans and Kaufman, 1981;

21

Martin, 1981). These ES cells resemble mouse EC cells with respect to morphology, *Oct-4*, alkaline phosphatase and surface antigen expression. ES cells differentiate readily into multiple tissue types both *in vitro* in embryoid bodies (Doetschman *et al.*, 1985), and *in vivo* in teratocarcinomas (Evans and Kaufman, 1981). Mouse ES cells participate fully in normal embryogenesis when reintroduced to host blastocysts, and can extensively contribute functional differentiated progeny to all somatic tissues and to the germline (Bradley *et al.*, 1984). In fact, formation of an entire viable fetus can be achieved following aggregation of ES cells and tetraploid embryos, providing the normal somatic status of imprinted genes has been faithfully propagated in ES cell culture (Nagy *et al.*, 1993; Dean *et al.*, 1998). These properties have revolutionized mutagenesis of the mouse by enabling the introduction of defined genetic modifications, as will be discussed in section 1.5, and has led to the widespread use of ES cells in developmental research (reviewed by Smith, 1992).

Mouse ES cells initially required a layer of mitotically inactivated mouse embryonic fibroblast feeder cells to support their survival, growth and inhibition of differentiation during culture (Evans and Kaufman, 1981; reviewed by Robertson, 1987). However, this need was subsequently obviated by the identification of the cytokine LIF (or DIA) which, in a purified form is able to sustain the self-renewal and propagation of ES cells in culture without compromising developmental potential *in vitro*, or *in vivo* (Gearing *et al.*, 1987; Smith and Hooper 1987; Smith *et al.*, 1988; Williams *et al.*, 1988). In fact, germline competent ES cell lines have since been established by direct culture of embryos in medium supplemented with LIF (Nichols *et al.*, 1990; Pease *et al.*, 1990).

Only a few strains of inbred mice have proven to be permissive for isolating ES cell lines from blastocyst explant cultures. Most of the ES cell lines currently in use are derived from a congenic 129/Sv strain of mice which is characterized by a high incidence of spontaneous testicular teratomas and teratocarcinomas (Evans and Kaufman, 1981; Martin, 1981; Robertson, 1987). The ability to overcome species and strain limitations for the isolation of ES cell lines will open up the scope to elucidate gene function via ES cell transgenesis (see section 1.5.3) and may provide appropriate animal models for human genetic diseases. Some success has now been achieved with this following the derivation of ES cell lines from inbred and well characterized strains including C57BL/6J (Ledermann and Burki, 1991; Kontgen *et al.*, 1993), DBA/1lacJ (Roach M. *et al.*, 1995), and BALB/cJ (Noben-Trauth *et al.*, 1996). More recently,

22

microsurgically isolated epiblasts and single, dissociated epiblast cells have been demonstrated to yield ES cell lines at a substantantially higher frequency than intact blastocysts, not only for the 129/Sv strain, but also for the previously nonpermissive CBA/Ca strain (Brook and Gardner, 1997).

(b) EG cells derived from fetal primordial germ cells

Following the isolation of EC and mouse ES cell lines it was found that mouse PGCs, when cultured on feeder layers with the addition of LIF, Steel factor and FGF-2 (FGF2), give rise to cells that resemble undifferentiated blastocyst-derived ES cells (Matsui *et al.*, 1992; Resnick *et al.*, 1992; Koshimizu *et al.*, 1996; reviewed by Buehr, 1997). These karyotypically normal embryonic germ (EG) cell lines have been derived from the gonadal ridges of 8.0-8.5 dpc embryos and also from older PGCs up to 12.5 dpc, for both sexes (Labosky *et al.*, 1994a; Tada T. *et al.*, 1998). Like mouse ES cells, EC cells and PGCs, EG cells express alkaline phosphatase, *Oct-4* and the SSEA-1 epitope. EG cells appear to have an equivalent pluripotentiality to ES cells as they can be induced to differentiate extensively in culture, form teratocarcinomas when injected to ectopic sites, and have the capacity to form chimaeras and populate the germline (Labosky *et al.*, 1994a; Stewart C. *et al.*, 1994; Tada T. *et al.*, 1998). EG cell lines have reportedly been derived from the culture of 5-, 6-, 7-, and 11-week postfertilization human PGCs on mouse STO feeders, in the presence of human FGF2, human LIF and forskolin (Shamblott *et al.*, 1998). Human EG cells resemble mouse ES and EG cells in morphology and marker gene expression, but also display carbohydrate cell surface markers found on human EC cells (section 1.4.2). Teratoma formation from these cells has not yet been reported, but following *in vitro* differentiation human EG cells form embryoid bodies composed of derivatives of the three germ layers, and express a broad range of mRNA and protein markers associated with a number of different cell lineages (Shamblott *et al.*, 1998, 2001; reviewed by Thomson and Odorico, 2000).

Although EG cell lines demonstrate a remarkable developmental potential, there appear to be differences between mouse EG and ES cell lines as a result of genomic imprinting. Unlike mouse ES and normal somatic cells, mouse EG cells show a similar epigenotype to that of PGCs which is characterized by genome-wide demethylation (Monk *et al.*, 1987). Changes in methylation for several imprinted loci have been demonstrated for some EG cells from 8.0-8.5 dpc embryos, and for most if not all EG cells from 11.5-

12.5 dpc embryos, with comparable epigenetic modifications for both male- and female-derived cells (Labosky *et al.*, 1994b; Tada T. *et al.*, 1998). These modifications are likely to reflect the *in vivo* erasure of allele-specific imprints and reprogramming events that normally occur during germ cell development (reviewed by Tilghman, 1999). Strikingly, EG cells are capable of effecting extensive heritable epigenetic modification of a somatic nucleus in EG-somatic cell hybrids, so that it resembles the germ cell nucleus (Tada M. *et al.*, 1997). It has been demonstrated, however, that a PGC nucleus cannot give rise to a viable embryo following transplantation to an enucleated oocyte (Kato *et al.*, 1999). Despite the implied developmental limitations for PGC-derived cells, their potentially imprint-free status may provide a more readily programmable pluripotential stem cell source for the derivation of imprint-silenced somatic cells, in the types of human somatic cell transplantation therapies discussed in section 1.4.6.

(c) Primate and human ES cells

Since the first description of mouse ES cells, it has been recognized that the derivation of human ES cells would provide a unique resource for functional studies of the human genome. A diploid pluripotential ES cell line with extensive capacity to differentiate was isolated in 1995 from rhesus monkey blastocysts (Thomson. *et al.*, 1995). This was followed soon after by the isolation of an ES cell line from the common marmoset, a primate considered to have more favourable reproductive characteristics for experimental embryology (Thomson *et al.*, 1996). Characterization of these primate ES cell lines showed that they resemble human EC cells more closely than their mouse counterpart in morphology, marker expression and lack of response to LIF *in vitro* (section 1.4.2). In 1998, James Thomson and his co-workers reported the derivation of five karyotypically normal human blastocyst-derived pluripotential ES cell lines, a finding which has fueled enormous interest and debate in stem cell biology. While ethical considerations prevent a confirmation of germline competency, characterization of these and subsequent feeder-dependent human ES cell lines (Reubinoff *et al.*, 2000) has demonstrated their ability to differentiate readily *in vitro* and give rise to teratomas containing derivatives of all three germ layers after injection into SCID mice. Human ES cells express *Oct-4*, alkaline phosphatase and surface epitopes found on human EC cells. As well, they express high levels of telomerase, an enzyme associated with immortal cell lines, and are capable of prolonged undifferentiated proliferation

(Thomson *et al.*, 1998; Reubinoff *et al.*, 2000; reviewed by Thomson and Odorico, 2000).

The availability of human ES cell lines now provides an *in vitro* model for the study of human development, in the context of a normal genetic background. However, present culture systems do not support routine clonal or large-scale growth of human ES cells. It is not yet clear whether growth is limited by slow cell division or a high incidence of spontaneous differentiation, or by cell death *in vitro*.

1.4.4 Regulation of pluripotency in ES cells

The molecular mechanisms that regulate the propagation of ES cells and underpin pluripotency, while not yet fully elucidated, have started to unfold in recent years (reviewed by Burdon *et al.*, 1999a). LIF promotes the proliferation of undifferentiated mouse ES cells through the activation of a heteromeric complex containing the class I cytokine receptor subunits gp130 and LIF-receptor (LIF-R), (Gearing *et al.*, 1991; Gearing and Bruce, 1992; Davis *et al.*, 1993). LIF is secreted in both a diffusible form as well as a matrix-bound form, to localize its action, and is known to exhibit a wide range of biological activities both *in vitro* and *in vivo* (Williams *et al.*, 1988; reviewed by Hilton and Gough, 1991; Smith *et al.*, 1992). The LIF-related cytokines, oncostatin M (OSM), cardiotrophin-1 (CT-1) and ciliary neurotrophic factor (CNTF) act through the same receptor complex, and each can similarly sustain self-renewal of ES cells (Conover *et al.*, 1993; Rose *et al.*, 1994; Wolf *et al.*, 1994; Pennica *et al.*, 1995). The gp130 receptor alone can also support the derivation and propagation of ES cell lines when homodimerized in response to interleukin-6 (IL-6) and a soluble form of IL-6 receptor (Nichols *et al.*, 1994; Yoshida *et al.*, 1994).

Biological responses to ligands that activate gp130 appear to be principally mediated by the signal transducer and activator of transcription (STAT) and mitogen-activated protein kinase (MAPK) signaling pathways, and depend upon the activation of Janus kinases (JAKs) and phosphorylation of Src-homology-2 (SH2) cytoplasmic domains on the receptor (Darnell *et al.*, 1994; Stahl *et al.*, 1995; Fukada *et al.*, 1996; Sheng *et al.*, 1997; reviewed by Darnell, 1997). A recent study reported that ES cells chimaeric for gp130 and the related granulocyte colony-stimulating factor receptor (G-CSF-R) are unable to engage the STAT3 transcription factor and are incapable of self-renewal, while the over-expression of a STAT3 interfering mutant causes ES cells to differentiate

25

(Niwa *et al.*, 1998). Conversely, suppression of the MAPK effectors extracellular regulated kinase (ERKs) 1 and 2 via elimination of the gp130 phosphotyrosine binding site for the most N-terminal SH2 domain of the protein tyrosine phosphatase SHP-2, has been shown to promote self-renewal of ES cells in the presence of sustained STAT3 activation (Burdon *et al.*, 1999b). The lack of an essential requirement for gp130-dependent SHP-2 recruitment and ERK signaling in ES cell propagation points to the possibility that STAT3 activity alone may be sufficient for blocking the differentiation of ES cells (Burdon *et al.*, 1999b), and is in contrast to the situation observed in many somatic cell types where stimulation of the ERK pathway is required for progression through the G1/S phase of the cell cycle (Pages *et al.*, 1993). That ES cells have a rapid doubling time of 8-12 h and a shortened G1 phase, and possess few of the G1-associated control mechansims, may provide a rationale for this observation (Mac Auley *et al.*, 1993; Savatier et *al.*, 1994, 1996; reviewed by Burdon *et al.*, 1999a).

Despite its apparent redundancy in self-renewal signaling, SHP-2, a homolog of the *Drosophila* protein corkscrew, may play an important role in regulating the normal differentiation process of ES cells. Overexpression of catalytically inactive SHP-2 proteins in ES cells inhibits their differentiation both in monolayer culture and in embryoid bodies (Burdon *et al.*, 1999a, 1999b). Furthermore, ES cells mutant for the gp130 binding site in both copies of the *Shp-2* gene are compromised in their capacity to differentiate upon withdrawal of LIF, or when aggregated to form embryoid bodies (Qu *et al.*, 1997; Qu and Feng, 1998). Differentiation of ES cells *in vitro* is associated with the induction of G1 cyclin expression and the re-establishment of G1 cell cycle control mechanisms (Savatier *et al.*, 1996). This reflects the transition that also occurs in the epiblast at gastrulation, and there is increasing evidence that ERK activation is an important effector in regulating these differentiation events. However, the exact stage(s) at which SHP-2 or ERK signaling is required for the process of ES cell commitment and differentiation is not yet known. On the basis of current evidence it appears that a fine balance between the STAT3 and ERK effectors may determine the regulation of ES cell pluripotency (Burdon *et al.*, 1999b; reviewed by Burdon *et al.*, 1999a).

Factors unrelated to the gp130 signaling pathway may also be implicated in supporting the growth of pluripotent stem cells. For example, self-renewal of ES cell colonies can be sustained by the paracrine production of a soluble, macromolecular, trypsin-sensitive

26

activity termed ES cell renewal factor (ESRF), derived from a LIF-deficient parietal endoderm-like cell line (Dani *et al.*, 1998).

Another gene implicated in the regulation of the pluripotent phenotype of normal or malignant embryonic stem cells is the novel winged-helix transcriptional repressor *Genesis*. Isolated from an embryonal carcinoma cDNA library *Genesis* expression is restricted to ES and EC cell lines, and declines rapidly with the induction of differentiation, even moreso when cells are exposed to retinoic acid than with the removal of LIF (Sutton *et al.*, 1996). As well, over-expression of *Pem*, a gene implicated in regulating early transition of ES cells to specific extraembryonic lineages, has been demonstrated to block the *in vitro* and *in vivo* differentiation of ES cells (Fan *et al.*, 1999). Finally, *Oct-4*, which as discussed earlier is crucial for establishment of the ICM in developing embryos (section 1.3.1), has now been established as a master regulator for ES cell pluripotency following the demonstration that a precise level of expression for this transcription factor is required to sustain self-renewal. Up-regulation of *Oct-4* in ES cells causes differentiation to primitive endoderm and mesoderm, while down-regulation induces dedifferentiation to trophectoderm (Niwa *et al.*, 2000). Interestingly, both *Oct-4* and *Genesis* bind to the retinoic acid-repressible E2 embryonic stem cell enhancer sequence through which *Oct-4* can activate transcription, indicating that they may also share genetic regulatory function (Sutton *et al.*, 1996).

1.4.5 *In vitro* generation of somatic lineages from ES cells

Mouse ES cells grown as aggregates *in vitro* spontaneously give rise to complex embryoid body structures containing a variety of differentiated cell types including extraembryonic yolk sac, cardiomyocytes, haematopoietic cells (Doetschman *et al.*, 1985; Keller, 1995), skeletal myocytes (Rohwedel *et al.*, 1994), epithelial cells (Bagutti *et al.*, 1996), and adipocytes (Dani *et al.*, 1997). Current research efforts in ES cell biology are focused on determining the *in vitro* requirements for efficiently and selectively directing both ES and EG cells towards specific cell lineages. Although strategies for doing this are currently limited, particularly for human cell lines, by the inability to generate pure populations of desired lineages, considerable progress in mouse ES cell differentiation to neural, hematopoietic, endothelial and cardiac tissue has been made, with many aspects of the lineage-specific differentiation programs

27

observed within embryoid bodies reflecting those found in the embryo (Keller *et al.,* 1993; Bain *et al.,* 1995; Keller, 1995; Klug *et al.,* 1996; Brustle *et al.,* 1997).

Retinoic acid has been shown to induce or repress differentiation programs in ES cell cultures in a time- and concentration-dependent manner (Wobus *et al.,* 1994). The early treatment of embryoid bodies with retinoic acid for a precise period of time is a requirement for the efficient generation of clusters of ES cell-derived mature adipocytes in outgrowths (Dani *et al.,* 1997). Exposure of later stage aggregation cultures to retinoic acid for a longer period of time induces a sizeable percentage of the differentiating cells to become mature neurons, astrocytes and oligodendrocytes, as well as proliferative precursors that give rise to functional neurons and glia (Bain *et al.,* 1995; Fraichard *et al.,* 1995; Strübing *et al.,* 1995; Finley *et al.,* 1996). Retinoic acid treatment inhibits ES cell cardiogenesis and the expression of mesodermal marker genes such as *Brachyury,* zeta-globulin and cardiac specific isoforms of actins (Wobus *et al.,* 1994; Bain *et al.,* 1996). As well, a switch from cardiogenesis to skeletal myogenesis and adipogenesis has been observed after retinoic acid treatment of embryoid bodies (Wobus *et al.,* 1994; Dani *et al.,* 1997). In a first report for somatic cell derivation from human ES cells, neuronal progenitor cells isolated from cultures grown for extended periods at high density were induced to form mature neurons when cultured on an appropriate substrate, and in a serum-free medium supplemented with retinoic acid (Reubinoff *et al.,* 2000). Retinoic acid is, however, a strong teratogen which perturbs neural patterning and neuronal identities *in vivo,* and as such may not always be a preferred choice for neural induction where specification of a particular neuronal characterisitic, such as neurotransmitter choice, is required for neuroscience research or therapeutic applications (reviewed by Kawasaki *et al.,* 2000).

Enrichment of populations can be achieved by the addition of growth factors and media supplements that favor the survival of desired cell types in embryoid body cultures. For example, highly enriched populations of neuroepithelial progenitor cells can be derived from mouse ES cells cultured in defined serum-free media containing FGF2, and sequential combinations of FGF2, EGF and PDGF. Upon the withdrawal of growth factors, these precursors readily differentiate to the three principal lineages of the nervous system (Okabe *et al.,* 1996; Brustle *et al.,* 1997; Brustle *et al.,* 1999). The co-culture of mouse ES cells with certain bone marrow stromal cell lines, as well as the addition of Steel factor and various cytokines at the later stages of embryoid body

28

differentiation enhances hematopoietic differentiation and the derivation of erythroid lineages (Keller *et al.*, 1993; Palacios *et al.*, 1995). In a step towards achieving directed *in vitro* differentiation of human ES cells, a recent study reported the differentiation into cells with different epithelial or mesenchymal morphologies following exposure to eight growth factors. While none of the factors directed differentiation exclusively to one cell type, it was observed that Activin-A and TGF-β1 mainly induced mesodermal cells, retinoic acid, EGF, bone morphogenic protein (BMP)-4 and FGF2 activated ectodermal and mesodermal markers, while β nerve growth factor (βNGF) and hepatocyte growth factor (HGF) allowed differentiation into the three embryonic germ layers (Schuldiner *et al.*, 2000).

Recent developments in 2-dimensional culture systems for ES cell differentiation now provide more amenable systems for the investigation of ES cell growth and differentiation, compared with the disorganized and less accessible differentiation that occurs within embryoid body structures (Nishikawa *et al.*, 1998). A recent study has reported the efficient differentiation of mouse ES cells to mesencephalic dopaminergic neurons following simple 2-dimensional co-culture of ES cells with stromal cells derived from skull bone marrow, in serum-free media. These stromal cells produce an activity that promotes neural differentiation from ES cells, without the need for embryoid body formation or retinoic acid treatment. Further, the addition of BMP4 to the co-cultures suppressed the stromal cell neuralization activity, and instead promoted epidermogenesis (Kawasaki *et al.*, 2000). The specification of mouse ES cells to hematopoietic and endothelial cell lineages has likewise been demonstrated to be independent of a requisite for 3-dimensional embryoid body culture or feeder cells. ES cells spreading 2-dimensionally on a type IV collagen matrix can differentiate and be sorted by a panel of cell surface markers to identify intermediate stage FLK1$^+$/E-cadherin$^-$ mesoderm cells from other cell lineages, for purified re-culture and induction of either vascular or blood cells in the presence of various recombinant growth factors (Nishikawa *et al.*, 1998; Yamashita *et al.*, 2000).

The homogeneous formation from mouse ES cells of a distinct type of pluripotent cell termed the primitive ectoderm-like (EPL) cell, demonstrates for the first time, the uniform and reversible differentiation of ES cells in response to biologically derived factors without aggregation culture (Rathjen *et al.*, 1999). The EPL cell closely resembles the 5.5-6.0 dpc primitive ectoderm lineage, and is able to adopt distinct

interchangeable states depending on the absence or presence of two separable factors derived from medium conditioned by a human hepatocellular carcinoma line. As demonstrated for the precise regulation of ES cell fate by *Oct-4* (Niwa *et al.*, 2000), it may be that other signals determining cell lineages act according to a quantitative expression mechanism, which will require transgenic or cell culture strategies to mimic the appropriate levels of expression for these signals.

While defined culture conditions will preferentially promote the differentiation of ES cells into certain lineages, the resulting population of cells is usually a heterogeneous array of differentiated types. For a few cell lineages of interest, transgenic strategies (see section 1.5.3) have now been described for generating pure cultures of these from a mixed population of differentiating mouse ES cells. Using an α-cardiac myosin heavy chain promoter to drive a selectable reporter gene in transfected ES cells, an essentially pure population of cardiomyocytes can be derived following differentiation (Klug *et al.*, 1996). Similarly, neuronal differentiation of selectable ES cell clones harboring a reporter gene under the control of *Sox-2*, which is expressed in embryonic neuroepithelium, led to the derivation of a pure population of neuroepithelial progenitor cells (Li *et al.*, 1998). Recently, the selection of an ES cell-derived insulin-secreting clone has been reported, following differentiation in low-glucose culture of cells transfected with a selectable reporter gene under the control of the regulatory regions of the human insulin gene (Soria *et al.*, 2000). These studies clearly demonstrate the future potential to exploit differentially expressed selectable marker genes for the efficient *in vitro* purification of desired cell types.

1.4.6 Applications for ES cells

ES cell-based research is facilitating a rapid advancement in our knowledge of mammalian biology, and emerging as a powerful tool for a wide range of applications in functional genomics and possibly human therapeutic medicine.

The types of research discussed in section 1.4.5 hold promise for the future derivation of purified populations of specified human ES cell-derived somatic cell types. The ability to generate such populations in large numbers would be of invaluable use for *in vitro* growth factor and drug discovery programs as well as for toxicity testing. Their application to cell-based transplantation therapies in the treatment of human disorders caused by a dysfunction or degeneration of a particular cell type can also be envisaged.

For example, different types of neurons for the treatment of spinal cord injuries and devastating degenerative disorders such as Parkinson's disease, Alzheimer's disease, multiple sclerosis and muscular dystrophy; heart muscle cells to replace infarct tissue or to treat congenital disorders of the heart (an organ that completely lacks a progenitor stem cell), and insulin-secreting pancreatic islet cells for the treatment of certain types of diabetes.

Several recent reports describe the successful transplantation and integration of mouse ES cell-derived somatic cells in animal models, though the functional significance of the transplanted cells is not in all cases clear. To date, the best example for this is the transplantation of ES cell-derived glial cells in a rat model for a human myelin disease, demonstrating their successful interaction with host neurons and the efficient myelination of axons in the brain and spinal cord (Brustle *et al.*, 1999). ES cell-derived cardiomyocytes have also successfully formed stable intracardiac grafts in the hearts of adult dystrophic mice (Klug *et al.*, 1996), and insulin-secreting cells have been able to normalize blood glucose when transplanted to streptozotocin-induced diabetic mice (Soria *et al.*, 2000). While these studies demonstrate the exciting potential for human ES cell-based therapies, numerous challenges are yet to be met. It will first be necessary to achieve pure populations of specific cell types to minimize the associated risk of teratoma or teratocarcinoma transformation for transplanted cells, and to also ensure the prevention of immune rejection (reviewed by Fuchs and Segre, 2000; Thomson and Odorico, 2000).

Human stem cell technology is currently under serious ethical scrutiny due to the use of donor human preimplantation embryos or aborted fetuses for the derivation of ES and EG cell lines. The alternate use of nuclear transfer technology for reprogramming a differentiated cell, as has been achieved in cattle, mice and sheep (Cibelli *et al.*, 1998; Wakayama *et al.*, 1998; Wilmut *et al.*, 1997), would require the use of an enucleated human oocyte and an adult somatic cell to produce a embryo. This·procedure would have the added benefit of providing ES cells that are genetically matched for all nuclear genes of the specific nuclear donor, but remains a procedure that holds serious moral considerations for society.

Stem cell therapies based on multipotent stem cells derived from adult tissues would circumvent many ethical issues associated with ES cell-based therapies, if suitable culture conditions can be established to maintain their proliferative capacity and direct

31
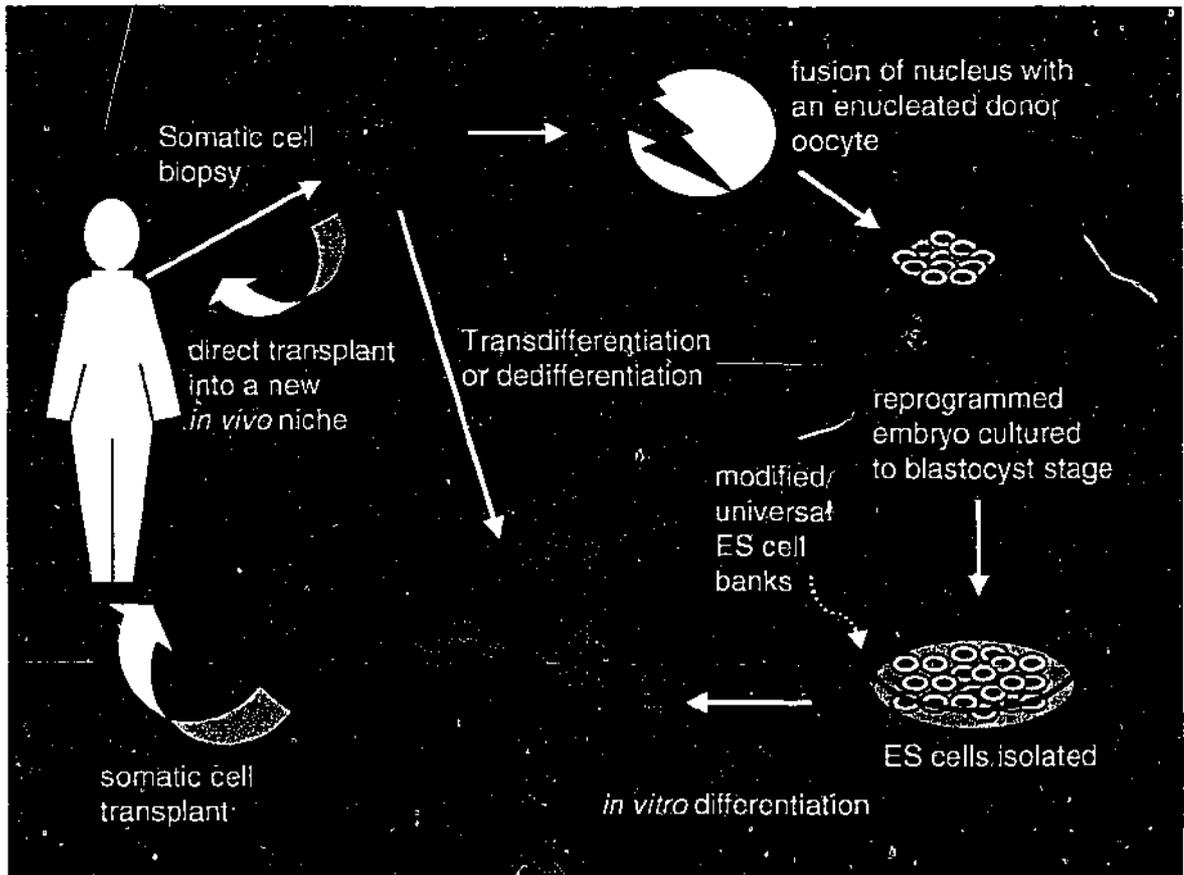
Figure 1.3: Model pathways to reprogramming somatic cells for the treatment of human disorders. In the shortest route, multipotent stem cells isolated from a biopsied somatic tissue are directly implanted into a new somatic tissue for *in vivo* reprogramming. In an intermediate route, such multipotent stem cells are cultivated and trans- or de-differentiated *in vitro* into a new progenitor stem cell type. The new cell type is then transplanted to a new *in vivo* niche where they can repopulate damaged or defective somatic tissue. In the longest route depicted, an enuclated donor oocyte is used to reprogram the nucleus from a biopsied somatic cell and ES cells are isolated from the ICM of the resulting culture-derived blastocyst. The ES cells are differentiated *in vitro*, along the cell lineage of choice, and transplanted to the appropriate *in vivo* niche. Alternatively, universal banks of genetically modified ES cells that are compatible with the recipient are used as a source for somatic cell derivation.

them to specified lineages. Recent reports have also suggested a previously unrecognized plasticity in the developmental capacity of some adult mouse multipotent stem cells. These reports include the ability for clonally derived adult neural stem cells to repopulate the hematopoietic system and produce a variety of blood cell types in sublethally irradiated mice (Bjornson *et al.*, 1999), for adult neural stem cells to contribute to chimaera formation in both chick and mouse embryos, giving rise to cells of all germ layers (Clarke *et al.*, 2000), for marrow stromal cells to generate astrocytes in the brain (Kopen *et al.*, 1999), for hematopoietic stem cells to give rise to myogenic progenitors which will participate in muscle regeneration (Ferrari *et al.*, 1998; Gussoni *et al.*, 1999), and muscle satellite cells to generate blood cells (Gussoni *et al.*, 1999; Jackson *et al.*, 1999). Possibly the most convincing evidence of stem cell plasticity comes from a recent study reporting that bone marrow transplantation in a mouse model of a lethal hereditary liver disease leads to hepatic regeneration and rescue of the mice (Lagasse *et al.*, 2000). In this study, donor-derived hepatocytes were found in large repopulating nodules in about 30-50% of the liver mass and were further demonstrated to be derived only from the hematopoietic stem cell population in adult bone marrow.

It remains debatable whether the experiments discussed above demonstrate environmentally induced transdifferentiation or dedifferentiation of truly plastic adult stem cells, or the selection of rare uncommitted cell types residing in the donor tissues. The possibility may be opened up for redirecting some progenitor cell types to other less easily obtainable types if stem cells from different adult tissues do prove to be more similar than previously thought (for reviews see Fuchs and Segre, 2000; Thomson and Odorico, 2000; Watt and Hogan, 2000; Weissman, 2000).

The potential pathways discussed above for reprogramming both ES and somatic cells in the treatment of human disorders are shown schematically in Figure 1.3. In order to realize the potential applications for human ES cells, it remains to establish efficient protocols for generating and manipulating pure populations of desired cell lineages *in vitro*. This will hinge greatly on an improved understanding of the intrinsic regulators of both ES and multipotent stem cells, as well as the extrinsic influences of the microenvironments, or niches, where stem cells normally reside.

1.5 MUTATIONAL ANALYSIS OF MOUSE DEVELOPMENT

Mutational analysis is a powerful tool which allows the biological function of a single gene product to be studied *in vivo*. By designing breeding programs for different mutations, the combined effect of multiple mutations can also be studied. Probably the most informative mutagenic screens have been performed in *Drosophila* where the systematic isolation and molecular analysis of mutations affecting the formation of body segments, or segment identity, has revealed many of the genes involved in embryonic pattern formation (Nüsslein-Volhard and Wieschaus, 1980; Akam, 1987). Similar large-scale screens have also been performed for *C. elegans* (Kemphues, 1988), *Arabidopsis* (Mayer *et al.*, 1991), and more recently for the zebrafish *Danio rerio* (Driever *et al.*, 1996; Haffter *et al.*, 1996). Many features of the zebrafish such as its availability in large numbers, and the external and rapid development rate of its transparent embryos, lend this vertebrate well to large-scale mutation screens and phenotypic analyses. However, the number of mutations that could be identified and mapped in the latter screens has been limited by current knowledge of the zebrafish genome (reviewed by Lardelli, 2000). Detailed mapping of the mouse genome presently makes the mouse a more feasible candidate for identifying vertebrate genetic mutations.

The following discussion summarizes the alternative strategies used to generate a resource of mouse mutants as models for the study of normal mammalian development, with a more detailed discussion being given in section 1.6 on the use of entrapment techniques in ES cells (for reviews see Jaenisch, 1988; Rossant and Hopkins, 1992; Faust and Magnuson, 1993; Hogan *et al.*, 1994; Bedell *et al.*, 1997).

1.5.1 Classical mouse mutants

Genetic defects that have arisen either spontaneously or in the offspring of mice exposed to X-rays have been an important source for identifying genes involved in mouse development. Identified on the basis of visible mutant phenotypes such as skeletal abnormalities, behavioural defects or coat color and hair morphology, a number of these mutants have been characterized molecularly by mapping cloned genes to the locus of interest.

Many hundreds of genetic loci have been mapped to specific regions of chromosomes in mouse mutants (Lyon *et al.*, 1996). Several mutant genes have been assigned a role in

development by the "candidate" approach where cloned mutant loci in the vicinity of a newly mapped gene are reviewed to determine if any cause a phenotype that might be expected from an alteration in the mapped gene. The genes encoding the c-Kit receptor and its ligand Steel factor (section 1.3.2) were identified from two spontaneous mouse mutants by this approach, revealing a role for c-Kit-mediated signal transduction pathways in the development of hematopoietic, melanogenic and germ cell lineages (Chabot *et al.*, 1988; Huang *et al.*, 1990). Similarly, mapping of the paired box-containing gene *Pax-1*, which is expressed in sclerotome cells and later in intervertebral disks, led to the identification of mutations of this gene in three independent alleles of *Undulated* mutants displaying abnormal vertebrae, and suggests a role for *Pax-1* in sclerotome patterning (Balling *et al.*, 1988; Kessel and Gruss, 1990). In the absence of candidate genes some mutations have been identified by positional cloning directly from mapping information, such as for the *T* mutant (Herrmann *et al.*, 1990) and the *Sry* mutant (Gubbay *et al.*, 1990), but as discussed in the following section (1.5.2), this can be a difficult task when applied widely to the mouse genome.

Analysis of spontaneous retroviral gene insertion sites in the mouse genome has also led to the identification of a number of important genes including *Dilute* (Mercer *et al.*, 1991), *Hairless* (Stoye *et al.*, 1988), and *Pink-eyed unstable* (Brilliant *et al.*, 1991). In some cases identification of such retroviral insertions has provided a molecular handle for the identification of nearby mutant loci. For example, positional cloning from the *Dilute* mutation on chromosome 9, which affects pigmentation by disturbing normal melanocyte function (Mercer *et al.*, 1991), subsequently led to the identification of the *Short-ear* locus which regulates skeletal differentiation as the gene encoding BMP5 (Kingsley *et al.*, 1992).

1.5.2 "Saturation" mutagenesis

The large-scale screens undertaken in *Drosophila* approached saturation for chemically induced mutations whose phenotypes indicated involvement in developmental processes (Nüsslein-Volhard and Wieschaus, 1980). Such large-scale mutagenesis schemes are possible, albeit far more costly in the mouse due to the slower generation times, small litter sizes, inability to easily observe and manipulate embryos at all stages, and the large genome size. Chemical mutagenesis has been the usual method of choice in the mouse with the most effective germ cell mutagens being N-ethyl-N-nitrosourea (ENU)

34

and chorambucil (Russell W. *et al.*, 1979; Hitotsumachi *et al.*, 1985; Russell L. *et al.*, 1989; reviewed by Rinchik, 1991). X-ray irradiation is also widely used (Russell, W., 1951; Searle, 1974).

X-rays and chlorambucil most often cause large lesions, such as deletions and translocations that can involve multiple genes, whereas ENU treatment is frequently associated with intragenic mutations, such as point mutations (Rinchik and Russell L., 1990; Rinchik *et al.*, 1990). After treatment of male mice with either chemical, mutation rates per locus are estimated at 13-150 x 10^{-5} per gamete, compared with a spontaneous mutation rate of 0.5-1.0 x 10^{-5}, while X-rays are about 10-fold less efficient (Rinchik, 1991; reviewed by Rossant and Hopkins, 1992).

The chemical saturation mutagenesis approach coupled with a breeding protocol and an entirely phenotype-driven screen (Brown S. and Peters, 1996) works best with dominant mutations, or with those that are recessive but viable, due to the lack of a pre-existing locus tag. Determining the primary function of genes in development from their mutant phenotype remains complex, however, and finding recessive lethal mutations (into which category fall many developmental loci), is a huge undertaking. Positional cloning of mutations by walking from closely linked markers by using yeast or bacterial artificial chromosomes (YACs or BACs), or phage clones (P1s) that contain murine chromosomal DNA, and then identifying the gene of interest from the mapped interval is a labor-intensive task. Large-scale ENU mutagenic screens have been accomplished for specific chromosomal regions of the mouse genome. For example, in the *t* complex of chromosome 17 new recessive embryonic lethal mutations have been identified, aided by a number of linked markers and an inversion that prevents recombination (Shedlovsky *et al.*, 1988; King *et al.*, 1989). A more powerful approach utilizing a deletion stock of mice for breeding with ENU mutagenized males has identified new lethal loci in the region surrounding the *albino (c)* locus of chromosome 7 (Rinchik, 1991). Other large-scale ENU mutagenesis experiments have screened for mutations that cause specific phenotypes, such as cataracts (Favor *et al.*, 1990), phenylketonuria (Shedlovksy *et al.*, 1993) and circadian behaviour (Vitaterna *et al.*, 1994).

Scaling up saturation screens to a genome-wide basis for the mouse is clearly a formidable task, considering the haploid complement of twenty chromosomes, compared with four in *Drosophila*. Development of dense genetic linkage maps for the mouse and the positioning of simple sequence length polymorphisms (SSLPs) or

35

microsatellite markers has been greatly advanced in the past decade (Love *et al.*, 1990; Copeland *et al.*, 1993; Dietrich *et al.*, 1996). With the development of YAC-based physical maps of the mouse genome and databases approaching saturation for expressed sequence tags (ESTs), the ease of positionally cloning genes mutated in saturation screens is improving, and renders such programs in the mouse more cost- and time-effective (reviewed by Bedell *et al.*, 1997).

1.5.3 Insertional mutagenesis

The introduction of exogenous DNA sequences, although a less efficient means of mutagenesis than the chemical methods discussed above, has the attractive advantage of providing a molecular tag to simplify the cloning of mutant loci. Insertional mutations have been introduced to the germline of mice by a number of routes including direct microinjection of cloned DNA into the pronucleus of zygotes, retroviral infection of early embryos or ES cells, and by electroporation of DNA into ES cells. Approximately 5-15% of transgenic lines and retroviral insertions cause a recessive mutation of an endogenous gene, which results in either embryonic lethality or a visible mutant phenotype in the mouse (reviewed by Gridley *et al.*, 1987; Jaenisch, 1988; Hogan *et al.*, 1994).

Pronuclear injection of cloned DNA has been used to generate transgenic mice for a variety of purposes (reviewed by Jaenisch, 1988), and prior to the advent of ES cells, has been the most commonly used technique for generating mutant mouse lines (Gordon *et al.*, 1980; Brinster *et al.*, 1981; Costantini and Lacy, 1981; Gordon and Ruddle, 1981; Harbers *et al.*, 1981; Wagner E. *et al.*, 1981; Wagner T. *et al.*, 1981). A large number of mutant loci generated by DNA microinjection have been cloned, including previously unindentified loci and those that are allelic with classical mutations (Meisler, 1992). However, analysis of the primary molecular defect that caused the mutant phenotype has often been complicated by deletions, duplications and rearrangements that frequently occur at the site of transgene integration. In addition, microinjected DNA often integrates as multiple copies into random sites within the genome, making it difficult to isolate single copy flanking sequences (reviewed by Wilson *et al.*, 1990). Characterization of mutations using pronucleus injected transgenes as molecular tags has consequently proven both difficult and time-consuming, and accounts for the

handful of new genes identified by this route (Woychik *et al.*, 1985, 1990; Singh *et al.*, 1991; Hodgkinson *et al.*, 1993; Brown A. *et al.*, 1995).

In contrast to the complex integration events that may occur following microinjection of transgenes, integration of replication deficient retroviruses leads to a short direct duplication of host sequences at the site of the single proviral insert, but does not result in gross rearrangements in the host genome (reviewed by Jaensich, 1988). While retroviruses can be introduced to the germline by infection of embryos at various developmental stages (Jaenisch, 1976; Jaenisch *et al.*, 1981), and isolating the flanking host sequences of a proviral insert is more straightforward than for transgenes, the main drawbacks for their use are a size limitation for transfected DNA and problems of reproducibly expressing the transduced gene in the animal (reviewed by Jaenisch, 1988). There is also some evidence that retrovirus integration is not entirely random, but occurs preferentially into regions close to DNase I hypersensitive sites and actively transcribed regions of chromatin (Vijaya *et al.*, 1986; Rohdewohld *et al.*, 1987; Scherdin *et al.*, 1990). Although new mutations have been identified and characterized following retroviral infection of early embryos (Schnieke *et al.*, 1983; Harbers *et al.*, 1984; Soriano *et al.*, 1987; Weiher *et al.*, 1990; Gridley *et al.*, 1991), screening for mutated genes generated by both this route and by pronuclear microinjection has been greatly limited by the number of insertional events that can be analysed, and the accurate and reproducible expression of heterologous genes in transgenic animals.

The ability to generate large numbers of ES cell transformants and pre-select specific mutations based on *in vitro* criteria underscores the potential use of ES cells in a large-scale screen for genes important in mouse development. Exogenous DNA can be introduced into ES cells by electroporation, lipofection and retroviral infection, allowing the *in vitro* selection for clones of a specific phenotype, whilst still retaining the ability to contribute to all somatic tissues and germ cells of the developing embryo (Gossler *et al.*, 1986; Robertson *et al.*, 1986; Hooper *et al.*, 1987; Kuehn *et al.*, 1987). For a specific known locus, inactivation, deletion, or subtle mutation may be introduced to the germline by homologous recombination in ES cells (Doetschmann *et al.*, 1987; Thomas and Capecchi, 1987; Thompson *et al.*, 1989). Candidate mammalian genes for site-directed ("targeted") loss-of-function analysis are chosen on the basis of phenotypes for their homologs in other organisms, or according to a pre-existing expectation of possible involvement in known biological processes (reviewed by Capecchi, 1989;

Kessel and Gruss, 1990; Wolpert *et al.*, 1998). Just about any kind of desired mutation can now be introduced into a mouse gene or genome including null point or conditional mutations, as well as complex chromosomal rearrangements such as large deletions, translocations or inversions (Capecchi, 1989; Hasty *et al.*, 1991; Valancius and Smithies, 1991; Wu *et al.*, 1994; Ramirez-Solis *et al.*, 1995). Although this strategy has successfully disrupted many hundreds of genes (Brandon *et al.*, 1995), preparing targeting constructs as well as isolating and mapping genomic clones, and generating transgenic mice for each candidate gene involves time- and resource-consuming work. Homology approaches are also limited by specific sequence motifs and a bias towards abundantly expressed genes in cDNA libraries, which rarely provide good representation of mRNAs in tissues that are difficult to obtain, such as those of the mammalian embryo.

The occurrence of early embryonic lethality with some germline mutations has made it impossible to study the effects of mutations at late stages of embryonic development, or in the adult. However, it is now possible to generate targeted somatic mutations that are tissue specific and/or inducible at different stages in embryonic or adult development, which address this limitation. One approach in use is the Cre-lox system for site-specific recombination (Lasko *et al.*, 1992; Orban *et al.*, 1992; Gu *et al.*, 1993, 1994). In this system, the *loxP* recognition sequences for P1 Cre recombinase are introduced by homologous recombination in ES cells to sites flanking a genomic domain of interest. Transgenic mice containing the desired *loxP* sites are crossed to transgenic mice that express Cre recombinase under the control of tissue-specific or inducible regulatory elements. Recombination occurs at the *loxP* sites when Cre is expressed, so that the intervening sequences are deleted and the resulting mutation occurs in a lineage-specific or conditional manner (reviewed by Chambers, 1994). This system has also been used to induce large deletions, specific inversions, duplications and translocations in the mouse (Ramirez-Solis *et al.*, 1995; Van Deursen *et al.*, 1995). A similar strategy for tissue-specific DNA recombination involves using yeast FLP recombinase and recombinase-specific flanking sequences (O'Gorman *et al.*, 1991).

The efficiency of insertional mutagenesis has been increased with the development of entrapment vectors, which rely on the *cis*-acting regulatory sequences of an expressed endogenous gene for the activation of a reporter gene (reviewed by Skarnes, 1990). Insertions into or near active transcriptional units create random mutations within the

tagged gene and allow its expression to be followed when the transcription orientation and translation frame of the reporter gene are matched correctly. In this way, mutations for novel genes that show interesting spatial or temporal expression patterns during embryogenesis can be detected.

The entrapment strategy was originally devised to detect and identify transcriptionally active endogenous operons in bacteria using the *Eschericia coli* reporter gene *lacZ* within a transposable element of the bacteriophage Mu genome (Casadaban and Cohen, 1979). *lacZ* encodes the enzyme β-galactosidase (βgal) which provides an easily detectable blue product in eukaryotic cells when exposed to the substrate 5-bromo-4 chloro-3-indolyl-β-D-galactoside (X-gal). The use of modified vectors from which the translation initiation codon of the reporter had been deleted, therefore requiring gene fusion for translation initiation and expression, demonstrated that the βgal protein can accommodate large amino-terminal fusions whilst retaining its enzymatic activity (Casadaban *et al.*, 1980). A similar approach using P-element transposon vectors has been used to detect developmentally regulated genes in *Drosophila*, especially those that direct nervous system development, based on expression patterns rather than their mutant phenotypes (O'Kane and Gehring, 1987; Bellen *et al.*, 1989, 1990; Bier *et al.*, 1989; Wilson *et al.*, 1989). For the many thousands of lines generated in these studies, activation of *lacZ* in a spatially and temporally restricted pattern during embryogenesis has been observed in about 65% of P-element insertions. As well as identifying valuable markers for cells, tissues and organs, approximately 15% of the insertions also caused recessive mutations that resulted in visible phenotypes, presumably as a consequence of the disruption of genes (Cooley *et al.*, 1988; Bier *et al.*, 1989; Wilson *et al.*, 1989).

Similar developmental activation of *lacZ* entrapment constructs has been shown to occur in transgenic mice generated by microinjection into the zygote (Allen *et al.*, 1988; Kothary *et al.*, 1988). Using a *lacZ* reporter gene under the control of a weak herpes simplex virus *thymidine kinase* (HSV*tk*) promoter, five of twenty transgenic lines were found to exhibit unique restricted patterns of *lacZ* expression during development (Allen *et al.*, 1988). For one of seven transgenic lines generated with a construct using a truncated version of the mouse heat shock protein *Hsp68* gene, a restricted pattern of *lacZ* expression was detected in the developing spinal cord (Kothary *et al.*, 1988). However, it was not for some years that the interrupted allele was cloned from DNA

39

sequences flanking the site of insertion and identified as the classical mouse neurological mutant *dystonia musculorum (dt)*, (Brown A. *et al.*, 1995). The design of entrapment vectors for introduction into mouse ES cells has since provided an effective strategy for the large-scale screening of randomly generated and selectable mutations for developmental genes of interest (reviewed by Skarnes, 1990; Joyner *et al.*, 1992; Takeuchi, 1997; Evans, 1998; Voss *et al.*, 1998a). This mutagenesis and cloning strategy provides the basis for much of the research undertaken in this study and is reviewed in greater detail in the following section (1.6).

1.6 ENTRAPMENT TECHNOLOGY IN MOUSE ES CELLS

Coupled with the use of ES cell technology the entrapment strategy allows pre-selection for integration events and reporter activity *in vitro* prior to producing mouse lines to assess *in vivo* consequences of mutations (Gossler *et al.*, 1989; Friedrich and Soriano, 1991). Screening for restricted expression patterns is anticipated to lead to the identification of developmentally regulated genes of functional significance, including those which would not be identified in conventional screens for mutant phenotypes, for example, genes which are made redundant by the functional substitution of a gene product from a second locus (Cooke *et al.*, 1997). Entrapment vectors fall into two functional groups, the "enhancer trap" vectors and the "promoter trap" vectors, commonly referred to as "gene trap" vectors. These two classes of vectors differ in the elements required from the endogenous locus in the production of reporter gene activity.

1.6.1 Enhancer traps

Enhancer trap vectors are designed with a minimal promoter element placed upstream of a reporter gene (Figure 1.4A), with this promoter remaining silent unless the vector inserts near an endogenous enhancer element. Enhancer trap insertions do not create fusion genes and need not insert inside a transcription unit to be activated by an enhancer. This approach was first developed to capture cellular enhancer elements capable of regulating tissue-specific expression from a distance independent of their orientation relative to transcription, and which were capable of acting upon heterologous promoters (Fried *et al.*, 1983; Weber *et al.*, 1984; Hamada, 1986a, 1986b). It was enhancer trap vectors that were used to generate the entrapment mutations described in section 1.5.3, for *Drosophila* and microinjected mouse zygotes.

40

Figure 1.4: Design and mechanisms of reporter gene activation in entrapment vectors. (A) The enhancer trap vector consists of a minimal promoter element (*P), including the TATA box for correct initiation of transcription, fused to a reporter gene which has its own translation initiation start site (ATG). The enhancer element (e) of an endogenous gene X is postulated to activate transcription for both the endogenous gene and reporter gene. The promoter trap vectors (B-D) rely on endogenous *cis*-acting enhancer (e) and promoter (P) elements for activation and expression of the reporter gene resulting in the generation of a novel fusion transcript with the host gene X. Reporter activity is detected when the insertion is in the correct orientation and the reading frame of the endogenous ATG is maintained. Insertions in 5'-noncoding regions can also be detected when the reporter gene includes its own ATG. (B) Exon traps contain only the coding sequences of the reporter gene and require integration into an exon of an endogenous gene for generation of a fusion transcript. (C) Poly A traps are designed to provide selectable antibiotic resistance (neo) under the control of a constitutive promoter where they acquire an endogenous polyadenylation site (pA). Integration into an endogenous transcription unit further provides for expression of a reporter gene. (D) Gene traps contain a splice acceptor site (SA) fused to the 5' end of the reporter. Integration of the vector within an intron of an endogenous gene results in the generation of a spliced fusion transcript. Insertions can also be detected in exons, in which case the endogenous SA and the introduced SA are competing.

A. Enhancer trap vector

ATG
reporter pA

e — P gene X

ATG
e — P reporter pA — P gene X

transcription →A(n)

translation

reporter protein   gene X protein

B. Exon trap vector

(ATG)
reporter pA

e — P gene X

e — P gene X — reporter pA

(ATG)
transcription →A(n)

translation   fusion protein
                       or
                       reporter protein if insertion is
                       5' of endogenous ATG and
                       vector ATG is used

C. Poly A trap vector

transcription

translation

fusion protein
(high frequency of
reporter gene activity)

neo protein
(antibiotic resistance)

D. Gene trap vector

transcription

translation

fusion protein

or

reporter protein if insertion is
5' of endogenous ATG and
vector ATG used

The use of an entrapment vector for electroporation into ES cells was first reported for a construct consisting of a minimal *Hsp68* promoter driving the *lacZ* gene upstream of the bacterial *neomycin phophotransferase* gene for neomycin resistance (*neo*) under the control of a HSV*tk* promoter (Gossler *et al.*, 1989). Although this study and the earlier studies for zygote injections have shown a high frequency of *lacZ* activation following integration into the mouse genome, the enhancer trap approach has proven less successful in the mouse than in *Drosophila*. This is largely due to the large distances over which enhancers may act, making characterization of the insertion site difficult in the more complex mammalian genome. For two novel sequences cloned from mouse ˜ʳ ʳell enhancer trap insertions, expression of the endogenous genes has been shown to be more widespread than the corresponding βgal activity, suggesting that the reporter gene responded only to a subset of the endogenous regulatory elements (Soininen *et al.*, 1992; Neuhaus *et al.*, 1994). As well, neither integration showed a homozygote phenotype following germline transmission, reflecting the fairly low probability for such an event as enhancer trap integrations are expected to leave the endogenous promoter and coding sequence intact.

1.6.2 Promoter traps

Greater success has been obtained with the promoter traps, and in particular its gene trap subclass, than with the enhancer traps (reviewed by Skarnes, 1990; Joyner *et al.*, 1992; Takeuchi 1997; Evans, 1998; Voss *et al.*, 1998a). The promoter traps are designed to generate fusion transcripts between the reporter gene and the endogenous gene present at the site of integration (Brenner *et al.*, 1989; Gossler *et al.*, 1989; Kerr *et al.*, 1989; Friedrich and Soriano, 1991). Lacking any promoter element of its own, activation of the reporter gene requires the integration of the construct within an active transcriptional unit to create a novel fusion transcript with the endogenous mRNA.

These DNA integrations occur randomly within the mouse genome (Chowdhury *et al.*, 1997) and are predicted to be highly mutagenic, as by definitior they interrupt the normal endogenous coding sequence. For several promoter trap studies to date the mutation of an endogenous gene at the site of integration has given rise to an overt mutational phenotype in up to half of the number of trapped lines analysed (Friedrich and Soriano, 1991; Skarnes *et al.*, 1992, 1995; von Melchner *et al.*, 1992; Chen Z. *et al.*, 1994; DeGregori *et al.*, 1994; Takeuchi *et al.*, 1995; Camus *et al.*, 1996; Forrester *et al.*,

1996; Serafini *et al.*, 1996; Hicks *et al.*, 1997; Couldrey *et al.*, 1998; McClive *et al.*, 1998; Stoyakova *et al.*, 1998; Voss *et al.*, 1998a). Furthermore, cloning a portion of the endogenous gene directly from the reporter fusion transcript facilitates a more rapid identification of the mutated genes than the previously described methods (Gossler *et al.*, 1989; Friedrich and Soriano, 1991). Reporter gene expression patterns have also been found to faithfully reflect expression of the endogenous trapped gene in several screening studies (Skarnes *et al.*, 1992, 1995; Takeuchi *et al.*, 1995; Chen J. *et al.*, 1996; Torres *et al.*, 1997). This, together with the aforementioned properties make the promoter trap vector strategy well suited for a large-scale genetic screen for developmentally important genes. Characterization of promoter trap mutations generated from several studies has revealed a range of tissue and temporally restricted expression profiles for both known and novel genes. Examples for such entrapment mutations are listed in Table 1.1.

Promoter traps can be sub-divided into three groups, exon traps, polyadenylation (poly A) traps and gene traps (Figure 1.4B-D). Exon trap vectors (sometimes referred to as "promoter traps") are equivalent to the previously described bacterial promoter traps (Casabadan and Cohen, 1979) and contain only the coding sequences of a reporter gene. To produce reporter activity, exon traps require integration into the coding region of an endogenous gene in the correct orientation and reading frame. Insertions in 5'-noncoding regions can also be detected when the reporter gene has its own initiation of translation codon (ATG), (von Melchner and Ruley, 1989; Friedrich and Soriano, 1991; Macleod *et al.*, 1991; Reddy, 1991). Poly A trap vectors are designed to allow access to genes that are not expressed in ES cells by the use of a selectable marker gene which is under the control of a constitutive promoter and is devoid of a poly A addition signal sequence. Poly A traps require the acquisition of an endogenous poly A site for isolation of transformed ES cell clones, in addition to trapping a transcription unit for expression of the reporter gene. Use of such traps has led to a 2.5- to 5-fold increase in the frequency of trap events over vectors containing a poly A site, although large deletions or rearrangements spanning more than 10 kb in the 3'-flanking regions associated with some lines have made it difficult to identify the genes responsible for mutations (Niwa *et al.*, 1993; Yoshida *et al.*, 1995). However, the ability of poly A traps to yield a high frequency of integrations into transcription units that are not necessarily active in undifferentiated ES cells is likely to make them very useful when

42

combined with an *in vitro* system, directed to specific pathways (Salminen *et al.*, 1998; Zambrowicz *et al.*, 1998; Ishida and Leder, 1999).

Promoter trap	Gene	Gene product	Homozygous phenotype	Reference
IRESβgeo	*Bodenin*	unknown/novel gene	none	Faisst & Gruss, 1998
pGTi	*Gt/2*	unknown/novel gene	dwarfism – maternal	S-Gossler *et al.*, 1998
PT1-ATG	*R.140*	unknown/novel gene	lethal/embryonic	Forrester *et al.*, 1996
PT1-ATG	*Aquarius*	unknown/novel gene	none	Sam *et al.*, 1998
pGT4.5	*cordon-bleu*	unknown/novel gene	none	Gasca *et al.*, 1995
pGT4.5	*Gt4-1*	unknown/novel gene	lethal/perinatal	Skarnes *et al.*, 1992
pGT1.8geo	*Taube nuss*	founding TBN protein	lethal/preimplantation	Voss *et al.*, 2000
pGT1.8geo	*MAP-4*	microtubule assoc prt	none	Voss *et al.*, 1998b
pGT1.8geo	*αE-catenin*	cell-cell adhesion	lethal/preimplantation.	Torres *et al.*, 1997
pGT1.8TM	*PTPₖ*	protein-tyrosine phos.	none	Skarnes *et al.*, 1995
pGT1.8TM	*LAR*	protein-tyrosine phos.	none	Skarnes *et al.*, 1995
pGT1.8TM	*Netrin-1*	axon guidance cue	not germline	Skarnes *et al.*, 1995
TV2	*Jumonji*	unknown/novel gene	lethal/<E15.5 dpc	Takeuchi *et al.*, 1995
ROSAβgeo	*α-Enolase*	glycolytic enzyme	lethal/early postimplant.	Couldrey *et al.*, 1998
ROSAβgeo	*TEF-1*	transcription factor	lethal/E11-12 dpc	Chen Z. *et al.*, 1994
ROSAβgeo	*BTF-3*	transcription factor	lethal/early postimplant.	Deng & Behringer, 1995
U3neo	*fug1*	unknown/novel gene	lethal/early postimplant.	DeGregori *et al.*, 1994
U3βgeo	*Eck*	recep. tyrosine kinase	none	Chen J. *et al.*, 1996
U3His	*REX-1*	transcription factor	not germline	von Melchner *et al.*, 1992
pPAT	*PAT-12*	unknown/novel gene	none	Yoshida *et al.*, 1995

Table 1.1: Characterized mutations identified in promoter trap studies. **Listed are examples for some of the genes, both novel and known, that have been identified by characterization of promoter trap insertions in mouse ES cells. The ROSAβgeo and U3 constructs are retroviral gene traps, pPAT is a poly A trap, and the remainder are plasmid-based gene traps.**

1.6.3 Design and delivery of a gene trap vector

The essential feature of a gene trap vector is the placement of a splice acceptor site upstream of the reporter. Gene trap insertion is typically within an intron, creating a spliced fusion transcript and detectable reporter gene activity where the transgene maintains the reading frame of the endogenous intiation codon – a one in three event (Gossler *et al.*, 1989; Friedrich and Soriano, 1991). Again, if the reporter includes an ATG codon insertions can be detected in 5'-untranslated regions. Gene traps may also

43

be detected in exons, in which case the endogenous and introduced splice acceptor sites are in competition (McClive *et al.*, 1998; Voss *et al.*, 1998a).

In the first study to report the use of a gene trap vector, the *lacZ* reporter gene was placed downstream of the splice acceptor consensus sequence derived from the mouse *Engrailed-2 (En-2)* gene and introduced into ES cells to identify actively transcribed genes in these cells. Generation of chimaeric animals subsequently gave information as to the temporal and spatial pattern of expression of the locus of integration (Gossler *et al.*, 1989). A more advanced gene trap was developed consisting of a *lacZ* and *neo* gene fusion which provides a single bifunctional protein with both βgal and neo resistance activities (Friedrich and Soriano, 1991). This reporter/selectable marker gene, termed *βgeo*, was linked to the adenovirus splice acceptor and successfully used in ES cells to trap loci which were developmentally regulated in resultant transgenic mouse lines (Friedrich and Soriano, 1991). *βgeo* fusion protein expression from trapped loci allows not only for visualization of βgal activity but also for selection of ES cell clones harboring functional gene trap insertions by exposing ES cell cultures transfected with the vector to the selective agent G418. While this study demonstrated that 95% of the neomycin resistant cells were also *lacZ* positive, limitations of correct reading frame insertion and a proportion of non-*lacZ* positive gene trap clones were evident. One explanation for the *neo* positive/*lacZ* negative clones identified is inhibition of lacZ protein tetramere formation (required for lacZ activity) by amino terminal fusion sequences originating from the endogenous gene (Friedrich and Soriano, 1991).

Similarly, vectors for the reporter fusion protein of hygromycin and βgal (Natarajan and Boulter, 1995), and for phleomycin and βgal (Camus *et al.*, 1996) have been shown to work effectively in ES cells. The application of more sensitive reporter systems such as the vital marker green fluorescent protein are likely to further enhance the use of gene trapping strategies (Zernicka-Goetz *et al.*, 1997; Ishida and Leder, 1999; Zheng and Hughes, 1999).

In order to access a larger pool of detectable gene trap events with a high correlation between selection and reporter activities, conventional gene trap vectors have been modified to increase the probability for translation of the reporter occurring in the correct reading frame. One strategy for achieving this has been to introduce a picornaviral internal ribosome entry site (IRES) element between the splice acceptor

44

and the reporter gene, providing position and reading frame independent translation of the *βgeo* coding sequence (Mountford *et al.*, 1994). IRES elements were first isolated from the 5'-untranslated region of the naturally uncapped picornaviral mRNAs (Pelletier and Sonenberg, 1988). The IRES element from the encephalomyocarditis virus (EMCV) was shown to act as a ribosome landing pad, allowing the internal cap-independent initiation of translation of a selectable marker or reporter gene product in mammalian cells (Jang *et al.*, 1988; Pelletier and Sonenberg, 1988; Jang and Wimmer, 1990; Ghattas *et al.*, 1991; Molla *et al.*, 1992). The EMCV-IRES element has been widely used for the construction of bicistronic vectors in mammalian transgenesis (reviewed by Mountford and Smith, 1995), and applied successfully for the expression of a *lacZ* reporter throughout whole mouse embryos and tissues following transfection of ES cells (Kim *et al.*, 1992).

As *βgeo* fusions to some classes of genes has been shown to cause a loss of βgal activity (Skarnes *et al.*, 1995), the translation of *βgeo* from an IRES, independent of the host gene, is predicted to detect gene classes that might be absent or under-represented with conventional gene trapping. A large-scale study reporting the use of an IRES*βgeo* gene trap vector suggests that this vector integrates randomly, without any apparent bias, and is able to capture a wide range of genes expressed in a wide variety of tissues and developmental stages, as well as trapping genes expressed at very low levels in ES cells (Chowdhury *et al.*, 1997). The number of G418 resistant colonies detected in this study was increased 3-fold, with an overall enrichment of about 15-fold for the number of detectable gene trap events, when compared with a conventional *βgeo* gene trap vector (Friedrich and Soriano, 1991). As well, use of the IRES*βgeo* vector allowed the production of chimaeric mice and germline transmission at the same efficiencies as for the conventional vector (Chowdhury *et al.*, 1997; Bonaldo P. *et al.*, 1998).

Other reported strategies to overcome reading frame constraints for detection of the reporter are to use the Moloney murine leukemia virus splice acceptor from the *env* gene which splices in all reading frames (Schuster-Gossler *et al.*, 1994), or to include an ATG upstream of the *lacZ* sequence which has been demonstrated to yield a 3-fold increase in the number of βgal-expressing and G418 resistant colonies compared with a vector lacking the initiation ATG (Hill and Wurst, 1993). The ATG vector is predicted to detect integrations into 5'-untranslated regions and has been shown to allow the identification of out-of-frame integrations as well as insertions into exons (McClive *et*

al., 1998). While the discussed strategies will generate more reporter-expressing clones, it is important to consider that these are still subject to the limitations of ease of cloning and the question of whether they truly represent endogenous expression for genes of interest.

Both plasmid and retroviral vectors have been successfully used for the introduction of promoter vectors to ES cells (Gossler *et al.*, 1989; Friedrich and Soriano, 1991; von Melchner *et al.*, 1992). Although retroviruses provide the advantages of efficient integration into the genome with a defined vector structure and little rearrangement at the site of insertion (section 1.5.3), their use in promoter trapping may hold some limitations. Inaccurate reflection of endogenous expression for a mutated gene by the proviral reporter has been observed (Deng and Behringer, 1995), which has generally not been the case following electroporation of plasmid traps. This might be associated with a repression of transcription due to *de novo* methylation of the provirus extending into flanking sequences (Jähner and Jaenisch, 1985). It has also been observed that retroviral entrapment vectors tend to integrate near the 5'-end of genes yielding short sequences following 5'-rapid amplification of cDNA ends (5'-RACE), and often in 5'-untranlsated regions (Friedrich and Soriano, 1991; Baker *et al.*, 1997). While there are several reports suggesting that most of the genome is accessible to retroviral integration (Chang *et al.*, 1993; Withers-Ward *et al.*, 1994), as alluded to earlier (section 1.5.3) there is some evidence that integration events may not be entirely random, with a subset landing in "hotspots" (Shih *et al.*, 1988; reviewed by Craigie, 1992). However, the examples given in section 1.6.5 for repeated trapping of the same gene in independent studies do not suggest a prejudice due to the transformation process used.

1.6.4 Screening for genes of interest

Creating insertional mutations in ES cells by entrapment techniques, holds the major benefit of being able to screen and pre-select ES cell lines for the generation of germline transmitting chimaeras on the basis of sequence analysis for the disrupted gene, as well as by the transgenic expression profile of the reporter gene *in vitro*. Genes disrupted by gene trap and poly A trap vectors can be efficiently identified from fusion transcripts by 5'- and 3'-RACE respectively, while those genes harboring exon trap insertions can be identified using plasmid rescue of genomic DNA flanking the insertion site (Frohman *et al.*, 1988; Chowdhury *et al.*, 1997; Hicks *et al.*, 1997; Townley *et al.*, 1997; Voss *et al.*,

1998a; Zambrowicz *et al.*, 1998). Conventional entrapment vectors have been used successfully by several laboratories to "tag" and inactivate cellular genes that are expressed in both undifferentiated ES cells and in restricted patterns in developing embryos and the entire animal, rather than more ubiquitously in homogeneous or heterogeneous populations of cells *in vivo* or *in vitro* (Gossler *et al.*, 1989; Friedrich and Soriano, 1991; Korn *et al.*, 1992; von Melchner *et al.*, 1992; Skarnes *et al.*, 1992, 1995; Wurst *et al.*, 1995; Scherer *et al.*, 1996). Tailoring of trap designs such as those in which the *neo* transcription is under the control of a constitutive promoter (Skarnes *et al.*, 1992; Wurst *et al.*, 1995; Forrester *et al.*, 1996), and the previously discussed poly A traps (section 1.6.2), are finding much broader applications where they allow for selection of cell populations in which a trapped locus is either active or inactive.

In one approach using the *βgeo* reporter, a "secretory trap" was developed to identify insertional mutations in genes encoding both secreted and type I membrane-spanning proteins, a class of genes previously missed by conventional gene trap vector (Skarnes *et al.*, 1995). Vectors were constructed to express the N-terminal cleavable signal sequence as well as the transmembrane domain of CD4 type I membrane protein fused to the *βgeo* reporter, and ES cells selectively pre-screened prior to chimaera generation. This study demonstrated the ability to detect genes expressed at very low levels in ES cells, and indicates the feasibility of designing such vectors for identifying genes which express specific classes of gene products, such as secreted proteins, upon ES cell differentiation. In a similar manner, a retroviral gene trap vector including a human placental alkaline phosphatase reporter gene has been used in conjunction with an embryoid body screen to select for genes which are expressed in a highly restricted manner during early postimplantation development (Xiong *et al.*, 1998). Recent reports for the use of gene trap constructs incorporating the Cre-mutated lox system for reversible site-specific expression (section 1.5.3) demonstrate the future feasibility for analyzing specific Cre-expressing alleles of interest (Araki *et al.*, 1999; Michael *et al.*, 1999; Hardouin and Nagy, 2000).

Several reports have emerged in recent years demonstrating the feasibility of screening for gene trap insertions into genes that are expressed after induction by specific differentiation protocols, and/or in specific cell lineages. For example, genes trapped in ES cells with the PT1-ATG gene trap vector (Hill and Wurst, 1993) were pre-screened for *in vitro* responsiveness of the *lacZ* reporter to retinoic acid, and in all but one line of

47

20 (9 induced and 11 repressed), showed unique spatially or tissue-specific patterns of expression between 8.5 and 11.5 days of embryogenesis (Forrester *et al.*, 1996). A large proportion of integrations that were repressed by retinoic acid in this study (8/11) were also found to be expressed in the developing heart *in vivo*. Similarly, ES cell clones trapped with an IRESβgeo vector (Chowdhury *et al.*, 1997) showed a strong enrichment for restricted patterns *in vivo*, following pre-selection for responsiveness *in vitro* to NGF, retinoic acid and Follistatin in order to enrich for genes that may play a role in the control of neural development (Bonaldo P. *et al.*, 1998; Stoykova *et al.*, 1998).

Further reports for pre-selection strategies include *in vitro* differentiation in suspension culture followed by selection for trapped clones repressed during exposure to hematopoietic growth factors in methylcellulose cultures (Muth *et al.*, 1998), *in vitro* differentiation in suspension culture followed by immunohistochemical examination for neuron-specific proteins (Shirai *et al.*, 1996), extended culture under G418 and low LIF concentration for restricted expression patterns expressed in aspects of the nervous system (Voss *et al.*, 1998a), and rapid direct sequence analysis of 5'-RACE products to identify those clones yielding correctly spliced single integration fusion sequences for gene classes of interest (Townley *et al.*, 1997). One study reports a comprehensive strategy which combines *in vitro* differentiation in suspension cultures exposed to retinoic acid, and pre-selection on the basis of both *in vitro* βgal staining and *in vivo* staining for co-expression of βgal and cell lineage markers of interest. This screen further includes 5'-RACE sequence analysis for novel genes, and analyses for gene trap insertion copy number, alternate splicing events and replacement of the endogenous transcript by gene trap transcript (Baker *et al.*, 1997).

These types of strategies may allow the detection of genes that are silent in undifferentiated ES cells but are activated at later stages of development, although genes with spatially restricted expression profiles are unlikely to be identified during morphogenetically disorganized ES cell differentiation *in vitro*. Co-localization of reporter gene expression with specific lineage markers, and scaling up induction gene trap screens using cytokines and growth factors at different time frames during embryoid body differentiation, will allow gene trap integrations in particular cell types or lineages to be targeted. Interestingly, the gene trap approach has been used to isolate genes in other cell types, for example genes involved in the differentiation of the

48

myeloid progenitor cell line (Jonsson *et al.*, 1996), and in the growth and fusion of skeletal myoblasts (Gogos *et al.*, 1996).

1.6.5 Perspectives and limitations for gene trapping

Gene trapping in ES cells has shown a great deal of application in the past decade, however, molecular analyses of the gene trap clones described frequently highlight limitations of the strategy for large-scale generation of gene trap mouse mutants.

The insertion of gene trap vectors into the genome of cells has been demonstrated to occur in a random manner detecting all major classes of genes (Chowdhury *et al.*, 1997), and as such should not be influenced by the relative chromosomal location of the gene in ES cells. However, deliberate and in some instances unforeseen bias for specific classes of genes can be imposed by differences in vector design and selection strategy. For example, gene trapping imposes an inherent bias in favor of detecting genes composed of large intronic regions as non-coding intron sequences make up the vast majority of the size of genes, and are the likely sites of integration of molecular reporters integrating at random (Casadaban and Cohen, 1980; Chu and Sharp, 1981; Weber *et al.*, 1984). It could be argued that a complete gene trap screen strategy would need to include an exon trap version of the vector of interest in order to recover insertions in smaller transcription units composed of few or no introns for the gene class of interest. A further simple example of bias is the selection for higher expressing genes by increasing the antibiotic concentration in selection media.

The accessibility of the specific gene loci in ES cells in culture could also potentially influence the number of gene loci that are susceptible to integration of a gene trap vector. The repeated trapping of some genes by entirely independent groups using different constructs does suggest the existence of integration "hot spots" (reviewed by Evans, 1998). Gene trap integrations in the *Jumonji* gene were first described for an ES cell clone trapped with a retroviral *neo*IRES*lacZ* vector (Takeuchi *et al.*, 1995), then subsequently by other laboratories using the retroviral ROSAβ*geo* gene trap (Friedrich and Soriano, 1991; Baker *et al.*, 1997) and the plasmid β*geo* vector pGT1.8*geo* (Skarnes *et al.*, 1995; Voss *et al.*, 1998a). Further examples of such reports include trapping *R-PTP-κ* with both the IRESβ*geo* gene trap (Chowdhury *et al.*, 1997) and the pGT1.8TM secretory trap (Skarnes *et al.*, 1995), and trapping *α-Enolase* with IRESβ*geo* (Chowdhury *et al.*, 1997) as well as with ROSAβ*geo* (Couldrey *et al.*, 1998).

Interestingly, the gene *Tiam-1* has been trapped in ES cells (Voss *et al.,* 1998a) as well as in a gene trap screen in BW5147 T-lymphoma cells (Habets *et al.,* 1994). Such genomic hotspots are likely to include regions with more open chromatin conformation, making these areas more accessible to vector insertion. It also remains possible that there is some degree of homology between these genomic loci and vector sequences, although there is little hard evidence at present to support this possibility.

While gene-trap screens can provide important clues as to the function of a gene based on expression patterns, some caution is required in selecting gene trap events by virtue of the βgal staining pattern, which is not in all cases a true reflection of endogenous gene expression. In some cases *in vitro* βgal expression has appeared to be more restricted than the expression of the endogenous gene (Skarnes *et al.,* 1992; Deng and Behringer, 1995; Voss *et al.,* 1998a). Conversely, screening for restricted reporter gene expression patterns would miss detecting genes that are widely expressed but developmentally important, as is the case for expression of the *Notch* gene in the *Drosophila* embryo (Hartenstein *et al.,* 1994). Also, expression of the reporter gene could show a false pattern where the integration of the trap vector disrupts the transcriptional regulatory sequence. This possibility underlies the need to routinely confirm expression of the non-disrupted endogenous allele by *in situ* hybridization.

Although by definition promoter trap integrations interrupt the endogenous coding sequence and numbers of published results have demonstrated that both the gene trap and exon vectors are mutagenic at a genomic level, it is clear that these integrations do not always create null alleles in homozygote animals (for example, Gasca *et al.,* 1995; Skarnes *et al.,* 1995; Chen J. *et al.,* 1996; Forrester *et al.,* 1996; Faisst and Gruss, 1998; McClive *et al.,* 1998; Muth *et al.,* 1998; Sam *et al.,* 1998; Voss *et al.,* 1998a, 1998b). An insertion into the 3'-end coding region of a gene may not necessarily disrupt the function of a trapped gene, while disruption of genes that are not developmentally important or have functional redundancy might not cause phenotypic abnormalities in any case.

Although an insertion in the novel gene *Cordon-bleu* resulted in an exquisite spatially restricted expression pattern (Gasca *et al.,* 1995), and a retroviral insertion in the receptor tyrosine kinase gene *Eck* caused a severe deficiency in Eck protein in adult mice (Chen J. *et al.,* 1996), overt phenotypes were not observed in homozygote animals in either case. In these and other studies where homozygotes lacked overt phenotypes,

50

the presence of low levels of wild type transcript or protein has been suggested as one possible method of curtailing the mutant gene trap effect. This event is probably the result of alternative splicing around the gene trap integration or a failure to use the construct's poly A signal (Faisst and Gruss, 1998; McClive *et al.*, 1998; Sam *et al.*, 1998; Voss *et al.*, 1998b). In contradiction to these findings, homozygotes carrying an insertion into the axon guidance molecule *Netrin* were found to die at birth and exhibited defects in spinal commissural axon projections, despite the presence of low levels of wild type transcript (Serafini *et al.*, 1996). Similarly, a gene trap insertion in the receptor-linked protein tyrosine phosphatase *LAR* gene caused phenotypic abnormalities in the size of basal forebrain cholingergic neurons and in hippocampal cholinergic innervation, although trace expression of full length *LAR* transcripts was found (Yeo *et al.*, 1997). It is difficult to assess the frequency with which functionally mutated alleles occur in gene trap screens without analysis for wild type transcripts in a large number of non-phenotypic events. However, the occurrence of overt phenotypic abnormalities arising from the fortuitous insertion of a splice acceptor-trap construct into an exon sequence in several studies, has led to the suggestion that while exon trap vectors are less efficient in producing promoter trap ES cell clones *in vitro*, they may be more efficient in producing functionally mutated alleles following integration into an endogenous coding sequence (for example, Skarnes *et al.*, 1995; Forrester *et al.*, 1996; McClive *et al.*, 1998; Voss *et al.*, 1998a).

The relative ease of generating large numbers of fusion transcripts with endogenous genes is evident in a number of large-scale gene trap studies now reported (Wurst *et al.*, 1995; Chowdhury *et al.*, 1997; Hicks *et al.*, 1997; Townley *et al.*, 1997; Voss *et al.*, 1998a; Zambrowicz *et al.*, 1998). The ease of cloning these has, however, been often complicated by the complexity of products obtained. In addition to the aberrant integration events, the sensitivity of 5'-RACE techniques can also yield multiple products which are difficult to interpret. The analysis of more than 150 independent gene trap cell lines by a rapid protocol for direct sequencing of 5'-RACE products demonstrated amplifiable sequence and correctly spliced events for just 37% (57/153) of these lines (Townley *et al.*, 1997). The remaining lines in this study had either made inefficient use of the splice acceptor, yielded two or more sequences 5' of the splice site, or had failed to amplify by RACE probably as a result of deletions in the vector. In some cases, use of a cryptic splice donor site within the vector sequence has been

51

observed (Townley *et al.*, 1997; McClive *et al.*, 1998). As discussed above, alternate splicing events can occur around the gene trap integration making a phenotype less likely in transgenic animals, and it is also possible that the vector sometimes makes use of multiple splice donor sites upstream in the gene or even in a different gene, particularly where tandem copies of the vector are integrated. While the design and delivery of the vector used in each study will determine the likely occurrence of such events, these observations heighten the advantage of using the automated analyses now available for directly sequenced products (Zambrowicz *et al.*, 1998), as well as the *in vitro* pre-screening strategies that combine both sequence analysis and induction screens (Baker *et al.*, 1997).

Although it is not possible to entirely predict the *in vivo* expression pattern of trapped genes from *in vitro* pre-screening of cell lines, the approaches discussed in section 1.6.4 using specific vectors and induction culture conditions demonstrate the potential for efficiently identifying interesting gene trap clones for further analysis. As it has been observed that the gene expression patterns seen in embryoid bodies mimic well those occurring *in vivo* during early development (Scherer *et al.*, 1996), with further advances for defined ES cell *in vitro* pre-selection protocols and improved design for vectors that will disrupt gene function more reliably and allow the use of vital reporters, the gene trap approach is likely to stand as a powerful means for simultaneously identifying and mutating genes of importance in the mammalian developmental processes.

1.7 CONCLUSION

Pluripotential cells derived from the preimplantation mammalian embryo have the capacity to give rise to differentiated progeny representative of the three embryonic germ layers, the germline and some extraembryonic tissues. Current research efforts in developmental biology are rapidly advancing knowledge of the genetic mechanisms controlling early mammalian development, as well as the conditions required to mimic this system *in vitro*. ES cells are currently providing a powerful tool for the identification, genetic modification and functional analysis of genes involved in normal mammalian development, and an enormous potential is now evident for the application of human ES and multipotent stem cell biology towards therapeutic molecular medicine. In order to realize this potential, the fundamental mechanisms regulating ES cell self-renewal and differentiation to more restricted somatic cell lineages will need to be fully

elucidated. Genes which are expressed in ES cells then downregulated as they differentiate and lose competence for multi-lineage commitment, are prime candidates for identifying such mechanisms. As gene trapping strategies in ES cells rely on integration into genes that are transcriptionally active in ES cells, the use of an entrapment approach for the identification and investigation of such candidate genes is feasible. The following study reports the use of a random gene trap cloning approach and *in vitro* pre-screen to identify novel genes that are expressed in ES cells and downregulated upon differentiation *in vitro*.

Chapter Two

MATERIALS AND METHODS

2.1 MATERIALS

Unless stated otherwise, general laboratory reagents were of analytical grade and supplied by BDH Laboratory Supplies (Poole, UK) or Sigma Chemical Co., (Sigma), (St. Louis, MO). Phenol saturated in Tris-HCl buffer was supplied by Life Technologies, Inc., (Life Technologies), (Grand Island, NY). Deoxynucleotide triphosphates (dNTPs), supplied by Promega (Madison, WI), were mixed at 2.5 mM each in deionized distilled water (ddH$_2$0) and stored in small aliquots at -20°C. Bacterial media constituents were supplied by DIFCO laboratories (Detroit, MI) and electrophoresis grade agarose was supplied by Progen Industries (Darra, Australia). Synthetic oligonucleotide primers were supplied by Oswel (Boldrewood, UK) and also by Pacific Oligos (Lismore, Australia). Radioisotopes were supplied by Amersham (Amersham, UK) and X-ray film by Eastman Kodak Co., (Kodak), (Rochester, NY).

Recipes for commonly used solutions are either provided in Appendix I, referenced, or described with their methodology in this chapter. As required, solutions were sterilized either by autoclaving or by filtration through a 0.22 μm filter unit (Millipore, Bedford, MA). Supplier locations are given with the first reference made to each company.

2.1.1 DNA Vectors

<u>Commercially available:</u>

pCR-Blunt (Invitrogen, Groningen, The Netherlands)

pGEM-T (Promega)

<u>Constructed by Dr. Peter Mountford (Stem Cell Sciences, Melbourne, Australia):</u>

pGT1.8 (OPT) IRES*βgeo* gene trap (see Chapter three)

pGT1.8 (OPT/Nuclear) IRES*βgeo* gene trap (see Chapter three)

2.1.2 DNA probes for hybridization

Gapdh   0.8 kb *Hind* III/*Eco*R I fragment of *Gapdh* coding sequence, [plasmid DNA kindly provided by Dr. Kate Loveland, Monash Institute of Reproduction and Development, Melbourne, Australia, (MIRD)].

βgeo   Approximately 2.5 kb *Sph* I/*Cla* I fragment of *lacZ/neo* coding sequence isolated from pGT1.8 (OPT/Nuclear) IRES*βgeo* gene trap vector.

En2   Approximately 1.2 kb *Eco*R I /*Bam*H I fragment of *En-2* intron sequence isolated from pGT1.8 (OPT/Nuclear) IRES*βgeo* gene trap vector.

Neo   0.8 kb *Xba* I/*Pst* I fragment of *neo* coding sequence isolated from pGT1.8 (OPT/Nuclear) IRES*βgeo* gene trap vector.

Oct4   1 kb *Hind* III/*Acc* I fragment of *Oct-3/4* coding sequence isolated from pKFKF/PGK-Oct-3, (plasmid DNA kindly provided by Dr. Peter Mountford, Stem Cell Sciences, Melbourne, Australia).

En41   41 nt oligonucleotide primer for the antisense strand of *En-2* coding sequence starting 41 nts downstream from the splice acceptor site in the pGT1.8 (OPT/Nuclear) IRES*βgeo* gene trap vector.

        5'-GCG ATC TGC GTT CTT CTT CTT TGG TTT TCG GGA CCT GGG AC-3'

Cob54Race   730 bp *Pst I/Sph* I fragment for endogenous sequence (see Figure 5.3) isolated from a cloned 960 bp 5'-RACE product (gt54#5/*pCR-BLUNT*) generated from the COB54 gene trap fusion transcript.

Cob54Pcr   817 bp RT-PCR product generated from wild type ES cell mRNA for endogenous COB54 sequence, extending from nt 213 to nt 1129 in the predicted 1.7 kb sequence (see Figure 5.3).

2.1.3 RACE and RT-PCR oligonucleotide primers

pGT 1.8 (OPT/Nuclear) IRES*βgeo* 5'-RACE & direct sequencing primers:

Oligo numbers refer to Figure 4.2. Laboratory stock names/numbers are in parentheses.

Oligo 1: (IRESRT2)   5'-CAA GCG GCT TCG GCC AGT AAC-3'

Oligo 3: (Jenrace 1)   5'-GCT CTG TCA GGT ACC TGT TGG-3'

Oligo 4: (Jenrace 2)   5'-ACT CAG CCT TGA GCC TCT GGA-3'

Oligo 2-dT: (#56)   5'-GGT TGT GAG CTC TTC TAG ATG G(T)$_{17}$-3'

Oligo 2: (#59) anchor   5'-GGT TGT GAG CTC TTC TAG ATG G-3'

Oligo 2-b*: (#105)   5'-*Biotin*GGT TGT GAG CTC TTC TAG ATG G-3'

Oligo 5: (#98) sequencing   5'-AGC AGT GAA GGC TGT GC-3'

COB54 RACE and PCR primers:

For primer positions in COB54 cDNA sequence refer to Figure 5.3

COB54F1   5'-GAA ACC AGC ACA GCG AGA AG-3'

COB54F2   5'-ACA TTG CCT CCC GAC ACT AC-3'

COB54F3   5'-GGC TTT CAG CTG CAT ATC G-3'

COB54RT   5'-GTA GTG TCG GGA GGC AAT GT-3'

COB54R1   5'-GAG CCC ATT ATG CTT CTC TCG-3'

COB54R2   5'-CCA GGA GGC CAG GAT TTA CA-3'

COB54R3   5'-CAC ATA TCC ATT GTT AAT GCC-3'

βgeo PCR primers:

BGEO5'   5'-ACT ATC CCG ACC GCC TTA CT-3'

BGEO3'   5'-TAG CGG CTG ATG TTG AAC TG-3'

General use primers:

E. coli poly A polymerase PCR primers (Life Technologies)

Random RT primers (Life Technologies)

2.1.4 *Escherichia coli* strains & culture media

E. coli strains:

XL1-Blue (Stratagene, La Jolla, CA)

JM109 (Promega)

One Shot TOP10 (Invitrogen)

Bacterial culture media:

The following were prepared as described in Appendix A, Sambrook *et al.*, (1989).

Luria-Bertani (LB) medium

LB agar plates

Terrific broth

0.17 M KH_2PO_4, 0.72 M K_2HPO_4 sterile salts solution for Terrific broth

SOC medium

Antibiotics:

Ampicillin (Sigma)

Kanamycin (Sigma)

For each, stock solutions of 50 mg/ml were prepared in ddH_2O, filter sterilized and stored at -20°C. Both stocks were used at a working concentration of 50 μg/ml by adding 1:1000 (v/v) to sterilized media where required.

2.1.5 Embryonic stem cells, culture media & solutions

<u>ES cell lines:</u>

The ES cell lines E14 (Handyside *et al.*, 1989) and E14Tg2a (Hooper *et al.*, 1987), derived from the 129/Ola strain of mice as described, were kindly provided by Prof. Martin Hooper (University of Edinburgh, UK). These karyotypically male lines were grown free of feeder fibroblast cells in the presence of LIF for routine ES cell culture, and used for gene trap electroporation experiments.

<u>Genetically modified ES cell lines:</u>

Kindly provided by Dr. Peter Mountford – Stem Cell Sciences, Melbourne, Australia.

OKO160 (Mountford *et al.*, 1994)

ZIN40 (Mountford and Smith, 1995)

DIA RK022 (Li *et al.*, 1995)

DIA-*βgeo*pA 63 (Mountford *et al.*, 1994)

GTIRES 177 (P.S. Mountford, unpublished)

Kindly provided by Prof. Bill Skarnes – University of California, Berkeley, CA, USA.

Gene trap ES cell lines harboring *βgeo* insertions in genes encoding secretory and membrane-spanning proteins.

ST514, ST519, ST497, ST534 (Skarnes *et al.*, 1995).

Generated in this study -

54 gene trap ES cell lines: COB1-54

<u>Tissue culture grade water:</u>

RO grade water was collected from the MIRD *Milli-Q biocel* tissue culture system (Millipore) and filter sterilized into 500ml tissue culture grade Schott bottles (Schott Duran, Mainz, Germany).

<u>Phosphate-buffered saline (PBS):</u>

Prepared and sterilized as described in Appendix I.

Fetal calf serum (FCS)

Supplied by Commonwealth Serum Laboratories, Parkville, Australia. Batch tested for ability to sustain the viability, growth and differentiation of ES cells at a clonal density in the presence and absence of LIF. Stored at -20°C.

ES cell culture medium:

80 ml	5x Glasgow minimum essential medium (GMEM) stock solution - prepared by dissolving 25 g GMEM powder (Life Technologies) in 400 ml ddH$_2$0, filter sterilized and stored at 4°C.
14.7 ml	7.5% (w/v) sodium bicarbonate in ddH$_2$0, filter sterilized and stored at room temperature (RT°C) for 1 month.
4.0 ml	100 mM sodium pyruvate (Life Technologies), stored at -20°C.
4.0 ml	10 mM 1x MEM non-essential amino acids (Life Technologies), stored at 4°C.
4.0 ml	200 mM L-glutamine (Life Technologies), stored at -20°C.
45 ml	FCS
0.45 ml	0.1 M β-mercaptoethanol - 70 μl 2-mercaptoethanol (Sigma) made up to 10 ml in PBS, filter sterilized and stored for 1 month at 4°C.
300 ml	Sterile ddH$_2$0.

Stored at 4°C for up to 1 month, adding 2 mM L-glutamine supplements every 14 days.

Antibiotics:

A solution containing 10,000 U/ml Penicillin G and 10,000 μg/ml Streptomycin sulfate (Life Technologies) was stored at -20°C and added 1:200 (v/v) to cell culture media as required.

Geneticin (G418 50mg/ml, Life Technologies) was stored at 4°C and used at a working concentration of 200-400 μg/ml in culture media for the selection of neomycin resistant ES cell colonies.

Leukemia Inhibitory Factor (LIF): (Differentiation Inhibiting Activity, DIA), (Smith *et al.*, 1988, Williams *et al.*, 1988).

Mouse-derived recombinant LIF at 4×10^7 U/ml (kindly provided by Dr. Lindsay Williams, MIRD), was batch titrated against clonal ES cell cultures to confirm that 100 U/ml activity is the minimal dilution required for 100% inhibition of differentiation. Routinely, titrated LIF was added to ES cell culture media at 10^3 U/ml from 1000x stock aliquots stored in ES cell culture medium with penicillin/streptomycin for up to 3 months at 4°C, or indefinitely at -20°C.

Trypsin solution: (0.25% Trypsin/1 mM EDTA/1% chicken serum):

100 ml   2.5% trypsin sterile stock in PBS

100 ml   10 mM ethylenediaminetetra-acetic acid (EDTA) sterile stock in PBS

10 ml   Chicken serum (stored at -20°C, Life Technologies)

Made up to 1 litre in sterile PBS, swirled to mix and stored in 20 ml aliquots at -20°C. After thawing, working stocks were kept at 4°C for no longer than 2 weeks.

To prepare a 2.5% (w/v) trypsin stock solution 10 g trypsin (Life Technologies) was dissolved in 400 ml PBS, filter sterilized and stored in 100 ml aliquots at -20°C.

To prepare a 10 mM (0.4%) EDTA stock solution 1.6 g Na_2EDTA salt was dissolved in 300 ml PBS adjusting the pH to 7.6 with 10 M NaOH, made up to 400 ml, autoclaved and stored at RT°C.

1% gelatin:

5 g of gelatin (Sigma) was added to 500 ml ddH_2O, autoclaved immediately and stored in 15 ml aliquots at -20°C. For a 0.1% working solution, 15 ml of the 1% stock was made up to 150 ml in sterile ddH_2O and kept at RT°C.

2.1.6 Mouse strains

Animals were provided by Monash University central animal housing facility (Clayton, Australia) and also by the Walter & Eliza Hall Institute (WEHI), (Parkville, Australia) SP1-pathogen free facility. Mice were maintained in the animal housing facilities at Monash Medical Centre (Clayton, Australia) on a diurnal light/dark cycle of 12hrs each, with the middle of the dark cycle (0200 h) deemed to be the ovulation point.

C57Bl/6J (inbred, from Monash University and WEHI)

129/Ola (inbred, from Monash University)

Monash University "wild type" F1 = C57Bl/6J female x CBA male

WEHI "wild type" F1 = CBA female x C57Bl/6J male

ZIN40 = transgenic strain incorporating a random *lacZ*-IRES-*neo* gene trap insertion, providing ubiquitous nuclear localized expression of βgal activity in embryonic and adult mouse tissues (Mountford and Smith, 1995); kindly provided by Dr. Peter Mountford, Stem Cell Sciences, Melbourne, Australia.

2.2 MOLECULAR BIOLOGY METHODS

Except where otherwise stated, the methods described in this section are based on those described in Sambrook *et al.*, (1989). An Eppendorf *5417C* microcentrifuge (Eppendorf, Hamburg, Germany) was used routinely for centrifugation in 0.5 – 1.5 ml tubes. For larger volume procedures, a *Biofuge stratos* centrifuge (Heraeus, Hanau, Germany), (Rotor #3335 - 10.70 cm maximum/5.6 cm minimum radii) and a *Sorvall RC-5* centrifuge (DuPont, Newtown, CT), (GSA rotor – 14.57 cm, SS-34 rotor – 10.70 cm maximum radii) were used. A *Lambda Bio 20 UV/Vis* spectrometer (Perkin-Elmer, Applied Biosystems Division, Foster City, CA) was used routinely for quantitation of RNA and DNA preparations.

2.2.1 General cloning techniques

Typically the following methods were used for the construction and subcloning of the gene trap vectors and for subcloning polymerase chain reaction (PCR) and RACE products into the plasmid vectors *pCR-BLUNT* (Invitrogen) and *pGEM-T* (Promega), for further analyses.

(a) Phenol/chloroform extraction of nucleic acids

For the removal of proteins phenol saturated in Tris-HCl buffer and chloroform-isoamyl alcohol (24:1 v/v) were added in a ratio of 1:1 to an equal volume of nucleic acid solution. Phases were vortexed vigorously for 10 seconds (s), and separated by centrifugation at 13,000 rpm for 5 minutes (min) at RT°C in a microcentrifuge, or at ~12,000 x g for 10 min in larger rotors. The upper aqueous phase containing the nucleic acid was transferred to a fresh tube and the extraction repeated with an equal volume of

chloroform to remove residual phenol. The final aqueous layer was transferred to a fresh tube for recovery of nucleic acid by precipitation (section 2.2.1b).

(b) Precipitation of nucleic acids

Unless otherwise stated, extracted DNA and RNA were precipitated by the addition of a 0.1 volume of 3 M sodium acetate (pH 5.2), (Appendix I), followed by 2.5 volumes of ice-cold 100% ethanol, mixed by gentle inversion and placed at -20°C for 30 min (or stored indefinitely). The precipitate was pelleted by centrifugation at 13,000 rpm for 15 min at 4°C in a microcentrifuge, or at ~12,000 x g for 30 min in larger rotors. To remove trace amounts of salts, the nucleic acid pellet was washed in a small volume of 70% (v/v) ethanol, recovered by 10 min centrifugation (as above), air dried at RT°C and resuspended as required in either buffer containing 10 mM Tris-HCl and 1 mM EDTA (pH 8.0), (TE buffer, Appendix I), or ddH$_2$0.

(c) Restriction digestion of DNA

Restriction endonucleases and reaction buffers were obtained from Promega, Boehringer-Mannheim (Mannheim, Germany) and New England Biolabs (Beverly, MA). Conditions recommended by the manufacturers for digestion with each enzyme were generally followed. For genomic DNA digests additional enzyme (> 2 units/μg DNA) and 1 mM spermidine [100 mM stock solution stored at -20°C, (ICN pharmaceuticals, Costa Mesa, CA)] were added to the reaction mix to improve the efficiency of digestion.

(d) Separation of DNA fragments on agarose gels

Restriction fragments and PCR products were fractionated by electrophoresis on 0.7-2% (w/v) agarose gels cast and run in either 1 x TAE or TBE buffer (Appendix I). Ethidium bromide [10 mg/ml stock, stored in dark bottle at RT°C, (Life Technologies)] was added to molten agarose at 0.5 μg/ml for visualization of DNA under ultraviolet (UV) illumination. DNA samples were mixed with a 0.1 volume of 10x DNA electrophoresis loading buffer (Appendix I) prior to loading gel lanes, and electrophoresis was carried out at voltages appropriate to agarose gel size, concentration and fragment separation required. DNA size markers used to determine fragment sizes were *1 kb Plus* DNA ladder (Life Technologies), bacteriophage λ DNA (Promega) digested with *Hind* III and

EcoR 1, and *pGEM* DNA markers (Promega). DNA was illuminated and photographed using a *Gel Doc 2000* gel documentation system (Bio-Rad, Hercules, CA).

(e)   Purification of DNA fragments from agarose gels

DNA fragments of interest were excised from agarose gels with sterile scalpels using a *FOTO/UV 26* transilluminator (Fotodyne, Hartland, WI), and DNA recovered from non-nucleic acid impurities with either the *QIAquick* gel extraction kit (Qiagen, Hilden, Germany) or the *Bresa-Clean* DNA purification kit (GeneWorks, Adelaide, Australia), according to the manufacturers' instructions. Purified DNA was resuspended in TE buffer or ddH$_2$0 for direct use in subsequent enzymatic reactions or cloning procedures.

(f)   Generation of blunt DNA ends

For the ligation of DNA fragments with non-identical termini *pCR-Blunt* (Invitrogen) plasmid vector was used. This vector is supplied linearized and blunt-ended at a unique *EcoR* I site in the polylinker, and purified for direct use in ligation reactions. *pCR-Blunt* provides selection of recombinants only via disruption of a lethal gene, such that background due to recircularization is low.

PCR products were end-polished with *Pfu* DNA polymerase to remove the 3' deoxyadenosine (A)-overhangs resulting from the exonuclease activity of *Thermus aquaticus (Taq)* DNA polymerase in PCR reactions. Prior to the end-polishing reaction, PCR and 3'-RACE products yielding a discrete band size were purified by gel purification (section 2.2.1e), while 5'-RACE products yielding various product sizes were purified by phenol/chloroform extraction and ethanol precipitation (sections 2.2.1a & b). The following end-polishing method was adapted from that published by Costa and Weiner, (1995).

A 10 µl *Pfu* polishing reaction mix was prepared with dNTPs at 2.5 mM each, 1x buffer [20 mM Tris-HCl (pH 8.8), 10 mM KCl, 10 mM (NH$_4$)$_2$SO$_4$, 2 mM MgSO$_4$, 0.1% Triton X-100, 0.1 mg/ml nuclease-free bovine serum albumin (BSA)], 2.5 units of cloned *Pfu turbo* DNA polymerase (10x buffer supplied, Stratagene) and typically <1 µg of purified DNA product. This was overlaid with mineral oil (Sigma) and incubated at 72°C for 30 min, then cooled on ice. End-polished DNA fragments were added

directly to a ligation reaction since *Pfu* DNA polymerase is inactive at temperatures below 50°C.

(g) Ligation of DNA fragments

Ligation reactions were usually set up with a vector to PCR insert ratio of 1:3 by molarity, although to achieve positive recombinants with 5'- and 3'-RACE products ratios in the order of 1:10-12 were required. A control ligation was always performed containing vector DNA alone with ligase to assess the extent of self-ligation.

For gene trap clones of interest (Chapter five) blunt ligations for 5'-RACE products were performed in a 10 µl reaction containing 25 ng linearized *pCR-BLUNT* vector (Invitrogen), 1x buffer with ATP [6 mM Tris-HCl (pH 7.5), 6 mM MgCl$_2$, 5 mM NaCl, 0.1 mg/ml BSA, 7 mM β-mercaptoethanol, 0.1 mM ATP, 2 mM dithiothreitol (DTT), 1 mM spermidine] and 4 Weiss units of T4 DNA ligase, (10x buffer supplied, Invitrogen). Blunt ligation mixes were incubated at 16°C overnight (o/n) and either kept on ice until transformation, or stored at -20°C.

For the COB54 gene trap clone (Chapter five) cohesive ligations for PCR products as well as 5'- and 3'-RACE products were set up with *pGEM-T* vector (Promega), which is supplied linearized at a unique *Eco*R V site in a polylinker, with 3'-terminal thymidine nucleotides (T) added to both ends and purified for direct use in ligation reactions. These single 3'-T overhangs at the insertion site prevent recircularization of the vector, and provide a compatible overhang for PCR products generated by *Taq* DNA polymerase. Ligations were performed in a 10 µl reaction containing 25 ng linearized *pGEM-T* vector (Promega), 1x buffer with ATP [30 mM Tris-HCl (pH 7.8), 10 mM MgCl$_2$, 1 mM ATP, 10 mM DTT], 3 Weiss units of T4 DNA ligase (10x buffer supplied, Promega) and additional ATP to a final concentration of 1.5 mM. Cohesive ligation mixes were incubated at 4-14°C o/n and either kept on ice until transformation, or stored at -20°C.

(h) Preparation of competent *E. coli* cells

JM109 (Promega) and *One Shot* TOP10 (Invitrogen) strains were purchased as competent cells ready for use. For XL1-Blue strain (Stratagene), an in-house stock of

competent cells was prepared by the following method, adapted from that published by Chung *et al.*, (1989).

1.   Bacteria were streaked from glycerol stocks onto an LB plate and grown o/n at 37°C. A single colony was inoculated into 5 ml LB medium, and incubated at 37°C o/n with vigorous shaking.

2.   A flask containing 250 ml LB medium (pre-warmed to 37°C for 30 min) was inoculated with the 5 ml o/n culture and incubated at 37°C with vigorous shaking until cells were at $OD_{600} = 0.3$-0.4.

3.   Cells were immediately chilled on ice for 5 min, transferred to a pre-chilled 250 ml bottle (Nalgene, Rochester, NY) for pelleting at 4000 x g for 10 min, at 4°C in a *Sorvall* GSA rotor and the supernatant discarded. All rotors, centrifuge bottles, tubes and pipettes/tips were kept at 4°C from this step onwards.

4.   The bacterial pellet was gently resuspended at $1/10^{th}$ of the original volume in ice cold transformation and storage solution (TSS), (10% w/v PEG, 5% v/v DMSO, 50 mM $MgCl_2$ made up to final volume with LB medium, pH 6.5), using gentle swirling action.

5.   200 µl cell aliquots were snap frozen on dry ice/ethanol in pre-chilled sterile cryovials (Nunc, Roskilde, Denmark) before storing at -80°C.

As TSS is a low salt storage medium, the addition of a salts solution (section 2.2.1i) was required when transforming these competent cells.

(i)   Transformation of competent *E. coli* cells

Gene trap vectors (Chapter three) were transformed into both XL1-Blue and JM109 competent cells. The 5'-RACE products for gene trap clones of interest (Chapter five) were transformed into *One shot* Top 10 cells, while the 5'-RACE, 3'-RACE and PCR products for the COB54 gene trap clone (Chapter five) were transformed into JM109 cells.

Typically for each transformation, a 50 µl aliquot (100 µl/XL1-Blue) of competent cells was thawed on ice for 5-10 min, gently mixed with 10 ng of plasmid DNA or 2 µl of ligation mixes in pre-chilled Eppendorf 1.5 ml tubes and incubated on ice for 30 min,

(20 min/JM109). For XL1-Blue transformations, the DNA mix was made up to a volume of 80 µl ddH$_2$0 and added to 20 µl of a 5x KCM salts buffer (0.5 M KCl, 0.15 M CaCl$_2$, 0.25 M MgCl$_2$), prior to mixing with cells. For *One Shot* Top 10 cells, 2 µl of 0.5 M β-mercaptoethanol was stirred gently into the thawed cells with a pipette tip before mixing with DNA. Cells were heat shocked at 42°C for 45 s and cooled on ice for 2 min before adding SOC medium - (800 µl/XL1-Blue, 950 µl/JM109, 250 µl/*One Shot* Top 10), pre-warmed to 37°C. Cultures were transferred to Falcon 15 ml (17 x 100 mm) tubes (Becton Dickinson, Lincoln Park, NJ), and incubated at 37°C with shaking (~200 rpm) for 1 h, (1.5 h/JM109). Cells were spread on LB plates containing 50 µg/ml of ampicillin (50 µg/ml kanamycin/*One Shot* Top 10 cells) typically from 20 µl, 50 µl and 100 µl of the SOC cultures, and incubated o/n at 37°C for growth of transformants. For RACE products, where higher numbers of colonies were often desired, the culture was pelleted by centrifugation at 1000 x g for 10 min and resuspended in 200 µl of SOC medium before plating.

2.2.2 Isolation of nucleic acids

(a) Small-scale preparation of plasmid DNA

Small-scale DNA plasmid preparations were made from 5 ml LB cultures (with ampicillin or kanamycin added to 50 µg/ml), incubated o/n at 37°C with shaking. Cells were harvested by centrifugation at 5000 rpm for 5 min in a 1.5 ml Eppendorf tube and plasmid DNA recovered from cellular debris and chromosomal DNA by the standard alkaline lysis protocol (Sambrook *et al.*, 1989), followed by purification with phenol/chloroform extraction and ethanol precipitation (sections 2.2.1a & b). The DNA was resuspended in 50 µl TE buffer containing 10 µg/ml RNase A (Appendix 1), or ddH$_2$0 for further applications. The *Plasmid Mini* (Qiagen) and *High Pure* Plasmid Isolation (Boehringer-Mannheim) kits were also used, according to the manufacturers' instructions. These kits employ modified alkaline lysis methods to release plasmid DNA, and spin-columns with specific resins to bind the DNA for purification and wash steps before eluting into TE buffer.

(b) Large-scale preparation of plasmid DNA

Gene trap vectors required for electroporation into ES cells (Chapter three) were prepared by caesium chloride (CsCl) gradient purification of larger amounts of plasmid DNA, according to the following method.

1. A 15-20 ml o/n LB/ampicillin culture of the required colony was used to inoculate 200 ml pre-warmed sterile Terrific broth (10% sterile salts added following autoclaving – section 2.1.4) in a 1 litre flask, and incubated o/n at 37°C with shaking at 240 rpm.

2. Cells were pelleted at 5000 rpm for 20 min at 4°C (*Sorvall* GSA rotor) and plasmid DNA recovered by the standard scaled-up alkaline lysis method followed by isopropanol precipitation as described in Sambrook *et al.,* (1989). The nucleic acid pellet was resuspended in 10.5 ml TE buffer, ensuring proper dissolution before purification.

3. 10.5 g of CsCl was added to the DNA solution, briefly warmed to 30°C and mixed gently to dissolve the salt. This was followed by the addition of 800 µl of 10 mg/ml ethidium bromide (Life Technologies), placing the DNA mixture in a 65°C waterbath for 5 s then gently mixing by inversion.

4. Using a pasteur pipette a *Quick-Seal* tube (Beckman, Fullerton, CA) was completely filled with the mixture, heat-sealed avoiding air pockets and centrifuged at 70,000 rpm for 18 h at 20°C in a Beckman *Ti70* rotor ultracentrifuge.

5. The lower band corresponding to closed circular plasmid DNA was recovered by puncturing the side of the tube with a 21-gauge needle (see Sambrook *et al.,* 1989) and ethidium bromide removed by 4-6 extractions with an equal volume of isobutanol saturated with CsCl-buffered ddH$_2$0, until no traces of pink colour were visible in the aqueous phase.

6. To remove CsCl from the DNA solution, 3 volumes of ddH$_2$0 were added, followed by precipitation in 4 volumes of 100% ethanol. Incubation was performed at 4°C (-20°C will precipitate CsCl salts) for 1-2 h, after which the nucleic acid was pelleted at 10,000 rpm for 15 min at 4°C (*Sorvall* SS34 rotor), air dried at RT°C and resuspended in TE buffer at 1 µg/µl.

The *Plasmid Maxi* kit with Qiagen-tip 500 (Qiagen) was also used to prepare subsequent large-scale plasmid DNA preparations, following the manufacturer's instructions. This kit employs a modified alkaline lysis procedure, followed by binding and purification of plasmid DNA on an anion-exchange resin. The eluted DNA is concentrated and desalted by isopropanol precipitation and redissolved in TE buffer as above. One Qiagen-tip 500 was used per 100 ml of Terrific broth culture.

(c) Isolation of genomic DNA from ES cells

The following methods for preparing genomic DNA (sections 2.2.2c & d) are adapted from the method published by Laird *et al.*, (1991).

1.   Media was aspirated from ES cell cultures grown to confluency (section 2.3.1) in Falcon 24- or 12-well tissue culture plates (Becton Dickinson), and cells rinsed twice with PBS. If not proceeding directly to the lysis step below, cells were trypsinized (section 2.3.1), washed and pelleted twice in PBS by centrifugation in 1.5 ml tubes at 1000 rpm for 5 min, and cell pellets stored at -80°C.

2.   500 µl of lysis buffer [100 mM Tris-HCl (pH 8.5), 5 mM EDTA (pH 8.0), 0.2% (v/v) SDS, 200 mM NaCl, with proteinase K added at 100 µg/ml just before use – Appendix I] was added to each well and incubated o/n at 37°C.

3.   Lysates were transferred to 1.5 ml tubes and extracted in 500 µl isopropanol with gentle shaking on a rocking platform (Ratek Instruments, Boronia, Australia) for 5-10 min at RT°C.

4.   Following centrifugation at 13,000 rpm for 10 min at RT°C, the supernatant was carefully removed. The DNA pellet was washed with 500 µl of 70% (v/v) ethanol and air-dried at RT°C.

5.   DNA was dissolved in 100 µl TE buffer (200 µl for 12-well cultures) for 30-60 min at 60°C, and stored at -20°C. Typically, yields of 10-30 µg were achieved from 24-well cultures.

(d) Isolation of genomic DNA from mouse tissues & embryos

1.   Tail biopsies (~1 cm in length) were taken from 3 week old weaned mice, and 10.5 dpc mid-gestation stage embryos were dissected (section 2.4.1) for genotype

analyses. These were collected into 1.5 ml Eppendorf tubes on dry ice - (embryos were rinsed in PBS) - and either stored at -80°C until required, or incubated o/n at 55°C in 500 µl of lysis buffer (section 2.2.2c) on a rocking platform (Ratek Instruments).

2.   Lysates were centrifuged at 13,000 rpm for 10 min and supernatants removed from tissue debris, transferring to new 1.5 ml tubes on ice.

3.   1 ml of isopropanol was added, the tube vortexed vigorously for 10 s, and centrifuged at 13,000 rpm for 20 min at RT°C. After carefully removing supernatant, the DNA pellet was washed with 500 µl of 70% (v/v) ethanol and air-dried at RT°C.

4.   Depending on the size of the precipitate, 100-200 µl TE buffer was added and tubes incubated for 1-2 h at 60°C with periodic vortexing, to help dissolve the DNA. Typically, yields of ~30-50 µg from tail biopsies and >100 µg from mid-gestation stage embryos were achieved.

To isolate a crude genomic DNA preparation for direct use in rapid PCR screens for transgenic mice, a quick method for ear punches was adapted from the above, omitting extraction and precipitation steps.

*   Ear punches were collected from weaning age mice into 1.5 ml tubes on dry ice and incubated with 100 µl of a low salts lysis buffer [10 mM Tris-HCl (pH 8.3), 50 mM NaCl, 0.2% (v/v) Tween 20 (Sigma), with proteinase K added to 100 µg/ml just before use – Appendix I] at 55°C for 1 h.

*   Lysates were incubated at 95°C for 5 min to inactivate the enzyme, centrifuged at 13,000 rpm for 5 min and 5 µl of supernatants used directly as PCR reaction templates (section 2.2.7a), or stored at –20°C and thawed briefly at 37°C before use.

(e)   Preparation of total RNA from ES cells and tissues

Standard precautions were followed to avoid RNase contamination during preparation and manipulation of RNA. Sterile plasticware was used wherever possible and glassware baked at 180°C. Sterile diethyl-pyrocarbonate (DEPC)-treated ddH_2O and solutions (Appendix I) were used throughout all RNA procedures.

For the preparation of total RNA from ES cells and tissues the following protocol was used, adapted from that published by Chomczynski and Sacchi (1987), with extra extraction steps added.

1.   ES cell cultures grown to ~85% confluency (section 2.3.1) were rinsed twice in cold sterile PBS and lysed by scraping in 1 volume (1.8 ml per 78.5 cm^2/6.8 ml per 175 cm^2 culture flasks) of denaturation solution [4 M guanidinium thiocyanate, 25 mM sodium citrate (pH 7.0), 0.5% sarcosyl, 0.1 M 2-mercaptoethanol]. Tissues were kept on ice at collection (section 2.4.4), rinsed in cold PBS and homogenized in 1 volume (1 ml for each 100 mg of tissue) of the above denaturation solution. Lysates were transferred to Falcon 15 ml polypropylene tubes (Becton Dickinson) and kept on ice.

2.   A 0.1 volume of 2 M sodium acetate (pH 4) was added and mixed by inversion.

3.   1 volume of water-saturated phenol was added and mixed with vigorous vortexing.

4.   A 0.2 volume of 49:1 chloroform/isoamyl alcohol was added, mixed with vortexing and cooled on ice for 20 min.

5.   The suspension was centrifuged at 10,000 x g for 20 min at 4°C (*Sorvall* SS34 rotor), and the upper aqueous phase transferred to a fresh polypropylene tube.

6.   To each aqueous yield of a 4 ml volume or less, 2 ml water-saturated phenol and 2 ml 49:1 chloroform/isoamyl alcohol were added and mixed by inversion.

7.   The suspension was centrifuged at 10,000 x g for 10 min at 4°C and the aqueous layer again transferred to a fresh polypropylene tube.

8.   The extraction described in steps 6 and 7 was repeated.

9.   1 volume of isopropanol was added and the RNA left to precipitate at -20°C for

1 h, (or o/n if desired).

10.   RNA was pelleted at 10,000 x g for 20 min at 4°C, briefly air-dried at RT°C and resuspended in 600 µl of denaturing solution (see step 1), transferring to a microcentrifuge tube.

11.   1 volume of isopropanol was added and RNA re-precipitated at -20°C for 30 min, (or o/n if desired).

12.   The precipitate was pelleted at 12,000 x g for 10 min at 4°C, washed with 1 ml 70% ethanol, vortexed and incubated at RT°C for 10 min to dissolve residual traces of guanidinium.

13.   RNA was collected by centrifugation at 12,000 x g for 10 min at 4°C, briefly air-dried at RT°C and dissolved in 50-100 μl DEPC-treated ddH$_2$0 (Appendix I).

14.   RNA solutions were stored at -80°C adding 3 volumes of ethanol for long-term storage. To recover the RNA, an aliquot was removed, sodium acetate (pH 5.2) added at 0.3 M, mixed well and centrifuged at 13,000 rpm for 15 min at 4°C.

(f)   Preparation of poly A⁺ mRNA

A method for isolating poly A⁺ mRNA on oligo(dT) cellulose - based on that published by Aviv and Leder, (1972) - was used in conjunction with a protocol for preparing ES cell homogenates directly for affinity chromatography (M. Pera, MIRD; *pers. comm.*). For tissue samples, poly A⁺ mRNA was also isolated on oligo(dT) cellulose columns following the initial isolation of total RNA as described in section 2.2.2e.

1.   4-10 x 10⁷ ES cells grown to ~85% confluency (section 2.3.1) were rinsed twice in cold sterile PBS and lysed by scraping in 5 ml cold DEPC-treated STE solution [0.1 M NaCl, 20 mM Tris-HCl (pH 7.5), 10 mM EDTA (pH 8.0) – Appendix I].

2.   The cell lysate was transferred to a Falcon 50 ml polypropylene tube (Becton Dickinson), quickly adding proteinase K at 100 μg/ml and SDS at 0.5% (Appendix I) prior to homogenizing at low speed for 45 s.

3.   The cell homogenate was incubated at 37°C for 40 min and prepared for binding with oligo(dT) cellulose beads by adding 5 M NaCl to a final concentration of 0.5 M (~440 μl), or stored at -80°C. Similarly, prepared total RNA for tissue samples (section 2.2.2e) was brought to the correct molarity for binding.

4.   For each 0.5 mg of total RNA, 100 mg oligo(dT) cellulose (New England Biolabs) was weighed asceptically into Falcon 15 ml polypropylene tubes (Becton Dickinson). Beads were prepared by washing 3 times in 4 ml of DEPC-treated loading buffer [0.4 M NaCl, 20 mM Tris-HCl (pH 7.5), 10 mM EDTA (pH 8.0), 0.2% SDS – Appendix I], centrifuging at 3000 rpm (*Biofuge stratos*, 3335 rotor) for < 1 min each time.

5.   Cell homogenates/tissue total RNA solutions were added to the prepared beads and mixed gently on a rotor wheel for 2 h at RT°C.

6.   The bead/RNA mix was spun at 3500 rpm for 1 min and supernatant discarded, leaving the oligo(dT) cellulose beads with bound poly A⁺ mRNA. The beads were washed in 2 ml of loading buffer and pelleted as for step 4, then gently resuspended in a fresh 2 ml aliquot of the loading buffer.

7.   The bead/buffer slurry was loaded onto a 0.8 x 4 cm chromatography column (Bio-Rad) by inversion of the tube and 1-2 ml of loading buffer used to rinse all beads onto the column. A further 1 ml of loading buffer was run through the column, discarding the run-off.

8.   The column was washed with 3x 1 ml of DEPC-treated oligo(dT) washing buffer [0.1 M NaCl, 10 mM Tris-HCl (pH 7.5), 1 mM EDTA (pH 8.0), 0.2% SDS – Appendix I], discarding the run-off.

9.   Poly A⁺ mRNA was eluted from the beads into a sterile 15 ml tube with 3x 1 ml of DEPC-treated elution buffer [1 mM NaCl, 1 mM EDTA (pH 8.0), 0.2% SDS – Appendix I], pre-warmed to 37°C.

10.   mRNA was recovered by precipitation (section 2.2.1b) using DEPC-treated NaOAc (Appendix I), an o/n -20°C incubation of the ethanol mix and centrifugation at 7000 rpm (*Biofuge stratos*, 3335 rotor) for 1 h at 4°C. The pellet was transferred to a microcentrifuge tube in 1 ml 70% ethanol, washed, respun at 13,000 rpm for 20 min and resuspended in 20 µl DEPC-treated ddH₂0 (Appendix I) after briefly air-drying at RT°C. Storage conditions were as described for total RNA solutions (section 2.2.2e).

2.2.3 DNA blot analysis

(a)   Bacterial colony blots

To identify individual bacterial colonies carrying recombinant plasmids for PCR and RACE products of interest, colonies were lysed at RT°C on nylon membranes and DNA prepared for hybridization with labelled probes. The method used is a modification of that originally published by Grunstein and Hogness, (1975).

1.   An LB plate containing colonies from an o/n culture was chilled for 1 h at 4°C.

2.　A Hybond-N+ positively charged nylon membrane (Amersham) was placed, numbered side down, on the surface of the agar medium in contact with the bacterial colonies, for up to 1 min. An 18-gauge needle was used to stab through the membrane edges and underlying agar to provide points of realignment for selection of positive recombinant colonies.

3.　Using blunt-ended forceps, the nylon membrane was peeled off the plate and placed colony side up on a 3 piece pad of 3MM paper (Whatman, Maidstone, UK) saturated with 10% SDS (Appendix I) for 3 min. The agar plate was incubated at 37°C for ~6 h or left at RT°C o/n to regenerate colonies, then stored inverted in Saran wrap at 4°C.

4.　The membrane was transferred to a second pad of Whatman 3MM paper saturated with denaturation solution (1.5 M NaCl, 0.5 M NaOH) for 5 min.

5.　The membrane was transferred to a third pad of Whatman 3MM paper saturated with neutralization solution [1.5 M NaCl, 0.5 M Tris-HCl (pH 7.5), 1 mM EDTA (pH 8.0)] for 5 min.

6.　The membrane was washed for 5 min on a fourth pad of Whatman 3MM paper saturated in 2x SSC (Appendix I), and blotted dry on 3MM paper for 30 min at RT°C.

7.　DNA was fixed to the charged membrane by placing, DNA face up, on a pad of Whatman 3MM paper saturated in freshly prepared 0.4 M NaOH for 30 min. The membrane was rinsed briefly in 2x SSC and stored between sheets of 3MM paper in Saran wrap, at 4°C.

8.　Prior to hybridization steps (section 2.2.6a) colony blots were floated on 2x SSC for 5 min followed by incubation in a pre-wash solution (5x SSC, 0.5% SDS, 1 mM EDTA) at 50°C for 30 min, to reduce background hybridization.

(b)　Southern blot

For analyses of gene trap ES cell lines and for genotyping mice, Southern blots (Southern, 1975) were prepared for hybridization of genomic DNA with labelled probes as follows:

1. Typically, 15-20 μg of genomic DNA was digested with the desired restriction enzyme (section 2.2.1c) in a reaction volume of 150 μl, concentrated by ethanol precipitation (section2.2.1b) and fractionated by electrophoresis on a 0.8% 1x TAE agarose gel (section 2.2.1d) run o/n at ~35 volts. The gel was photographed under UV illumination (section 2.2.1d) prior to blotting.

2. Where DNA fragments of interest were >10 kb in size, the gel was depurinated with gentle shaking for 20 min in 0.25 M HCl and rinsed in ddH_2O.

3. DNA was transferred to a Hybond N+ positively charged nylon membrane (Amersham) by capillary blotting, as described by Sambrook *et al.,* (1989). The transfer was run o/n at RT°C and under alkaline conditions, using a freshly prepared transfer buffer of 0.4 M NaOH. These conditions also provided for irreversible fixation of transferred DNA to the membrane.

4. Following transfer, the membrane was rinsed briefly in 2x SSC and stored for hybridization (section 2.2.6a) as for bacterial colony blots (section 2.2.3a).

(c) Alkaline dry blot

For the transfer of abundant low molecular weight RACE products (section 2.2.8) following gel electrophoresis (section 2.2.1d), a simplified Southern dry blot protocol was used as follows:

1. Following electrophoresis, the gel was gently shaken in a denaturation solution (1.5 M NaCl, 0.5 M NaOH) for 45-60 min at RT°C.

2. The gel was gently shaken in a less concentrated alkaline blot solution (1.5 M NaCl, 0.24 M NaOH) for 10 min at RT°C.

3. A simplified transfer sandwich was set up by inverting the gel onto a glass plate, placing a Hybond-N+ (Amersham) membrane on top avoiding trapped air bubbles, followed by 2 sheets of Whatman 3MM paper and a flat glass or plastic plate. A heavy book was placed on top to assist the dry "squash" transfer of DNA to the membrane.

4. Following transfer (3-4 h for small PCR products), the membrane was rinsed briefly in 2x SSC and stored for hybridization (section 2.2.6a) as for bacterial colony blots (section 2.2.3a).

2.2.4 Northern blot analysis of RNA

Northern blots were prepared for hybridization of ES cell and mouse tissue RNA samples with labelled probes, based on the method published by Fourney et al., (1988).

(a) Separation of RNA on agarose/formaldehyde gels

Extracted RNA samples (sections 2.2.2e & f) were electrophoresed on 1% (w/v) agarose/0.66 M formaldehyde denaturing gels cast and run in 1 x MOPS buffer (Appendix I). Total RNA (10-30 µg) or poly A$^+$ mRNA (1-3 µg) resuspended in a 5 µl volume of DEPC-treated ddH$_2$0 (Appendix I) was mixed with a 25 µl volume of 6x RNA electrophoresis loading buffer (Appendix I). Samples were denatured at 65°C for 10 min, followed by rapid cooling on ice and the addition of 1 µl ethidium bromide (10 mg/ml stock, Life Technologies) prior to loading, for UV visualization. Gels were pre-run at 60 volts for 5-10 min and loaded samples electrophoresed at 30-35 volts for ~18 h at RT°C, with buffer circulating. To estimate the size of RNA transcripts 10 µg of RNA marker *G319* (Promega) was used. RNA was visualized and photographed by exposure to UV illumination as described for DNA (section 2.2.1d).

(b) RNA blotting

1. The RNA gel (section 2.2.4a) was prepared for transfer by gently shaking in two changes of 10x SSC (Appendix I), for 20 min each at RT°C, to remove residual formaldehyde.

2. RNA was transferred to a Hybond N+ positively charged nylon membrane (Amersham) by capillary blotting, as described by Sambrook et al., (1989). The transfer was run o/n at RT°C, using a wick saturated in 20x SSC (Appendix I) and a transfer buffer of 10x SSC.

3. Following transfer, the membrane was rinsed briefly in 2x SSC and transferred RNA fixed by baking for 2 h at 80°C. The membrane was stored for hybridization (section 2.2.6b) between sheets of Whatman 3MM paper in Saran wrap, at 4°C.

2.2.5 Preparation of radiolabelled probes

(a) Double stranded DNA fragments

The *High Prime* and *RTS Radprime* kits (Boehringer-Mannheim) and the *Rediprime* kit (Amersham) were used to label double stranded DNA (dsDNA) probe fragments, according to manufacturers' instructions. These kits employ the random primer labelling method (Feinberg and Vogelstein, 1983, 1984) to incorporate $[\alpha^{32}P]$-dCTP into a newly synthesized complementary DNA strand and provide ready-to-use reaction mixes in dry, RT°C stable formats.

Typically, 15-25 ng of linearized, gel-purified DNA fragment in ddH_2O (sections 2.2.1c, d, e) was denatured at 95°C for 10 min and cooled on ice for 5 min. The DNA was added to the provided mixes containing Klenow fragment, deoxynucleotides (dATP, dGTP and dTTP), random primers (octamers or nanomers), reaction buffer and a dye, according to manufacturers' instructions. Then 50 µCi of $[\alpha^{32}P]$-dCTP (10 µCi/µl, Amersham) was added and the 20-50 µl mix incubated at 37°C for 10-30 min. The reaction was stopped with a 5 µl volume of 0.2 M EDTA, or by heating at 65°C for 10 min.

Non-incorporated nucleotides were removed from labelled DNA by spin chromatography of the final mix through a G50 Sephadex *eppi-column* (Amersham Pharmacia Biotech, Uppsala, Sweden) at 2000 rpm for 3 min in a microcentrifuge, collecting the eluate into a fresh tube. The volume was made up to 100 µl in ddH_2O and a 2 µl aliquot taken to determine the total incorporated radioactivity on a *Tri-Carb 1900 TR* liquid scintillation analyser (Packard, Meriden, CT). Probes were stored for up to 1 week at -20°C.

(b) Oligonucleotides

The enzyme polynucleotide kinase (PNK) catalyses the transfer of the terminal phosphate group of ATP to the 5'-OH terminus of DNA or RNA. To end-label an oligonucleotide primer, ~10 pmol of the single stranded DNA was mixed in a 0.5 ml PCR grade tube with 1x PNK buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 100 µm EDTA, 5 mM DTT, 100 µm spermidine, pH 8.2), 10 units of PNK (10x buffer supplied, Boehringer-Mannheim) and 40 µCi of $[\gamma^{32}P]$ ATP or $[\gamma^{33}P]$ ATP (10 µCi/µl,

Amersham), in a reaction volume of 25 µl in ddH$_2$0. The reaction was incubated at 37°C for 30 min and stopped by heating at 70°C for 5 min, then cooled on ice. Probes were stored for up to 1 week at -20°C.

2.2.6 Hybridization conditions

Hybridizations were performed in a rotating glass cylinder with 10 ml buffer volumes, in either an Xtron *HI 2002* (Bartelt Instruments, Heidelberg West, Australia) or a Hybaid *Shake n' Stack* (Hybaid, Ashford, UK) hybridization oven.

(a)   DNA blots

1.   Pre-hybridization was performed in DNA hybridization buffer [0.5 M Na$_2$HPO$_4$ (pH 7.2), 7% SDS (w/v), 1% BSA (w/v), 1 mM EDTA (pH 8.0) – Appendix I] at 65°C for 4-5 h for Southern blots (section 2.2.3b), or for at least 30 min for bacterial colony blots (section 2.2.3a) and alkaline dry blots (section 2.2.3c).

2.   dsDNA probes (section 2.2.5a) were denatured at 95°C for 5-10 min and cooled on ice for 2 min. The denatured probe was added to a fresh aliquot of DNA hybridization buffer at 1-2 x 10^6 cpm/ml. Where oligonucleotide primer probes were used, no denaturation step was required, and a whole end-labelling reaction volume of 25 µl (section 2.2.5b) was added to the buffer. Hybridization proceeded o/n at 65°C.

3.   Post-hybridization washes were performed at 65°C with 2 incubations in DNA wash buffer I [2x SSC, 0.1% (w/v) SDS – Appendix I] for 30 min each, followed by 1-3 incubations in DNA wash buffer II [0.1x SSC, 0.1% SDS (w/v) – Appendix I] for 15 min each, depending on the stringency required.

4.   Membranes were kept moist in Saran wrap and exposed to X-ray film (*X-omat AR5* or *Biomax MR1*, Kodak) between intensifying screens at -80°C.

5.   To re-hybridize DNA membranes, the bound probe was stripped by placing the filter in boiling 0.1% SDS, leaving until the solution cooled (~20-30 min) and rinsing briefly in 2x SSC (Appendix I).

(b) RNA blots

1. For Northern blots (section 2.2.4b), pre-hybridization was performed in RNA hybridization buffer [5x Denhardt's, 50% deionized formamide (v/v), 1% SDS (w/v), 3% dextran sulphate (w/v), 5x SSPE, 100 μg/ml denatured salmon sperm DNA – Appendix I] at 42°C for 4-5 h.

2. dsDNA probes (section 2.2.5a) were denatured at 95°C for 5-10 min and cooled on ice for 2 min. The denatured probe was added to a fresh aliquot of RNA hybridization buffer at 2×10^6 cpm/ml. Hybridization proceeded o/n at 42°C.

3. Post-hybridization washes were performed at 65°C with 3 incubations in RNA wash buffer I [1x SSPE, 0.5% SDS (w/v) – Appendix I) for 15 min each, followed by a 15 min incubation at 60°C in RNA wash buffer II [0.1x SSPE, 0.5%SDS (w/v) – Appendix I].

4. Membranes were kept moist in Saran wrap and exposed to X-ray film (*X-omat AR5* or *Biomax MR1*, Kodak) between intensifying screens at -80°C.

5. To re-hybridize RNA membranes, the bound probe was stripped by placing the filter in boiling 0.5% SDS, leaving until the solution cooled (~20-30 min) and rinsing briefly in 2x SSC (Appendix I).

2.2.7 Polymerase Chain Reaction (PCR)

Sequencing, genotype and expression analyses for ES cell lines, mouse embryos and tissues employed PCR amplification of genomic DNA templates and reverse transcriptase (RT)-PCR amplification from RNA templates. The PCR protocols used were modifications of the method originally published by Saiki *et al.*, (1988), employing the use of *Taq* DNA polymerase.

(a) PCR amplification of DNA

Typically, a PCR reaction mix was set up in a 0.5 ml PCR tube on ice containing 10-50 ng DNA template, a 1x PCR buffer [10 mM Tris-HCl (pH 9.0 at 25°C), 50 mM KCl, 0.1% Triton X-100 – 10x buffer supplied, Promega], 2.0 mM $MgCl_2$ (25 mM stock solution, Promega), dNTPs at 0.2 mM each, 25 pmol (~100 ng) each of forward and reverse oligonucleotide primers, 1 unit of *Taq* DNA polymerase (5U/μl, Promega) and

made up to a final volume of 50 µl in ddH$_2$0. The mix was overlaid with a 20 µl drop of mineral oil (Sigma) and reactions run in either an *Omnigene* (Hybaid), or a *PTC-100* (MJ Research Inc., Waltham, MA) programmable thermal cycler. Samples were transferred from ice directly to 95°C for a denaturation step of 5 min, followed by 30 cycles of 95°C/30 s, 52–56°C/30 s (determined empirically for each primer set), 72°C/2 min, with a final extension of 72°C/15 min.

Control reactions were included for each experiment using known DNA templates and a reaction omitting DNA. The presence and size of PCR products was assessed by agarose gel electrophoresis (section 2.2.1d), (typically loading 10 µl for visualization) and gel bands purified as required (section 2.2.1e).

(b) Reverse transcriptase (RT)-PCR

Reverse transcription was performed on prepared RNA templates (sections 2.2.2e & f) using either random primers to generate a pool of cDNAs for amplification, or a gene-specific primer to produce a specific cDNA strand for amplification cycles. Control reactions without the RT enzyme and without sample template were included for each experiment.

1.   Prior to the RT reaction, 0.5-1 µg of poly A$^+$ mRNA (or 1-3 µg of total RNA) was annealed with 10 pmol of a gene-specific primer or 3 µg of random primers (3 µg/µl, Life Technologies) in the presence of 10 units of RNase Inhibitor (10 U/µl, Life Technologies), adding DEPC-treated ddH$_2$0 (Appendix I) to a total volume of 11.5 µl in a 0.5 ml PCR tube. For annealing with a gene-specific primer, the mix was incubated at 70°C for 10 min and then 60°C for 20 min, followed by rapid cooling on ice. For annealing with random primers, the mix was incubated at RT°C for 10 min.

As an optional step prior to proceeding with the above primer annealing, the RNA templates were firstly incubated in a 9 µl volume with 1 unit of RNase-free DNase (1 U/µl, Promega) at 37°C for 30 min, to remove any genomic DNA contamination. The reaction was stopped by adding 1 µl of 0.25 M EGTA and incubating at 70°C for 10 min.

2.   For each template a 7 µl RT reaction mix was prepared containing 4 µl of a 5x 1st strand buffer [250 mM Tris-HCl (pH 8.3 at RT°C), 375 mM KCl, 15 mM MgCl$_2$,

(Life Technologies)], 2 µl of 0.1 M DTT (Life Technologies) and 1.5 µl of 10 mM dNTPs. The annealing mix was added to the RT reaction mix, and tubes incubated for 2 min at 42°C for gene-specific primers or 37°C for random primers.

3.   200 units of *Superscript II* MMLV reverse transcriptase (200 U/µl, Life Technologies) were added (excepting control reactions) and the 20 µl reactions incubated at 42°C for 1 h, then 50°C for 30 min for gene-specific primers or 37°C for 1 h for random primers. The enzyme was inactivated by incubation at 70°C for 15 min and the tube cooled on ice.

4.   Typically, 5 µl of the above RT reaction was used as a cDNA template for PCR amplification with gene-specific primers, as described in section 2.2.7a, and the remainder stored at -20°C. To assess the yield of cDNA templates from an RT reaction using random primers, a PCR amplification was performed using *poly A polymerase* PCR primers (section 2.1.3) with a T_a of 52°C and products visualized by agarose gel electrophoresis (section 2.2.1d).

(c)   Direct RT-PCR on single mouse eggs and embryos

A modified protocol was used for performing RT-PCR directly on lysates of single mouse eggs, preimplantation stage embryos and postimplantation stage embryos up to 7.5 dpc, adapted from the method published by Daniels *et al.*, (1997). Control reactions without the RT enzyme and without sample template were included for each experiment.

1.   Unfertilized and fertilized mouse eggs and embryos up to 4.5 dpc were collected (section 2.4.1), rinsed briefly in cold PBS containing 1% FCS and transferred individually in a minimal volume to 0.5 ml PCR tubes on ice containing 5 µl of a lysis buffer [1 U RNase Inhibitor (Life Technologies), 0.8% Igepal (Sigma), 5 mM DTT, (Life Technologies)]. For 5.5-7.5 dpc embryos, lysis was carried out in 10µl of the same lysis buffer, and poly A$^+$ mRNA isolated with *Dynabeads Oligo(dT)$_{25}$* (Dynal, Oslo, Norway), according to the manufacturer's instructions. Eluted mRNA was resuspended in 5 µl of DEPC-treated ddH$_2$0 (Appendix I). Lysates and mRNA preparations could be stored at –80°C.

2. For each 5 µl lysate or mRNA preparation, a 5 µl RT reaction mix was prepared on ice containing 0.2 µl RNase Inhibitor (10 U/µl, Life Technologies), 0.3 µl of random primers (3 µg/µl, Life Technologies), 2 µl of 5x 1st strand buffer [250 mM Tris-HCl (pH 8.3 at RT°C), 375 mM KCl, 15 mM MgCl$_2$, (Life Technologies)], 0.5 µl of 0.1 M DTT (Life Technologies), 0.75 µl of 10 mM dNTPs, 0.75 µl of DEPC-treated ddH$_2$O (Appendix I) and 0.5 µl of *Superscript II* MMLV reverse transcriptase (200 U/µl, Life Technologies).

3. The 5 µl lysate was incubated at 80°C for 5 min to denature the mRNA, chilled immediately on ice and spun briefly in a microcentrifuge. The 5 µl RT reaction mix was added, the tube transferred from ice directly to 37°C and left to incubate for 1 h. The enzyme was inactivated with a 15 min incubation at 70°C and the tube cooled on ice.

4. Typically, half (5 µl) of the RT reaction was used as a cDNA template for PCR amplification with gene-specific primers and the remainder stored at -20°C.

5. PCR reactions were prepared as described in section 2.2.7a except that 0.25 units of *Taq* DNA polymerase were added and the final reaction volume was 25 µl. The reactions were denatured at 95°C for 5 min followed by 50 cycles of 95°C/30 s, 56°C/30 s, 72°C/1 min and a final extension at 72°C/15 min. Half of the reaction products were electrophoresed on an agarose gel for visualization (section 2.2.1d).

2.2.8 Rapid amplification of cDNA ends (RACE)

Protocols for 5'- and 3'-end cDNA amplification were used to elucidate transcript sequences for gene trap ES cell clones of interest. These are based on the original RACE method described by Frohman *et al.*, (1988) with streamlined modifications adapted from Frohman (1995) and Townley *et al.*, (1997). Oligonucleotide primer numbers listed in parentheses refer to those on the 5'-RACE schematic in Figure 4.2. All primer sequences are given in section 2.1.3.

(a) 5'-end cDNA amplification

All 5'-RACE reactions were performed at least twice for each cell line in independent experiments, with repeated 2nd round PCR reactions run for each as required either to rerun sequencing reactions, or to confirm an absence of amplifiable product.

81

1.   The 1st strand cDNA was reverse transcribed from poly A$^+$ mRNA template (section 2.2.2f) annealed with a gene-specific RT primer (oligo 1) from the gene trap vector or endogenous sequence of interest, in a 20 µl volume as described in section 2.2.7b.

2.   The RT reaction was stopped by alkaline hydrolysis with the addition of a 0.1x volume of 1 M NaOH and incubating at 65°C for 20 min, then neutralizing with a 0.1x volume of 1 M HCl.

3.   RNA, salts and excess primer were removed by loading the 1st strand reaction onto a 0.025 µm microdialysis filter (Millipore) floating on 10 ml of TE buffer in a Falcon 6 cm petri dish (Becton Dickinson), and left at RT°C to dialyse over 4 h. The cDNA was collected in 20 µl of ddH$_2$0. Alternatively, the reaction was diluted to 400 µl in TE buffer and purified on an Amicon *Centricon-100* spin filter (Millipore) with 2x 400 µl TE washes, according to the manufacturer's instructions. The eluate was concentrated in a Jouan *RC10.10.* centrifugal vacuum drier (Acim Jouanin, Evreux, France) for ~5 h at 30°C, and the sample brought to 20 µl in ddH$_2$0.

4.   The 1st strand cDNA was A-tailed by adding 6 µl of a 5x terminal deoxynucleotidyl transferase (TdT) buffer [0.5 M potassium cacodylate (pH 7.2), 10 mM CoCl$_2$, 1 mM DTT, (Life Technologies)] and 2 µl of 2mM dATP, then incubating at 37°C for 2 min. 30 units of TdT (15 U/µl, Life Technologies) were added and the sample incubated for a further 5 min at 37°C, followed by 70°C for 2 min.

5.   Half (15 µl)of the A-tailed cDNA was used directly for 2nd strand synthesis in a 20 µl reaction containing 1x restriction enzyme buffer M (Boehringer-Mannheim), 0.5 mM dNTPs, 10 ng of a T-tailed anchor primer (oligo 2-dT) and 2 units of Klenow enzyme (2 U/µl, Boehringer-Mannheim). The reaction was incubated at RT°C for 30 min, then 37°C for 30 min and 70°C for 5 min.

6.   To remove salts and primers and also to size select cDNAs above 300 bp in length, the 2nd strand reaction was microdialysed as for step 3, but on a 0.1 µm filter (Millipore) and collected in 36 µl of ddH$_2$0.

7.   A 1st round nested PCR amplification reaction was prepared in a 50 µl volume as described in section 2.2.7a, with the addition of 10^{-5} M tetramethylammonium chloride

(TMAC) and 5 units of *Taq* DNA polymerase. The forward primer used was the anchor primer (oligo 2) and the reverse was a gene-specific primer (oligo 3) nested 5' of the RT primer. The reactions were amplified for 30 cycles of 94°C/90 s, 60°C/90 s, 72°C/3 min, with a final extension at 72°C/10 min.

8.   1st round PCR products were dialysed on a 0.1 µm filter as for step 6 and collected in 40 µl of ddH₂0.

9.   A 2nd round nested PCR amplification reaction was prepared as for step 7, using 5 µl of the dialysed 1st round products. Where the 2nd round products were to be cloned (section 2.2.1), the anchor primer (oligo 2) was again used as the forward primer. If the products were to be directly sequenced by capture onto streptavidin-coated beads (section 2.2.9a) a 5'-biotin-labelled anchor primer (oligo 2-b*) was used. The reverse gene-specific primer (oligo 4) was nested 5' of oligo 3 in both cases. PCR conditions were as for the 1st round of PCR (step 7).

10.   2nd round PCR products were dialysed on a 0.1 µm filter as for step 6, collected in 40 µl of ddH₂0 and stored at -20°C.

11.   To confirm the presence of gene-specific 5'-RACE products, 3 µl of the 2nd round yield was electrophoresed on a 1.2% agarose gel (section 2.2.1d), transferred to a nylon membrane under alkaline dry blotting conditions (section 2.2.3c), and hybridized (section 2.2.6a) with a ³²P end-labelled primer (section 2.2.5b) nested upstream of the 3'-end of PCR products.

(b)   3'–end cDNA amplification

1.   The 1st strand cDNA was reverse transcribed from poly A⁺ mRNA template (section 2.2.2f) annealed with the T-tailed anchor primer (oligo 2-dT) in a 20 µl reaction containing the same mix as described in section 2.2.7b, but with the following modifications. The T-tailed primer was not annealed directly with the RNA, but instead added to the RT reaction mix, while the template was incubated in the presence of 10 units RNase Inhibitor (Life Technologies) at 80°C for 3 min, cooled rapidly on ice and briefly spun in a microcentrifuge. RNA was added to the RT mix and incubated at RT°C for 5 min, then 42°C for 1 h and 50°C for 30 min.

2. The RT reaction was stopped with a 15 min incubation at 70°C and the RNA template destroyed by adding 1.5 units of RNase H (2 U/µl, Life Technologies) and incubating at 37°C for 20 min.

3. The 1st strand reaction mix was diluted to 1 ml with TE buffer and stored at 4°C as a 3'-end cDNA pool. Since no A-tailing reaction was necessary, amplification steps were directly proceeded to.

4. For 2nd strand synthesis and 1st round PCR amplification a 50 µl PCR reaction was prepared as described in section 2.2.7a with the addition of 10^{-5} M TMAC. *Taq* DNA polymerase was not included in the PCR cocktail, but 5 units were added under hot start conditions. The forward primer used was a gene-specific primer upstream in the sequence of interest and the reverse primer was the anchor primer (oligo 2).

1 µl of the 3'-end cDNA pool was added to the PCR mix, incubated at 98°C for 5 min and cooled to 75°C. The *Taq* DNA polymerase was added and the mix incubated at 56°C for 2 min. The cDNAs were extended at 72°C for 40 min. The reaction was amplified for 30 cycles of 94°C/60 s, 56°C/60 s, 72°C/3 min, with a final extension at 72°C/15 min.

5. A 1:1000 dilution in TE buffer of the 1st round amplification products was used as template cDNA for the 2nd round of nested PCR. The reaction was prepared as for step 4, but with a forward gene-specific primer nested downstream of that used in the previous step. The reverse primer was again the anchor primer (oligo 2). Amplification conditions were the same as for step 4, eliminating the 2 min annealing and 40 min extension steps.

6. 3'-end RACE products were visualized by agarose gel electrophoresis (section 2.2.1d) and purified for cloning (section 2.2.1).

2.2.9 DNA sequencing

Thermal cycle sequencing methods employing the dideoxy-mediated chain-termination sequencing technique (Sanger *et al.,* 1977) were used to sequence RACE and PCR products for gene trap ES cell clones of interest. Electrophoresis gels were manually analysed for products directly sequenced with a radioactive end-labelled primer from the gene trap vector, while an automated *ABI Prism 377XL* DNA sequencer system (Perkin-

Elmer Applied Biosystems) was used to analyse sequencing reactions for cloned products incorporating fluorescent dye terminators.

(a) Direct sequencing of 5'-RACE products

Rapid sequence analysis of gene trap ES cell clones was performed using a method for the direct sequencing of biotinylated 5'-RACE products, based on that published by Townley *et al.*, (1997).

1. To prepare streptavidin-coated *Dynabeads* (Dynal) a 2x bead wash solution [10 mM Tris-HCl (pH 7.5), 1 mM EDTA (pH 8.0), 2 M NaCl] was made. For each template, an *MPC* magnetic tube holder (Dynal) was used to remove the storage supernatant from 20 µl (200 µg) of beads in 0.5 ml PCR grade tubes. Beads were resuspended in 20 µl of 1x bead wash solution, mixing gently. The wash was removed using the magnet and beads resuspended in 40 µl of 2x bead wash solution.

2. Biotinylated PCR products in a 40 µl volume - [typically $1/10^{th}$ of 2nd round PCR products from 5'-RACE reactions (section 2.2.8a), diluted in ddH_2O] - were immobilized by adding to the 40 µl of washed beads. The samples were incubated at RT°C for 15-30 min with frequent mixing to keep the beads in suspension.

3. The supernatant was removed with the aid of the magnet and beads washed with 40 µl of 1x bead wash solution – (samples could be stored at 4°C for several weeks at this step). To denature DNA, the wash was removed using the magnet, and the beads were incubated in 8 µl of fresh 0.1 M NaOH for 10 min at RT°C.

4. To remove the unbound (non-biotinylated) DNA strand, the beads were washed once in 50 µl of 0.1 M NaOH, once in 40 µl of 1x bead wash solution and once in 50 µl of TE buffer, using the magnet. Beads were resuspended in 25 µl of ddH_2O.

5. 10 pmol of the gene trap vector sequencing primer (oligo 5, Figure 4.2 & section 2.1.3) was end-labelled with 40 µCi of $[\gamma^{32}P]$ ATP or $[\gamma^{33}P]$ ATP in a 25 µl reaction volume, as described in section 2.2.5b.

6. Using a Perkin-Elmer *AmpliCycle* sequencing kit (Roche Molecular Systems, Inc., Branchburg, NJ), 2 µl of each termination mix (ddGTP, ddATP, ddTTP, ddCTP) was dispensed into a 96-well thin-walled plate (Perkin-Elmer), and kept on ice.

7.   A mix was prepared with the 25 µl template DNA/beads suspension, 1 µl (0.4 pmol) of the end-labelled sequencing primer (step 5) and 4 µl of a 10x cycling mix [500 mM Tris-HCl (pH 8.9), 100 mM KCl, 25 mM MgCl$_2$, 0.25% (v/v) Tween, 1 unit of AmpliTaq CS DNA polymerase (Perkin-Elmer/Roche)]. 6 µl of this mix was dispensed to each of the four termination wells and the reactions covered with a microseal film (Perkin-Elmer).

8.   The tray was placed on a hotplate in a Hybaid Omnigene PCR instrument pre-warmed to 95°C, and reactions denatured for 2 min at 95°C. 30 cycles of sequencing were performed at 95°C/60 s, 60°C/60 s, 72°C/60 s and 4 µl of a stop solution [95% formamide, 20 mM EDTA, 0.05% bromophenol blue, 0.02% xylene cyanol FF, (Perkin-Elmer/Roche)] added when cycling was complete.

9.   Samples were heated to 95°C for 3 min and 2.5 µl of each was electrophoresed through a 0.2 mm 6% denaturing polyacrylamide gel, prepared with Sequagel products (National Diagnostics, Atlanta, GA) on a Bio-Rad apparatus according to manufacturers' instructions. Electrophoresis of samples was performed in 1x TBE buffer (Appendix I) for ~2 h at 50 watts after pre-running the gel at 100 watts for 10 min.

10.   Following electrophoresis, the gel was transferred to Whatman 3MM paper and dried under vacuum at 80°C on a gel drier (Hoefer Scientific) for 40 min. Sequencing products were manually analysed following o/n exposure to X-omat AR5 film (Kodak) at RT°C, without an intensifier screen.

(b)   Automated sequencing of cloned PCR products

RACE and PCR products cloned into pCR-BLUNT and pGEM-T plasmids (section 2.2.1) were prepared for sequence analysis using the ABI PRISM BigDye Terminator cycle sequencing ready reaction kits (Perkin-Elmer/Roche), according to the manufacturer's instructions. 250 ng of purified small-scale plasmid DNA (section 2.2.2a) was prepared in a 10 µl sequencing reaction volume with the BigDye Terminator ready mix - containing the fluorescent-labelled dye terminators, dNTPs and AmpliTaq FS DNA polymerase - and 0.5 µl (3.2 pmol) of the sequencing primer (for either the gene trap vector or plasmid sequences), in 0.5 ml PCR tubes. 25 cycles of sequencing

were performed at 96°C/30 s, 50°C/20 s and 60°C/4 min in a Hybaid *Omnigene* PCR instrument. Extension products were purified by ethanol precipitation as specified by the manufacturer and resuspended in the supplied loading buffer for electrophoresis and automated analysis on the *ABI Prism 377XL* DNA sequencer system (Perkin-Elmer). *ABI Prism* systems at both the Centre for Genome Research (Edinburgh University, UK) and Prince Henry's Institute (Clayton, Australia) were used.

(c) Sequence analysis

For analysis of nucleic acid and translated sequences the BLAST programs (Altschul *et al.*, 1990; Altschul *et al.*, 1997) were used to search GenBank, EMBL, dbEST, and Swiss-Prot databases. The Jackson Laboratory Mouse genome database and the TIGR genome databases were searched for homologies. Multiple sequence alignments were performed using the CLUSTAL W (Feng and Doolittle, 1987) and MACAW (Schuler *et al.*, 1991) alignment programs. Motif searches were kindly performed by Dr. George Rudy of the Genetics and Bioinformatics Group, WEHI, (Parkville, Australia).

2.3 ES CELL CULTURE METHODS

All ES cell cultures were maintained at 37°C with 5% CO_2 in both Heraeus *Heracell* and Sanyo *MCO-17A1* (Sanyo, Osaka, Japan) humidified incubators. Methods for routine culture and differentiation of ES cells are based on those described in Smith *et al.*, (1991), except where stated otherwise. All tissue culture procedures were performed in an Email *air handling* class II biological safety cabinet, (BTR Environmental, Regents Park, Australia) and standard procedures for asceptic and sterile technique adhered to. Unless otherwise indicated, tissue culture grade plasticware was supplied by Falcon (Becton Dickinson) and Nunc. Routinely, centrifugation steps were performed in either a Sigma *2-4* rotor (Sigma Laborzentrifugen, Harz, Germany) or a Beckman *TJ-6* rotor. Cell cultures were inspected by phase-contrast microscopy using Zeiss *Axiovert 25* (Carl Zeiss Inc., Thornwood, NY) and Nikon *Eclipse TE300* (Nikon Inc., Melville, NY) microscopes.

2.3.1 Routine expansion and passage of ES cells

All ES cell lines were routinely cultured without feeder layers in GMEM culture medium (section 2.1.5) supplemented with LIF (Smith *et al.*, 1988, Williams *et al.*,

1988) at 10^3 U/ml (section 2.1.5). All solutions were pre-warmed to 37°C in a water bath. Flasks, dishes and wells were coated with 0.1% (w/v) gelatin (section 2.1.5) for at least 15 min and aspirated prior to use.

1.   Cultures seeded at 3-4 x 10^4 cells/cm^2 were inspected under phase-contrast optics every day. Medium was changed every 24-48 h until cultures approached confluency, ($\sim 10^7$ cells per 25 cm^2 tissue culture flask).

2.   To passage cultures, the culture medium was aspirated and cells rinsed twice with sterile PBS.

3.   Trypsin solution (section 2.1.5) was added to give a thin film over the cells (0.5 ml/25 cm^2) and the cultures incubated for 2-3 min at 37°C. Flasks were tapped to completely dislodge cells from the surface and checked under low-power magnification for small ES cell clumps.

4.   Culture medium (9 ml/25 cm^2) supplemented with LIF was added to stop the trypsin digestion and cells resuspended by pipetting 3-4 times against the flask wall, using the medium speed setting on a Drummond Scientific Co. *express pipet-aid* (Becton Dickinson). The suspension was transferred to a Falcon 15 ml conical tube, or to a 30 ml conical tube (Sarstedt, Rommelsdorser, Germany) for larger cultures, and pelleted by centrifugation at 1000 rpm for 5 min at RT°C.

5.   The cell pellet was resuspended in 5-10 ml of culture medium supplemented with LIF and cells counted on a haemocytometer. A freshly gelatinized culture vessel containing culture medium supplemented with LIF (8-10 ml/25 cm^2) was seeded as in step 1 and transferred to the CO_2 incubator.

2.3.2 Freezing ES cells

1.   Cells were harvested and centrifuged as described in section 2.3.1. Cells were resuspended in 10 ml of culture medium supplemented with LIF and counted on a haemocytometer.

2.   The suspension was again centrifuged at 1000 rpm for 5 min and the cell pellet resuspended in cold freezing medium [culture medium containing LIF with DMSO freshly added at 10% (v/v) and filter sterilized], at a concentration of 1-3 x 10^6 cells/ml.

3.   The freezing suspension was quickly dispensed to cryotubes (Nunc) in 1 ml aliquots and placed in a polystyrene container at -80°C for 24 h before transferring to liquid nitrogen for long-term storage.

A convenient modification of the above protocol was devised for the short-term frozen storage of large numbers of ES cell clonal cultures generated by gene trap electroporation experiments (section 2.3.4).

*   Culture medium was aspirated from ES cell cultures grown in a 12-well plate (3.8 cm^2/well) and rinsed twice with sterile PBS.

*   100 μl of trypsin solution (section 2.1.5) was added to each well and decanted after 20 s, prior to dislodgement of cells from the surface. 1 ml of the above freezing solution (step 2) was added to each well and plates stored in a polystyrene container at – 80°C for up to 1 week, or transferred to the gaseous phase of a liquid nitrogen storage tank.

2.3.3 Thawing ES cells

1.   Frozen vials were retrieved from liquid nitrogen storage onto dry ice and quickly thawed in a 37°C water bath. Vials were wiped with 70% (v/v) ethanol and thawed cells gently transferred to a 15 ml tube containing 10 ml of culture medium supplemented with LIF. Cells frozen in 12-well plates were thawed in the same manner, except that the culture medium was added to the wells to completely dislodge and recover the cells.

2.   Cells were pelleted at 1000 rpm for 5 min, the supernatant aspirated, and cells gently resuspended in 10 ml of fresh culture medium supplemented with LIF. Cells were seeded into pre-gelatinized tissue culture vessels as required, and cultures transferred to the CO$_2$ incubator.

3.   After o/n incubation, the culture medium was changed to ensure complete removal of residual DMSO and cells maintained as described in section 2.3.1.

2.3.4 Transfection of ES cells by electroporation

Gene trap vectors were introduced into ES cells by electroporation - a means for gene transfer into eukaryotic cells, first described for lymphocytes by Potter *et al.*, (1984).

Electroporation involves exposure of a suspension of cells and cloned DNA to a high-voltage electric discharge, resulting in the formation of membrane pores large enough to allow the entry of macromolecules such as DNA.

(a)   Vector preparation

1.   Plasmid DNA for electroporation was prepared by the CsCl gradient purification of a large-scale preparation, as described in section 2.2.2b.

2.   150 µg of gene trap vector DNA was linearized by o/n incubation with the restriction enzyme *Sal* I in a digest reaction volume of 400 µl (section 2.2.1c).

3.   Digested DNA was precipitated by bringing the final salt concentration of the reaction mix to 0.3 M with 3 M sodium acetate (Appendix I), adding 2.5 volumes of cold absolute ethanol and incubating at -20°C for 30 min. The DNA was pelleted and washed twice with 70% (v/v) ethanol as described in section 2.2.1b and the ethanol drained off.

4.   After complete evaporation of ethanol traces by air drying in a hood, the pellet was resuspended in sterile PBS to give a 1 µg/µl solution of linearized DNA. This could be stored at -20°C if not being immediately transfected into cells.

(b)   Electroporation of ES cells

1.   A routine ES cell culture (section 2.3.1) was established in a 175 cm^2 flask containing 50 ml of culture medium supplemented with LIF and seeded with 2.5-5 x 10^6 cells. Media was changed every 24-48 h until the culture was ~85% confluent.

2.   Prior to electroporation, the culture was incubated for 2-3 h in a fresh 25 ml change of culture medium. Cells were harvested (section 2.3.1) using 2 ml of trypsin solution and the digestion inactivated with 8 ml of culture medium containing LIF. Cells were pelleted at 1000 rpm for 5 min, resuspended in 10 ml of sterile PBS and counted. After re-pelleting, cells were resuspended at 50 x 10^6 cells in 700 µl of sterile PBS.

3.   60 µg of linearized DNA (section 2.3.4a) dissolved in 100 µl of sterile PBS was mixed by gentle pipetting with the 700 µl cell preparation from step 2 and the suspension transferred to a 0.4 cm electroporation cuvette (Bio-Rad).

4.   The cuvette was pulsed in a *Gene Pulsar II* electroporation system (Bio-Rad) with an electrical discharge of 0.8 kV at 3 μF, which gave a time constant of 0.1 ms. Cells were left in the cuvette at RT°C for 5 min, then very gently transferred to a tube containing 9.2 ml of culture medium supplemented with LIF and placed at 37°C in a humidified CO_2 incubator for 20 min.

5.   The cell suspension was plated to pre-gelatinized 10 cm diameter tissue culture plates (78.5 cm^2) containing 10 ml of pre-equilibrated culture medium supplemented with LIF, at 5×10^6 cells per plate and incubated o/n.

6.   After 18 h, media was aspirated from the plates and replaced with culture medium supplemented with LIF and containing G418 at 200 μg/ml (section 2.1.5). This was repeated daily for 3-4 days until most of the cells had died and small G418 resistant colonies had started to appear. The selection medium was then changed every 48 h.

(c)   Picking G418 resistant colonies

G418 resistant ES colonies were well established after ~8 days of culture in selection medium and were individually picked for expansion, freezing and further analyses. Clonal cultures were maintained in culture medium supplemented with penicillin/streptomycin antibiotics (section 2.1.5) until frozen stocks were established and successfully thawed.

1.   For each colony to be picked, one well in each of duplicate 24-well plates (2.0 cm^2/well) and a 12-well plate (3.8 cm^2/well) were gelatinized. Culture medium containing LIF was equilibrated in each well (1 ml/2.0 cm^2, 2 ml/3.8 cm^2), and G418 added at 300 μg/ml (section 2.1.5) where selection experiments were performed.

2.   96-well plates (0.32 cm^2/well) containing 50 μl aliquots of trypsin solution (section 2.1.5) were pre-warmed in the incubator.

3.   Selection media was aspirated and the 10 cm transfection plate rinsed with 5-10 ml of sterile PBS. Using a Leica *MZ6* dissecting microscope (Leica Microscopy Systems, Wetzlar, Germany), colonies were quickly assessed and those to be picked were circled with a marker pen on the bottom of the plate. The plate was again rinsed with 5-10 ml of sterile PBS, aspirating off completely.

4.   Using a 10-100 µl Eppendorf pipettor and sterile 100 µl tip, a 10 µl drop was taken up from a 96-well containing trypsin solution and released against the edge of the colony to be picked. After 3 s the colony was easily aspirated and dispensed back into the trypsin well. 10-12 colonies were quickly picked in this manner and the plate of trypsin wells returned to the incubator for 3-5 min. The plate was tapped a few times to disperse cells and the digestion stopped by adding 100 µl of pre-warmed culture medium to each well.

5.   Each cell suspension was pipetted with the Eppendorf tip a few times to help break up the colony, split to the triplicate plates of pre-equilibrated culture wells (step 1), and incubated as for routine cultures. Cultures were grown to confluency with a change of culture medium every 24-48 h, adding G418 at 300 µg/ml for selection experiments. The 24-well cultures were used to identify βgal positive colonies with X-gal staining analysis (section 2.6.2), while the 12-well culture was maintained for the subsequent expansion (section 2.3.1) and freezing (section 2.3.2) of selected cell lines, and for further analyses.

2.3.5 *In vitro* differentiation of ES cells - embryoid body cultures

To analyse *in vitro* differentiation for gene trap ES cell lines of interest, embryoid bodies (EBs) were established in the absence of LIF from individual "hanging drop" aggregate cultures containing a small and defined number of cells. For Northern blot analysis of RNA (section 2.2.4) from differentiating cells, EBs were established on a larger scale by high density aggregation culture. Due to the exposed nature of EB work, penicillin/streptomycin antibiotics (section 2.1.5) were added to culture medium.

(a)   Hanging drop embryoid body culture

Individual aggregates were produced by plating ~300 cells in 20 µl microdrop volumes using a protocol modified from that published by Mountford *et al.*, (1998).

1.   A routine ES cell culture (section 2.3.1) was grown to ~80% confluency in either a 25 cm^2 flask or a 6-well plate (9.6 cm^2/well) and cells harvested following a 2-3 h incubation in a fresh change of culture medium containing LIF. Trypsinization was as described in section 2.3.1 except that the digestion was stopped with culture medium not supplemented with LIF.

2.   Cells were pelleted at 1000 rpm for 5 min, resuspended in a fresh 10 ml aliquot of culture medium without LIF, and counted. An aliquot containing 30,000 cells (for up to 100 drops) was transferred to a 5 ml tube and made up to a final volume of 2 ml (100 x 20 μl drops) in culture medium without LIF. Where selection experiments were peformed, G418 was added to the final volume at 300 μg/ml.

3.   Gently, cells were well resuspended and 40-50 x 20 μl microdrops plated to the upturned lid of a 10 cm bacteriological dish (Sarstedt), using an Eppendorf multipipettor with a sterile 500 μl tip.

4.   The lid was carefully inverted over the dish base, containing enough ddH_2O to cover the surface and maintain local humidity, preventing evaporation of the drops. The hanging drops were incubated for 3 days at 37°C with 5% CO_2.

5.   To prevent EBs adhering to tissue culture wells, these were coated with a 1% agarose solution prepared by mixing cooled molten 2% sterile agarose with an equal volume of a 2x preparation of culture medium (section 2.1.5) without LIF - (adding G418 at 600 μg/ml in the case of selection experiments). The agarose mix was quickly delivered to wells (1 ml per 12-well/3.8 cm^2 or 100 μl per 48-well/0.75 cm^2), allowed to set for 1 h in the laminar flow hood and plates stored in *Parafilm M* laboratory film (American National Can., Neenah, WI) at 4°C until day 3.

6.   After 3 days, aggregates had formed simple EB structures in the hanging drops and these were transferred with a wide-bore pasteur pipette to the agarose-coated plates, pre-equilibrated with culture medium not containing LIF - (adding G418 at 300 μg/ml for selection experiments). EBs were cultured either singly in a 48-well plate or pooled to a 12-well plate for a further 7 days, replacing the culture media each 48 h by swirling EBs to the centre of the wells and carefully aspirating off spent media.

7.   After 10 days of suspension culture, the enlarged and often cystic EBs were trypsinized (section 2.3.1) using 100 μl trypsin solution and a 5-10 min incubation at 37°C. The digestion was stopped with 200 μl of culture medium containing LIF and cell clumps pipetted vigorously with a 1000 μl Eppendorf tip to disperse cells. The suspension was plated to gelatinized tissue culture wells (24-well/2.0 cm^2) containing pre-equilibrated culture medium with LIF, and cultured for 48 h. Recovered cell types

were assessed histochemically by Leishman's staining (section 2.6.1) for cell morphology, or by X-gal staining for βgal expression (section 2.6.2).

(b)   High density aggregation culture

Large populations of undifferentiated ES cells were induced to differentiate in aggregation cultures, using a protocol adapted from that described by Robertson, (1987).

1.   A routine ES cell culture (section 2.3.1) was grown to ~80% confluency in a 175 cm^2 flask and cells harvested following a 2-3 h incubation in a fresh change of culture medium supplemented with LIF. Cells were harvested in 2ml of trypsin solution (section 2.3.1) and the digestion stopped by adding 13 ml of culture medium without LIF. In order to maintain ES cells in clumps, the harvest was handled with gentle and minimal pipetting.

2.   Cells were pelleted at 1000 rpm for 5 min in a 30 ml conical tube (Sarstedt), gently resuspended in a fresh 10 ml aliquot of culture medium without LIF, and counted. 8-10 x 10^6 cells were plated to a 10 cm bacteriological dish (Sarstedt) containing 10 ml of culture medium without LIF, and cultured at 37°C with 5%CO_2. Suspended aggregates formed within 24 h and plates were gently shaken every 24 h to prevent adhesion.

3.   Every 48 h, aggregates were sedimented by gravity after gently transferring to a 30 ml conical tube (Sarstedt). Media was aspirated, replaced with a fresh 10 ml aliquot and aggregates/EBs transferred to a new bacteriological dish.

4.   For a time course analysis of genes expressed in ES cells, poly A^+ mRNA was prepared (section 2.2.2f) from EBs collected after 2, 4, 6 and 8 days of culture without LIF. Ample RNA was yielded for Northern blot analysis (section 2.2.4) from 10 culture plates each for days 2 and 4, and from 6 plates each for days 6 and 8.

2.4   EMBRYOLOGICAL METHODS

Methods for the collection and culture of mouse embryos were based on those described by Hogan et al., (1994), unless otherwise stated. For hepes-buffered flushing media, M2 + BSA (Quinn et al., 1982) and Gardiner's HEPES-G1 medium (Barnes et al., 1995) were used. Gardiner's G1 and G2 media (Barnes et al., 1995) were used for in

vitro embryo culture, except for blastocyst explants (section 2.4.3b) which were cultured in a supplemented Dulbecco's modified Eagle's medium (DMEM). Gardiner's culture media were prepared by Ms. Megan Munsie (MIRD).

2.4.1 Embryo collection

Adult females were mated o/n and fertilization assumed to have occurred in the middle of the dark period (0200 h) preceding a vaginal copulation plugging. Where large numbers of preimplantation stage embryos were required for experiments, prepubescent females of 3-5 weeks age were administered 5 IU of pregnant mare's serum gonadotrophin (*Folligon*, Intervet, Boxmeer, The Netherlands) at 1600 h, followed 48 h later with 5 IU of human chorionic gonadotrophin (hCG), (*Chorulon*, Intervet), to induce superovulation prior to mating. In this case ovulation was presumed to have occurred 10-12 h after the hCG injection. For both natural and superovulated matings, embryos were staged as 0.5 dpc at 1400 h on the day on which the vaginal plug was observed.

0.5 dpc embryos were collected by slicing the oviduct ampullary region and dissecting cumulus from fertilized eggs in Gardiner's HEPES-G1 medium containing 300 µg/ml sterile hyaluronidase type II (Sigma). Fertilized eggs were rinsed in Gardiner's HEPES-G1 medium and cultured *in vitro* (section 2.4.3a) as required, for stages through to expanded blastocyst development.

3.5 dpc, 4.5 dpc and implantation-delayed (section 2.4.2) embryos were flushed from the uterine horns with M2 + BSA medium or with Gardiner's HEPES-G1 medium.

Postimplantation stage embryos up to 7.5 dpc were dissected from uterine decidua in cold PBS containing 1% FCS, as described by Beddington, (1987). Instruction in these techniques by Professor Marilyn Monk (visiting fellow, MIRD) is gratefully acknowledged. Later stages (9.5 dpc up to 15 dpc) were dissected from the uterus in cold PBS/1% FCS and separated from extraembryonic membranes, as described in Hogan *et al.*, (1994).

For each stage of development, collected embryos were either assessed for βgal activity by X-gal staining (section 2.6.2) or prepared for RT-PCR analysis (section 2.2.7c).

2.4.2 Delay of embryo implantation

Implantation-delayed blastocysts were obtained from plugged female mice which were ovariectomized at 2.5 dpc and injected subcutaneously with 1 mg of progesterone, (*Depo-Provera* 10mg/ml in sterile PBS, Sigma). Hatched blastocysts were recovered 4-7 days following ovariectomy by gentle flushing of the uterine horns with Gardiner's HEPES-G1 medium. βgal activity in implantation-delayed blastocysts was assessed by X-gal staining (section 2.6.2).

2.4.3 *In vitro* culture of embryos

(a)   Preimplantation stage embryos

Fertilized eggs collected at 0.5 dpc (section 2.4.1) were transferred to microdrops of Gardiner's G1 culture medium under sterile mineral oil (Sigma) in a 35 mm dish (Falcon) and cultured at 37°C with 5% CO_2 in a Heraeus *Heracell* humidified incubator. Embryos were cultured to the 4-cell stage of development (~48 h) in this media. For the ongoing culture of embryos through to the hatching blastocyst stage, 4-cell embryos were transferred to a new dish with equilibrated microdrops of Gardiner's G2 culture media and cultured for a further 48 h. Embryos at all stages were either assessed for βgal activity by X-gal staining (section 2.6.2) or prepared for RT-PCR analysis (section 2.2.7c).

(b)   Blastocyst explant culture

Blastocysts collected at 3.5 dpc (section 2.4.1) were transferred singly to 16 mm (4-well) Nunc dishes pre-coated with 0.1% gelatin (section 2.3.1) and cultured in 1x DMEM [5x DMEM stock (Trace), 2.2 g/L sodium bicarbonate, 2 mM L-glutamine (Life Technologies)] supplemented with 20% FCS (section 2.1.5), 0.1 mM β-mercaptoethanol (Sigma) and 1 mM non-essential amino acids (Life Technologies). Recombinant LIF (section 2.1.5) was added to culture media at 1000 U/ml and G418 at 400 µg/ml (Life Technologies) where required. Explants were cultured at 37°C with 5%CO_2 for up to 6 days after plating. Media was refreshed on day 3 following hatching of the blastocyst and attachment of explant growth to the surface of the dish. βgal activity in explant cultures was assessed by X-gal staining (section 2.6.2).

2.4.4 Dissection of mouse tissues

Tissues were dissected with sterilized instruments and rinsed in cold PBS/1% FCS for X-gal staining (section 2.6.2) and wax sectioning (section 2.6.3), or for the isolation of RNA (sections 2.2.2e & f). A range of tissues were dissected from transgenic and wild type mice at 2-3 days following birth (assisted by Mr. Joseph Bianco, MIRD), and from adult mice at 7-10 weeks of age, with reference to *The Atlas of Mouse Development* (Kaufman, 1992). Genital ridges and mesonephros were dissected from 11.5, 12.5 and 13.5 dpc embryos as described by Hogan *et al.*, (1994), in cold PBS/1%FCS (assisted by Dr. Anna Michalska, MIRD). Poly A$^+$ mRNA prepared from a range of postimplantation stage, neonate and adult wild type tissues was also kindly provided by Dr. Paul Hertzog's laboratory, (MIRD).

2.5   GENERATION OF CHIMAERIC MICE

Chimaeric mice were produced by blastocyst injection of gene trap ES cell lines, with modifications of the procedures outlined by Bradley (1987). Assistance from Dr. Anna Michalska (MIRD), particularly with embryo transfers, is gratefully acknowledged.

2.5.1 Preparation of the injection chamber

Blastocysts were flushed at 3.5 dpc (section 2.4.1) from naturally mated inbred C57Bl/6J mice (section 2.1.6), using M2 + BSA medium (Quinn *et al.*, 1982) and held at RT°C for injecting. Any unexpanded embryos were transferred to M16 + BSA medium (Whittingham, 1971) pre-equilibrated in microdrops under mineral oil and cultured for up to 24 h in a humidified incubator at 37°C with 5% CO_2, to allow expansion. For injection, expanded blastocysts were transferred to a concave slide chamber containing a microdrop of M2 + BSA medium overlaid with mineral oil and kept at RT°C. ES cells, trypsinized after a 2 h incubation in a fresh change of culture medium (section 2.3.1) were placed in the above drop as single cells or pairs and replenished during the injection procedure, where clumping occurred over time.

2.5.2 Blastocyst injections

Injections were performed with a rounded holding pipette and a bevelled injection pipette with a sharp spiculated tip, prepared by Ms. Megan Munsie (MIRD) from thin

walled glass capillary tubing, ($\varnothing_{external}$ = 100 µm, Clarke Electro Medical, Reading, UK). Holding pipettes were prepared manually by flaming capillary tubing over a bunsen burner, pulling apart from the ends and snapping the tip off at an external diameter of ~150-200 µm. The walls of the tip were melted using a *De Fonbrune* microforge (Beaudouin, Paris, France) to produce a smooth rounded tip with an internal diameter of 10-20 µm and a 30° bend placed ~4 mm from the tip. Injection pipettes were prepared by mechanically pulling the capillary tubing with a *Flaming/Brown P-87* horizontal micropipette puller (Sutter Instruments, Novato, CA). Settings used were 320°C for 200 µs with a pulling velocity of 40 units and a repeat exposure at the same temperature at 50 units velocity. This resulted in a pipette with a long tapered end (4mm tip with a diameter of <10 µm), which was removed using the microforge to give an internal diameter of 25 µm. Using a capillary tip grinder (Bachofer Laboratoriumsgerate, Reutlingen, Germany), a 45° bevel was ground onto the blunt end of the pipette and a sharp spicule added for zona pellucida penetration by melting the lower bevel edge with the microforge glass bead at a low heat setting. A 30° bend was placed ~4 mm from the pipette tip.

Holding and injection pipettes were each attached to Leitz instrument holders and operated by Leitz manipulators (Leica Microscopy Systems). The suction of both pipettes was controlled by Hamilton micrometre syringes connected by small-diameter silicone tubing filled with mineral oil. An *Eclipse TE300* inverted phase microscope (Nikon) was integrated into the micromanipulation assembly.

For each blastocyst 10-20 ES cells were injected into the blastocoel cavity with the bevel of the pipette facing upwards toward the inner cell mass. Injected embryos were allowed to recover for 2-3 h in GMEM ES cell culture medium (section 2.1.5) in a humidified incubator, before transferring to recipient females.

2.5.3 Uterine embryo transfer

Pseudopregnant recipient females were of an F1 stock from C57Bl/6J and CBA matings (section 2.1.6) and had been mated with a vasectomized F1 male of the same background 2.5 days earlier. Unilateral uterine transfers were performed with 5-12 injected embryos transferred to each side.

2.5.4 Germline transmission

Contribution of transgenic ES cells to resulting offspring was estimated by coat colour. As the ES cells are derived from the 129/Ola strain of mice (section 2.1.5) which carry the dominant agouti allele, and the C57Bl/6J mouse strain (section 2.1.6) is homozygous recessive at this locus, the appearance of agouti, sandy or white coloured hairs amongst the otherwise black hairs of the C57Bl/6J host blastocysts indicated successful ES cell incorporation into the resulting offspring. To test for germline transmission of the ES cells chimaeras were paired with C57Bl/6J mice (section 2.1.6). Generation of agouti offspring indicated successful transmission of germ cells derived from the ES cells, whereas black offspring were indicative of transmission from the host blastocyst. To identify those agouti offspring harboring a gene trapped allele (~50%), PCR analysis was performed (section 2.2.7a) on genomic DNA prepared from ear clips taken at weaning age (section 2.2.2d).

2.6 HISTOLOGICAL METHODS

2.6.1 Leishman's staining of ES cells

For morphological examination of ES cells, cultures were fixed and stained with Leishman's using a protocol adapted from that described by Smith, (1991).

1. A 0.15% (w/v) solution of Leishman's (Sigma) was prepared in 100% methanol, dissolved using a magnetic stirrer with heat, and filtered through Whatman 3MM paper.

2. Medium was aspirated from cultures and cells rinsed twice with PBS. A volume of the Leishman's solution was added to wells, enough to cover the surface of cells (~0.5 ml per 2 cm^2 well) and left at RT°C for 15 min.

3. Two volumes of ddH$_2$0 were added to the stain and left on the cells for 5 min. The stain mix was decanted, wells gently rinsed a few times with ddH$_2$0 and left to air dry.

4. For light box or microscropy examination, cells were covered with ddH$_2$0, or with sterile glycerol if storing for longer periods.

2.6.2 X-gal staining of cells, tissues and embryos

βgal activity was assessed for cells, tissues and embryos by X-gal staining, based on the protocol published by Beddington *et al.*, (1989).

The following solutions were prepared for X-gal staining:

<u>0.5 M phosphate buffer, pH 7.5</u>

59.6 g Na_2HPO_4 (Sigma)

11.0 g $NaH_2PO_4.2H_2O$ (Sigma)

Made up to 1 litre in ddH_2O and stored at RT°C.

<u>X-gal rinse buffer</u>

0.1 M phosphate buffer (pH 7.3)

2 mM $MgCl_2$

Stored at 4°C

<u>Fixative</u>

10 ml	0.5 M phosphate buffer (pH 7.3)	(0.1 M)
0.4 ml	25% gluteraldehyde (Sigma)	(0.2%)
1 ml	0.25 M EGTA (Sigma)	(5 mM)
0.1 ml	1M $MgCl_2$	(2 mM)
1.4 ml	37% formaldehyde	(1%)

Made up to 50 ml in ddH_2O and stored at 4°C for up to 2 weeks.

<u>Detergents</u>

10% deoxycholate stock solution

10% NP40 stock solution

<u>X-gal (5-Bromo-4-chloro-3-indolyl-β-D-galactopyranoside)</u>

50 mg/ml stock dissolved in dimethylformamide (Sigma). Stored in the dark at −20°C.

<u>Ferrous stock solutions</u>

0.5 M (1.65 g/10 ml PBS) potassium ferricyanide (Sigma)

0.5 M (2.11 g/10 ml PBS) potassium ferrocyanide (Sigma)

Stored in the dark at RT°C.

<u>X-gal staining solution</u>

Prepared by adding the following to 960 μl of X-gal rinse buffer:

10 μl	0.5 M potassium ferricyanide	(5 mM)
10 μl	0.5 M potassium ferrocyanide	(5 mM)
20 μl	50 mg/ml X-gal	(1 mg/ml)

The staining solution was prepared fresh and filtered through a 0.22 μm filter (Millipore) to remove crystals prior to use.

1.   Cells were washed in PBS and fixed for 5-20 min (depending on size) at 4°C. Embryos and tissues were dissected and rinsed in cold PBS (with 1% FCS added for small tissues and embryos up to 7.5 dpc requiring glass pipette transfer). Embryos/tissues were fixed - (with the addition of 0.02% NP40 for postimplantation stage embryos up to 11.5 dpc; 0.8% NP40 and 0.1% sodium deoxycholate for embryos older than 11.5 dpc and for tissues) - for 20 min (7.5 dpc), 30 min (9.5 dpc) up to 60 min (15.5 dpc and tissues). Embryos 12.5 dpc and older were cut sagitally for improved penetration of fixative and stain.

2.   Cells were rinsed 3x for 5 min each, embryos and tissues 3x for 20 min each in X-gal rinse buffer (with 1% FCS added for embryos/tissues requiring glass pipette transfer), at RT°C.

3.   Cells, embryos and tissues were placed in X-gal staining solution - (with the addition of 0.8% NP40 and 0.1% deoxycholate for embryos older than 11.5 dpc and for tissues) - in a humidified chamber placed at 37°C. Cells and embryos were left to stain o/n, while embryos older than 12.5 dpc and tissues were left for 36-48 h. Blue staining denoted βgal enzyme activity.

4.   Following staining, cells, embryos and tissues were rinsed in PBS and stored in 10% PBS buffered formalin at 4°C. Preimplantation embryos alternatively, were stored in PBS/1% FCS microdrops under paraffin oil at 4°C.

2.6.3 Wax sectioning of X-gal stained tissues

Following whole-mount X-gal staining for βgal activity (section 2.6.2), embryos and tissues to be analysed histologically were fixed for at least 24 h in 10% PBS buffered formalin at 4°C. Samples could be left as such for several months (if required), prior to processing for paraffin wax embedding. Wax sections were kindly prepared by Ms. Ann Davies, histology laboratory (MIRD).

(a)   Histological processing

Fixed tissue sections/embryos were placed in 3-4 changes of 70% ethanol over 3 days before placing in a Leica TP 1020 automatic tissue processor (Leica Instruments, Nussloch, Germany). Tissues were processed through the following solutions, for 30 min each for small tissues, and 60 min each for larger tissues.

1 x 70% ethanol

3 x 100% ethanol

3 x thinner solvent 3B/2026 (HIChem Industries, Hallam, Australia)

3 x wax

Sections were embedded in paraffin wax using a Leica EG 1160 embedder (Leica Instruments). Wax sections of 5-10 μm thickness were cut on a Leica RM 2135 microtome (Leica Instruments), and deposited on *Superfrost Plus* positively charged microscope slides (Erie Scientific Co., Portsmouth, NH).

(b)   Eosin counterstaining

As the gene trap vector used in this study provided nuclear localization of βgal activity for a trapped endogenous gene, eosin was used to counterstain the cytoplasm of whole-mount X-gal stained tissues. Sections were dewaxed in 2 changes of solvent 3B/2026 (HIChem Industries) for 4-5 min each, then rehydrated through an ethanol series of 2 x 100%, and 70% for 4-5 min each, and rinsed briefly in tap water. Sections were stained

for 5 min in 1% aqueous eosin (Amber Scientific, Belmont, Australia), and after another brief rinse in water, were dehydrated through 70% and 2 x 100% ethanol changes for 2 min each. Finally, sections were cleared in 2 changes of histolene solvent (Fronine, Riverstone, Australia), for 2 min each, then mounted in Gurr's DePeX mounting medium (BDH Laboratory Supplies), and coverslipped. Mounted sections were viewed and photographed under light-field optics, on a Lieica DMR digital microscopy system (Leica Microscopy Systems).

Chapter Three

GENE TRAP INTEGRATION & *IN VITRO* SCREENING FOR ES CELL-SPECIFIC GENES

3.1   INTRODUCTION

The growth and differentiation of ES cells is likely to be controlled by a group of key regulatory genes. It has been demonstrated that expression of the transcription factor *Oct-4* is necessary for establishing the pluripotential group of stem cells that make up the epiblast in the preimplantation embryo (Nichols *et al.,* 1998), and more recently it has been shown to play a key regulatory role in determining ES cell fate according to its precise level of expression (Niwa *et al.,* 2000). It is hypothesized that the activation of *Stat-3* by LIF (Niwa *et al.,* 1998; Burdon *et al.,* 1999a), or by an alternative self-renewal signal ESRF (Dani *et al.,* 1998), is necessary to stimulate the expression of a specific *Oct-4* partner such as *Rox-1* for maintenance of the pluripotent ES cell phenotype (Niwa *et al.,* 2000). Presently, a greater understanding of the fundamental molecular mechanisms controlling the self-renewal and differentiation of both ES and multipotent mouse stem cells is needed to facilitate establishing robust stem cell culture systems for a range of mammalian species and the widespread application of such systems to stem cell-based research and tissue therapies.

En-2 | **IRES** | **nβgeo** | **pA**

(1) ES cells electroporated with gene trap vector and cultured in media containing G418

neomycin resistant clones selected for analysis

(2) Three-tiered *in vitro* screen to select ES cell gene trap clones harboring integrations in ES cell-specific genes

(3) 5'-RACE and sequence analysis of mRNA fusion transcripts to identify endogenous gene sequences

(4) Chimaera generation for selected ES cell-specific gene trap clones

(5) Analysis of gene trap reporter gene expression during development

Figure 3.1: Experimental design for the identification and analysis of genes expressed in ES cells in vitro and downregulated upon differentiation. (1) ES cells are electroporated with the pGT1.8 (OPT/Nuclear) IRESβgeo gene trap vector and cultured in selection media. (2) G418 resistant colonies are cultured through three levels of *in vitro* screening assays to identify those clones harboring reporter integrations in genes which are specifically expressed in the ES cells, and not in their differentiating progeny. (3) For selected clones harboring candidate ES cell-specific genes, the endogenous gene is identified by sequence analysis of 5'RACE products generated from mRNA fusion transcripts. (4) ES cell lines harboring gene trap insertions in novel genes are microinjected into host blastocysts for chimaera generation and transmission through the germline. (5) The *in vivo* expression profile for the novel gene is analyzed by following nuclear expression of the reporter gene in transgenic animals during development.

The aim of this study is to design and implement a screening strategy to identify genes which are expressed in ES cells, but not in their differentiated progeny, and may therefore, play a key role in regulating the growth and/or differentiation of ES cells. This strategy, outlined in Figure 3.1, combines random ES cell gene trap cloning with a stringent three-tiered *in vitro* pre-screen to identify clones expressing the reporter gene in the appropriate stem cell-restricted fashion. Clones harboring candidate ES cell-specific gene trap integrations are then selected for 5'-RACE and direct sequencing analyses of cDNAs for mRNA fusion transcripts (see Chapter four), germline transmission (see Chapter five) and *in vivo* phenotypic analyses (see Chapter six).

This chapter describes the use of two new gene trap vectors which provide IRES-mediated translation of either cytoplasmic or nuclear localized *βgeo* reporter gene products when inserted into the intronic regions of transcriptionally active genes. These promoterless constructs are expected to insert randomly across the mouse genome into genes representative of all major classes, including those with even very low levels of expression. The incorporation of an IRES element abates the need for insertion of the reporter in alignment with the preceding endogenous reading frame, and as such was anticipated to increase the number of detectable gene trap events (Chowdhury *et al.*, 1997). In order to select for gene trap events of interest, the generated clones were subject to three levels of *in vitro* pre-screening prior to selection for further analysis. The design and optimization of this screening strategy as well as the results for a large-scale gene trap experiment, are also described in this chapter.

3.2   DESIGN OF AN OPTIMIZED IRES*βgeo* GENE TRAP VECTOR

Two gene trap vectors employing an optimized (OPT) IRES*βgeo* cassette were used with each providing IRES-mediated translation of a *lacZ-neo* fusion reporter gene (via a translation initiation codon in an optimized Kozak sequence, P. Mountford, unpublished) following transfection into ES cells.

Based on the previously described vector pGT1.8*geo* (Skarnes *et al.*, 1995), these 9.4 kb constructs (Figure 3.2) contain a 5'-end splice acceptor sequence and preceding intron buffer sequence from the mouse *En-2* gene (Gossler *et al.*, 1989) joined in frame with the promoterless *βgeo* reporter-selectable marker gene (Friedrich and Soriano, 1991). At the 3'-end this sequence is followed by the simian virus 40 (SV40) poly A signal sequence (Joyner *et al.*, 1989). The *βgeo* gene fusion of the parental vector lacks an
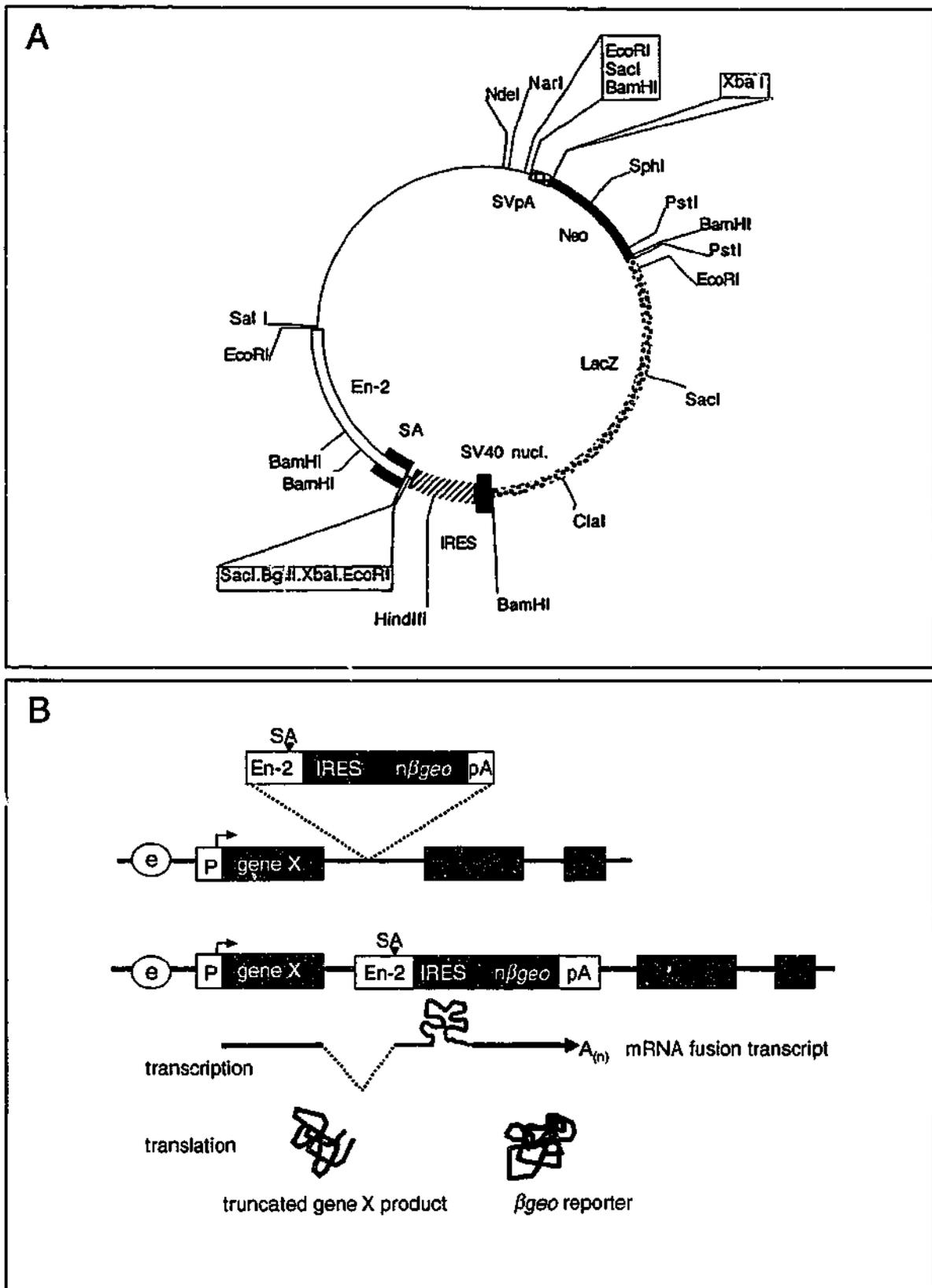
Figure 3.2: Design and mechanism of reporter gene activation in the pGT1.8 (OPT/Nuclear) IRESβgeo gene trap vector. (A) The 9.4 kb construct contains 5' intron buffer sequence upstream of a splice acceptor site (SA) in coding sequence from the mouse *En-2* gene (En-2). An optimized EMCV IRES sequence is ligated between En-2 sequence and the downstream *βgeo* reporter gene (LacZ + Neo). Nuclear localization of the *βgeo* reporter is mediated via fusion of the nuclear localization signal from the SV40 large T gene (SV40 nucl.) to the 5' LacZ terminus. Polyadenylation of gene trap fusion transcripts is provided by a downstream SV40 poly A signal (SVpA). (B) Where the construct integrates into an intronic region of an active transcriptional unit, the IRES sequence mediates cap-independent translation of a nuclear localized (n) *βgeo* reporter gene. The pGT1.8 (OPT) IRESβeo gene trap vector was identical to the above construct, except for the exclusion of the nuclear localization signal.

ATG codon and does not contain the *neo* point mutation known to reduce its enzymatic activity (Yenofsky *et al.*, 1990; Skarnes *et al.*, 1995).

The bicistronic IRES*βgeo* vectors were obtained by introducing the IRES element from the 5'-untranslated region of the EMCV mRNA (Jang *et al.*, 1988; Pelletier and Sonenberg, 1988) between the *En-2* splice acceptor and *βgeo* reporter gene sequences. Previous IRES*βgeo* constructs (Mountford *et al.*, 1994; Chowdhury *et al*, 1997; Bonaldo P. *et al.*, 1998) have made use of an IRES-*lacZ* fusion sequence in which the 594 bp IRES sequence was modified by mutagenesis of the native 11th translation initiation codon to create a convenient cloning site and force initiation of translation to occur 9 bp downstream at the 12th AUG codon (Ghattas *et al.*, 1991). In the pGT1.8 (OPT) IRES*βgeo* construct, initiation of translation was relocated to the native AUG codon, and the surrounding IRES nucleotide sequence modified to reflect an optimized Kozak consensus sequence (Kozak, 1989; P. Mountford, unpublished). The constructed OPT IRES-*lacZ* fusion plasmid was then cloned into the pGT1.8*geo* parental vector (P. Mountford, unpublished).

A second construct, pGT1.8 (OPT/Nuclear) IRES*βgeo*, was modified from the above construct by the insertion of an SV40 nuclear localization signal (Figure 3.2A) immediately downstream of the optimized AUG codon in the IRES element (P. Mountford, unpublished). This construct was designed to provide improved visualization of the reporter gene activity in cultured cells and in animal tissues.

These previously untested gene trap constructs were designed to enhance the frequency of detectable events by providing optimized internal initiation of translation of the reporter gene (Figure 3.2B). Any insertion into an active gene in the correct orientation should confer resistance to the selective agent G418 and result in the expression of a functional *lacZ* gene product by eliminating amino terminal fusion events that result in the loss of βgal, but not neo, activity.

3.3 COMPARISON OF pGT1.8 (OPT) & (OPT/Nuclear) IRES*βgeo* GENE TRAP VECTORS

To compare the efficiency of the cytoplasmic and nuclear localized IRES*βgeo* gene trap vectors, plasmid DNA was linearized with *Sal* I and introduced by electroporation into E14 ES cells (section 2.3.4). After 8-10 days of culture in medium supplemented with LIF and containing G418 at 200 μg/ml, G418 resistant colonies were counted and

primary colonies picked to assess those also staining for βgal activity (sections 2.3.4 &
2.6.2). Three independent electroporation experiments were performed in parallel using
50 μg of DNA for each vector, and equally dividing the ES cell harvest on each
occasion to give 40-50 x 10^6 cells for each electroporation. In the third experiment ~36
μg of pGT1.8 (OPT/Nuclear) IRES*βgeo* DNA was used, but this did not result in a
lesser yield of G418 resistant colonies compared with 50 μg for the pGT1.8 (OPT)
IRES*βgeo* vector. Results summarized in Table 3.1 show the overall number of G418
resistant colonies obtained for each vector, with an efficiency of 1:240,000 (4.1 x 10^{-6})
for the pGT1.8 (OPT) IRES*βgeo* vector, and 1:276,000 (3.6 x 10^{-6}) for the pGT1.8
(OPT/Nuclear) IRES*βgeo* vector. An equivalent number of primary colonies [77.9%
for the pGT1.8 (OPT) IRES*βgeo* vector and 71.2% for the pGT1.8 (OPT/Nuclear)
IRES*βgeo* vector] were positive for βgal activity following X-gal staining. The *lacZ*
expression patterns for both vectors varied widely from weak to marked expression in
some or all cells of a colony.

	pGT1.8 (OPT) IRES*βgeo*	pGT1.8 (OPT/Nuclear) IRES*βgeo*
# ES cells electroporated	137 x 10^6	137 x 10^6
# G418 resistant colonies	571	497
# stable gene trap integrations	1:240,000	1:276,000
# primary colonies picked	77	73
% *lacZ* positive colonies	77.9%	71.2%

*Table 3.1: Comparison of gene trapping efficiency and reporter gene expression profiles for two
(OPT) IRESβgeo gene trap vectors. Results are pooled for three electroporation experiments
performed in parallel for the pGT1.8 (OPT) IRESβgeo and pGT1.8 (OPT/Nuclear) IRESβgeo
vectors.*

On the basis of these preliminary results, all subsequent electroporation experiments
were performed using the pGT1.8 (OPT/Nuclear) IRES*βgeo* vector to provide nuclear
visualization of the reporter gene activity in future expression analyses for gene trap
events of interest.

3.4 ESTABLISHMENT OF A THREE-TIERED SCREENING SYSTEM TO SELECT FOR GENE TRAP INTEGRATIONS IN CANDIDATE ES CELL-SPECIFIC GENES

As gene trapping relies on random integration of the reporter-selectable marker gene fusion within transcriptionally active genes, it was important to design and implement a large-scale but stringent screening system to eliminate those clones which are not of interest. This section describes the development of a three-tiered screening system to select for those ES cell clones harboring gene trap integrations in genes that are specifically expressed in ES cells and downregulated upon differentiation.

3.4.1 Control ES cell lines to validate and optimize a screening system

A series of previously characterized ES cell lines harboring either targeted or random gene trap *lacZ-neo* insertions (section 2.1.5) were employed as positive and negative control cell lines to evaluate the proposed screening strategy for ES cell-specific gene trap insertions.

ES cell lines trapped with the secretory gene trap (see section 1.6.4), and harboring βgeo insertions within genes encoding N-terminal signal sequences, were used as negative control lines. These cell lines, ST514, ST519, ST497 and ST534, all stain positive for βgal activity in subsets of spontaneously differentiated ES cells (Skarnes *et al.*, 1995). In addition, an ES cell line with an IRESβgeopA cassette targeted at the 3'-untranslated region of the LIF gene, DIAβgeo pA63, is positive for βgal activity in differentiated cells but not in ES cells (Mountford *et al.*, 1994). Similarly, the cell line DIA RK022 has a targeted IRESβgeopA insertion in the ligand binding region of the LIF-R gene, and is positive for βgal activity in both ES and differentiated cells (Li *et al.*, 1995). Two other cell lines ZIN40 and GTIRES 177, generated from previous gene trap studies with a *lacZ*-IRES-*neo* vector, also provided negative controls. These both show ubiquitous βgal expression in differentiating ES cell cultures, and in adult mouse tissues for ZIN40 (Mountford and Smith, 1995).

To date, the *Oct-4* gene provides the best example for "ES cell-specific" gene expression *in vitro*. The OKO160 ES cell line - generated by a targeted replacement of the POU-specific domain and the homeodomain coding sequences in exons 2-5 of *Oct-4* with an IRESβgeopA cassette (Mountford *et al.*, 1994) - was used as the positive control in all experiments.

3.4.2 Optimization of culture and selection conditions

ES cell culture medium was initially optimized for the routine culture of ES cell lines, including batch titrations of LIF and FCS, and the addition of MEM non-essential amino acids (data not shown), (sections 2.1.5 & 2.3.1). The above control cell lines, eight negative and one positive, were then used to establish the three screening assays for ES cell-specific gene trap insertions, and to optimize the selection and culture conditions for each.

(a) Optimizing ES cell monolayer cultures

To optimize the selection of G418 resistant ES cells from transfection cultures, the OKO160 and ZIN40 control cell lines were cultured as described for electroporation experiments (section 2.3.4b), with the addition of G418 at concentrations of 0, 200, 400 and 1600 µg/ml, and examined morphologically following Leishman's staining (section 2.6.1), (data not shown). To further mimic a gene trap transfection experiment, the control cell lines were each mixed with a background of non-resistant E14Tg2a ES cells (section 2.1.5). A G418 concentration of 200 µg/ml was chosen as most suitable for the efficient elimination of non-transfected cells whilst allowing for the detection of integrations in genes with only low-level expression in ES cells.

Conditions were also optimized for the selection of neo resistant ES cells in expanded clonal cultures. All control ES cell lines (section 3.4.1) were grown as routine monolayer cultures from a seeding density of ~50,000 cells/cm^2 (section 2.3.1), with the addition of G418 at concentrations of 0, 200, 300, 400 and 600 µg/ml, and examined morphologically following Leishman's staining (section 2.6.1). Results summarized in Table 3.2 demonstrate the survival of differentiating cells for all negative control cell lines in the presence of G418, except for ST519, which did not readily differentiate in culture and showed a marked reduction in ES cell survival with the minimal selection level. In contrast, the positive control OKO160 was the only cell line to show survival of pure ES cell populations under selection, with differentiating cells dying at a G418 concentration at or above 200 µg/ml. Most control cell lines showed strong survival of neo resistant cells at the highest level of selection, although many showed a reduction in overall culture growth between 400 and 600 µg/ml G418. A G418 concentration of 300 µg/ml was chosen as most suitable for the efficient elimination of non-resistant cell

types from expanding transgenic ES cell cultures, while allowing detection of genes with low levels of expression in ES cells and sustaining cell growth for resistant cells.

G418 (μg/ml)	0	200	300	400	600
OKO160	ES +++ diff +++ grth ++++	ES +++ diff + grth ++	ES +++ diff - grth ++	ES +++ diff - grth ++	ES +++ diff - grth +
ZIN40	ES +++ diff +++ grth ++++	ES +++ diff +++ grth +++	ES +++ diff +++ grth +++	ES +++ diff +++ grth +++	ES +++ diff +++ grth +++
GTIRES 177	ES +++ diff +++ grth ++++	ES ++ diff +++ grth +++	ES ++ diff +++ grth +++	ES ++/+ diff +++ grth +++	ES + diff +++ grth ++
DIAβgeo pA63	ES +++ diff +++ grth +++	ES ++ diff +++ grth ++	ES ++/+ diff +++ grth ++	ES + diff +++ grth ++	ES - diff +++ grth +/-
DIA RK022	ES +++ diff +++ grth +++	ES ++ diff +++ grth ++	ES +/- diff +++ grth ++/+	ES - diff +++ grth +	ES - diff +++ grth +/-
ST514	ES +++ diff +++ grth ++++	ES +++ diff ++ grth +++	ES +++ diff + grth +++	ES +++ diff +/- grth ++	ES + diff - grth +/-
ST519	ES +++ diff - grth +++	ES + diff - grth +/-	ES - diff - grth -	ES - diff - grth -	ES - diff - grth -
ST497	ES +++ diff +++ grth ++++	ES + diff +++ grth +++	ES +/- diff +++ grth ++	ES +/- diff +++ grth ++	ES - diff +++ grth +
ST534	ES +++ diff +++ grth +++	ES + diff +++ grth ++	ES +/- diff +++ grth ++	ES +/- diff +++ grth ++/+	ES - diff +++ grth +

Table 3.2: G418 titration of control ES cell lines grown in monolayer cultures. Cells were plated at a density of 50,000 cells/cm^2 and cultured for 3-4 days in medium supplemented with LIF and 0-600 μg/ml G418. Cell morphology and growth was assessed and scored following Leishman's staining on a scale of - (none), +/- (weak), + (mild), ++ (moderate), +++ (marked) or ++++ (very marked). ES = ES cells; diff = differentiated cells; grth = culture growth. The positive control OKO160 is the only cell line that shows ES cell-specific neomycin resistance with efficient elimination of differentiating cells at a G418 level >200 μg/ml.

110

Control ES cell lines were also cultured routinely at a clonal density of ~30 cells/cm^2 in non-selective medium (section 2.3.1) and assessed for βgal activity following X-gal staining (section 2.6.2). *lacZ* expression patterns corresponded with the neo resistance profiles for each cell line (Table 3.2) and confirmed that the positive control OKO160 is the only cell line which demonstrates ES cell-restricted *lacZ* expression (data not shown). All negative control cell lines, except ST519, showed various levels and patterns of X-gal staining in both ES and differentiating cells. ST519 ES cells, which had failed to propagate well in the earlier monolayer selection assay (Table 3.2), showed few differentiated cells in this assay and did not stain for βgal activity. In the original study that generated this line, βgal activity was observed only in a subset of spontaneously differentiated cells and not in ES cells (Skarnes *et al.*, 1995), suggesting that such a subset has not arisen in this study. The DIA*βgeo* pA63 line was positive for βgal activity in differentiating cells but not in ES cells, as previously reported (Mountford *et al.*, 1994).

These experiments confirmed the ability to detect a cell line positive for ES cell-restricted *neo* and *lacZ* expression in monoclonal cultures, and demonstrated the variation in *lacZ* patterns seen for cell lines carrying randomly introduced or targeted *βgeo* insertions.

(b)   Optimizing ES cell suspension cultures

Aggregation induces ES cells to develop into highly differentiated embryoid body structures (Martin and Evans, 1975; Doetschman *et al.*, 1985). The accompanying elimination of pluripotential ES cells is due to the efficient induction of differentiation, and probably also due to programmed cell death (Coucouvanis and Martin, 1995; 1999). It has been previously demonstrated that continuous removal of differentiating cells from EBs by selection for ES cell-specific neo resistance, in the absence of LIF, results in the persistence and expansion of ES cells (Mountford *et al.*, 1998). This finding was evaluated as the basis for a stringent 3-dimensional screening assay to confirm ES cell-specific gene trap events.

To better control the size and, potentially, onset and progression of ES cell differentiation in aggregates, a series of experiments were initially undertaken to establish aggregates of predetermined size and optimized culture conditions (data not shown). Observations suggested that aggregates were best established with 200-300

cells in 20 μl hanging drops, and that the growth of EBs was more efficient after 7 days if aggregates were transferred from hanging drop to suspension culture after 2-3 days (section 2.3.5a). In order to prevent adhesion of differentiating EBs to the surface of cell culture vessels, suspension culture wells were coated with an agarose-media mix (section 2.3.5a). The established method used for all hanging drop/EB assays in this study is given in section 2.3.5a.

For all control ES cell lines (section 3.4.1) aggregates were grown in hanging drop cultures without LIF, and with the addition of G418 at 300 μg/ml on the 4th day when aggregates were transferred to suspension culture (section 2.3.5a). EBs were dissociated after a total of 10 days in suspension culture, and cells cultured for a further 2 days in media containing LIF but not G418. Cell morphology was assessed following Leishman's staining (section 2.6.1). Results, summarized in Table 3.3A, demonstrate that only the OKO160 positive control ES cell line showed recovery of an essentially pure ES cell population. Negative control ES cell lines showed the survival of substantial numbers of differentiated cell types and the effective elimination of ES cells from EBs. In some cases, scant ES cell clusters could be recovered from the centers of large negative control EBs formed by the joining up of a few aggregates in suspension culture wells. The ST514 and ST519 negative control lines, however, gave results that were unexpected. An ES cell-like culture with minimal differentiated cells was recovered from ST514 EBs, while the ST519 EBs did not yield a recoverable culture of any cell type. In both cases it would seem that ES cells have not efficiently undergone differentiation in the absence of LIF, possibly due to the selection of a subset of non-differentiating aggregates in this experiment. A clonal culture of the ST519 line in the presence and absence of LIF (without G418), did indeed confirm the inability for this ES cell line to differentiate or even survive in the absence of LIF (data not shown). Identification of a non-differentiating ES cell clone highlighted the need to confirm differentiation capacity by culturing selected clones in the absence of G418 selection.

In a second experiment, EB cultures were established from the OKO160 and ZIN40 cell lines with exposure to G418 at 0, 300, 400 and 600 μg/ml (section 2.3.5a). Selection was applied in the hanging drop cultures, as well as in suspension cultures, to optimize the efficiency of eliminating non-resistant ES or differentiating cells from EBs. Results, shown in Table 3.3B, confirmed the extinction of ES cells from fully differentiated cultures for both cell lines when cultured without G418, and the recovery of pure ES

cell cultures from OKO160 EBs cultured with 300 µg/ml G418. These results were also sustained at the highest selection level.

A ●————————————● B ●—————————————————————————————●

G418 (µg/ml) (no LIF)	days 4-10 300	days 0-10 0	300	400	600
OKO160	ES +++ diff +/-	ES - diff +++	ES +++ diff -	ES +++ diff -	ES +++ diff -
ZIN40	ES +/- diff +++	ES +/- diff +++	ES +/- diff +++	ES +/- diff +++	ES +/- diff +++
GTIRES 177	ES - diff +++				
DIAβgeo pA63	ES - diff +++				
DIA RK022	ES - diff +++				
ST514	ES +++ diff +				
ST519	ES - diff -				
ST497	ES +/- diff +++				
ST534	ES +/- diff +++				

Table 3.3: Selection for neomycin resistance in control ES cell aggregates. Hanging drop cultures were established with 300 cells/20 µl and aggregates cultured in the absence of LIF for 10 days. The proportion of ES cells (ES) or differentiated cells (diff) recovered from dissociated EBs after a further 2 days culture with LIF was visually assessed following Leishman's staining and scored on a scale of - (none), +/- (few scant), + (mild), ++ (moderate), +++ (marked). G418 selection was applied at (A) 300 µg/ml from day 4-10 for all control cell lines and (B) 0-600 µg/ml from day 0-10 for the OKO160 positive control and ZIN40 negative control cell lines. Aggregation culture without G418 selection confirms the ability of both cell lines to differentiate normally in the absence of LIF, while the OKO160 cell line displays selection for ES cell-specific neomycin resistance with exposure to G418.

Having optimized culture and selection conditions the following primary, secondary and tertiary *in vitro* screening protocols were established (sections 3.4.3-3.4.5) for use in a large-scale gene trapping experiment.

3.4.3 Primary screen for ES cell clones harboring gene trap integrations in potential ES cell-specific genes

Following electroporation of ES cells with the pGT1.8 (OPT/Nuclear) IRESβgeo vector, cells were cultured in medium supplemented with LIF and containing G418 at 200 µg/ml, for up to 10 days (sections 2.3.4a & b). G418 resistant colonies were

assessed by phase-contrast microscopy to select those ES cell colonies lacking peripheral differentiated cells (Figure 3.3A). All "ES only" colonies were selected based on this morphological criteria and expanded in monolayer culture (section 2.3.4c) for the next level of screening.

3.4.4 Secondary screen to confirm ES cell-restricted *βgeo* expression

Colonies selected from the primary screen were split into two adjacent 2.0 cm^2 culture wells for the secondary screen, and into a third 3.8cm^2 well for maintenance of the line (section 2.3.4c). Cells in the latter well were expanded in culture medium supplemented with LIF and containing G418 at 300 μg/ml, and either maintained for the tertiary screen or frozen (section 2.3.2) until results of the secondary screen were available.

For the secondary screen, the selected colonies were expanded in the presence of LIF, with and without the addition of G418 at 300 μg/ml. Cultures were grown for 4-6 days until confluency was approached, with media changes every 2nd day (section 2.3.1). Cultures were fixed and stained with X-gal (section 2.6.2) and assessed by both phase-contrast and light-field microscopy to confirm that the selected clones could differentiate normally in the absence of G418 and that the *βgeo* expression was indeed restricted to the ES cell population (Figure 3.3B).

3.4.5 Tertiary screen for ES cell recovery from embryoid bodies

Those clones confirmed as positive for ES cell-specific *βgeo* expression in the secondary screen were then screened for ES cell-specific *neo* expression in a 3-dimensional suspension culture assay.

This tertiary screen cultured the selected gene trap ES cell lines as aggregates in hanging drop culture for 3 days in the absence of LIF. For each line, 40-50 aggregates were established in parallel in the absence and presence of G418 at 300 μg/ml (section 2.3.5a). The resulting EB structures were pooled and transferred for further culture in 3.8 cm^2 tissue culture wells coated with agarose. Again, culture was in the absence of LIF, with or without G418 at 300 μg/ml. After a total of 10 days in suspension culture, the enlarged EBs were trypsinized and replated into 2.0 cm^2 gelatinized tissue culture wells and cultured in the presence of LIF for a further 2 days (section 2.3.5a). Adherent cell morphology was assessed by phase-contrast microscopy for recovery of undifferentiated ES cells (Figure 3.3C).

114

A

+LIF
+G418

pGT1.8 (OPT/Nuclear) IRES*βgeo*

positive control

negative control

discard

B

+LIF
+G418

+LIF
-G418

X-gal staining

negative control ➤

discard

positive control ➤
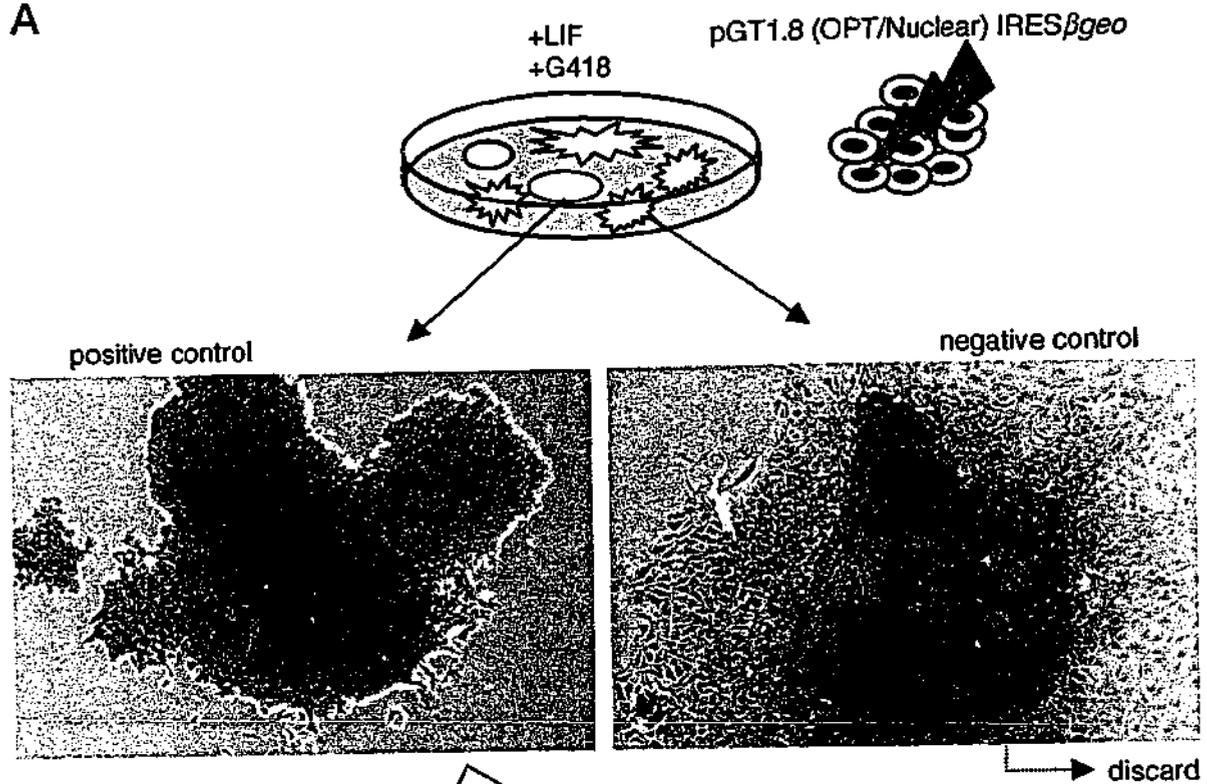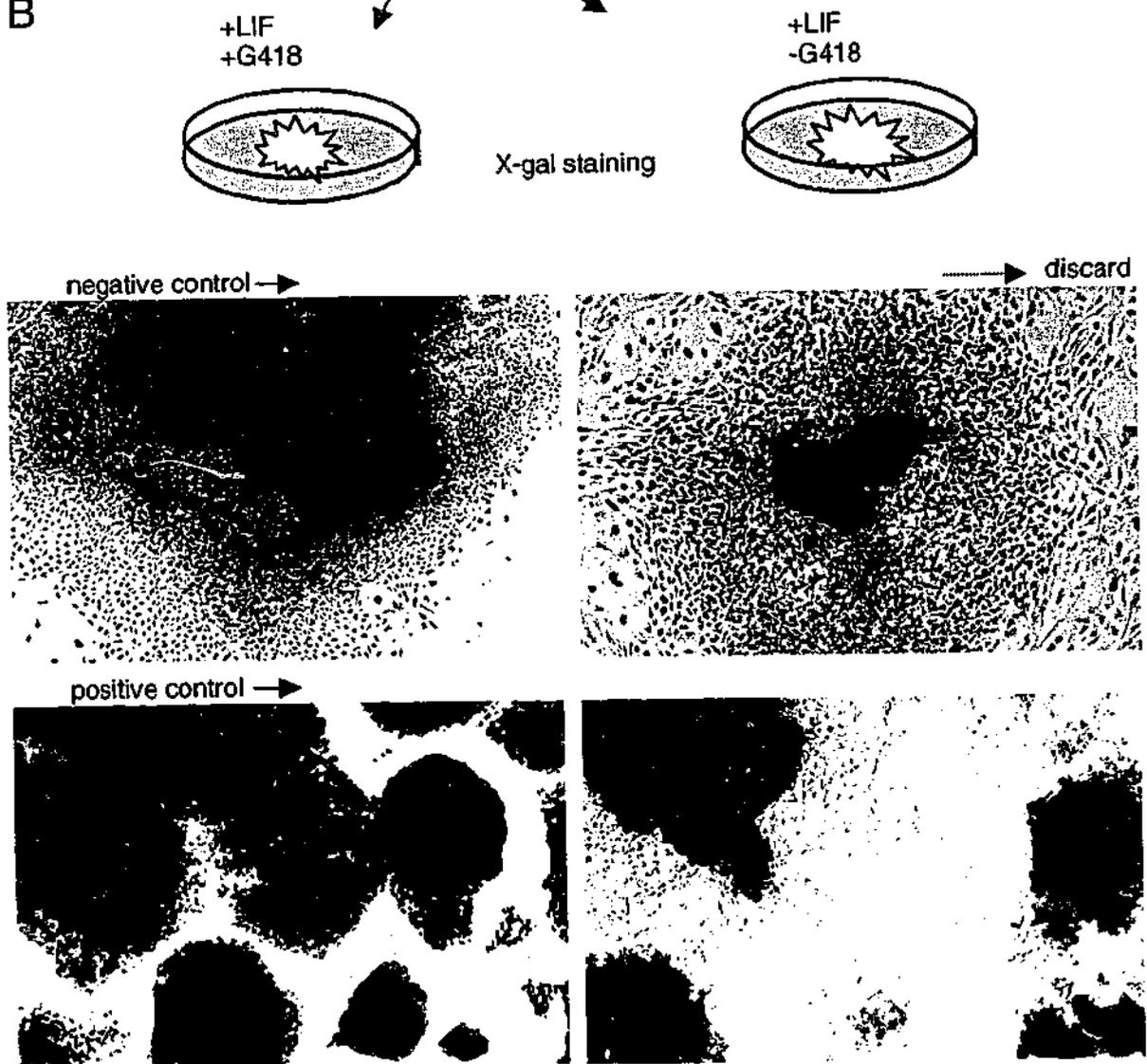
C
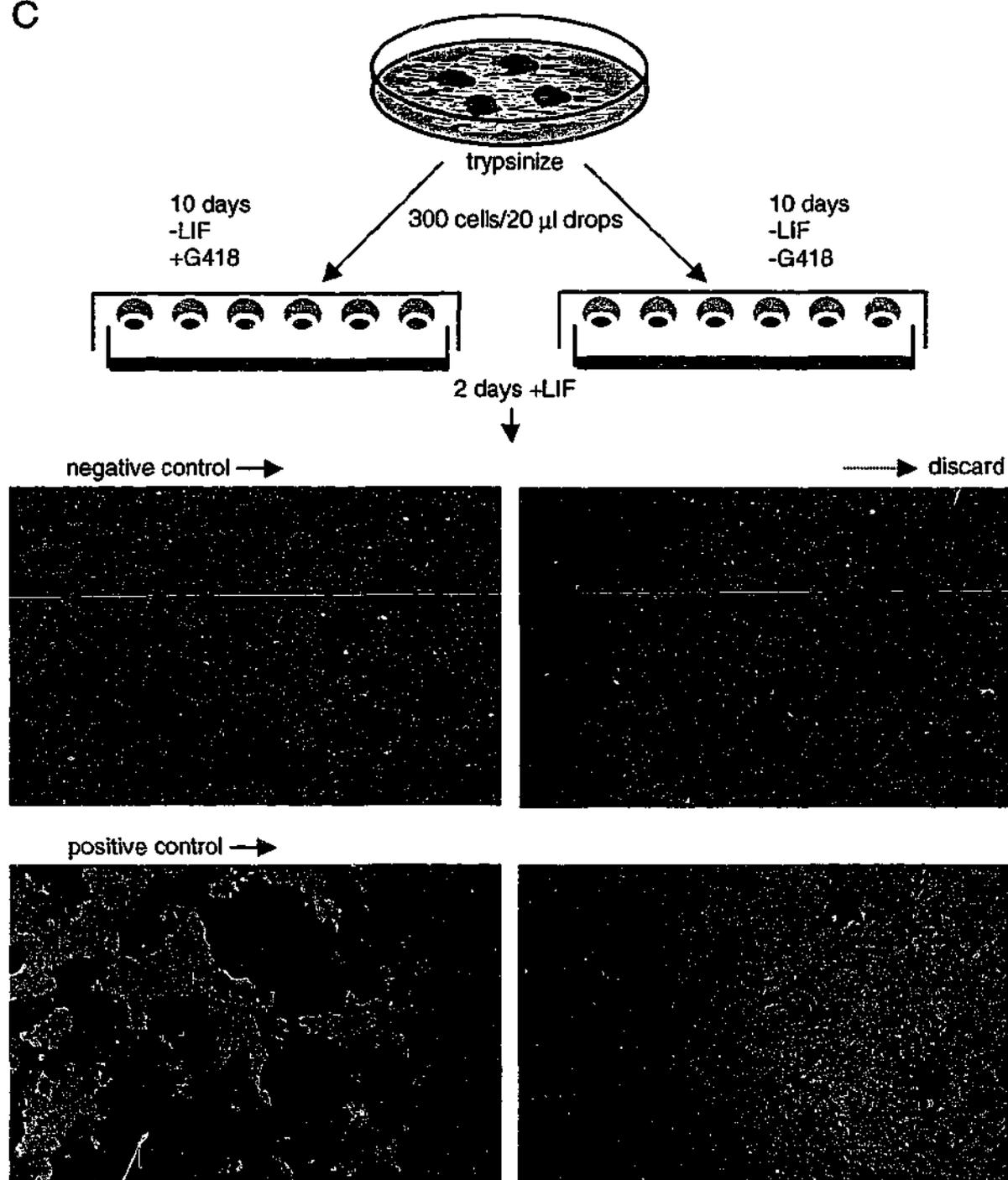
Figure 3.3: A three-tiered in vitro screening strategy to identify gene trap integrations in ES cell-specific genes. (A) Following electroporation with the gene trap vector, ES cells were cultured for up to 10 days in the presence of LIF and G418 at 200 µg/ml. In a primary screen for ES cell-specific clones, those colonies lacking peripheral differentiated cells were selected. Positive and negative control colonies are shown following Leishman's staining. (B) In a secondary screen, selected primary colonies were expanded in monolayer culture in the presence of LIF, with and without G418 at 300 µg/ml, and stained with X-gal after 4-6 days culture. This screen confirmed normal expansion and differentiation for gene trap ES cell clones and restriction of βgeo expression to the undifferentiated ES cell population in the positive control cell line. (C) In a tertiary screen, ES cells were cultured in the absence of LIF as aggregates in hanging drops, with or without G418 at 300 µg/ml. The resulting embryoid bodies were dissociated by trypsinization after 10 days and cultured for a further two days in monolayer culture in the presence of LIF, then stained with Leishman's. The ability to recover a high proportion of cells with undifferentiated ES cell morphology under these conditions confirmed selection for ES cell-specific neo expression. For each screen, the positive control ES cell line shown is OKO160 and the negative cell line is ZIN40. Photographed under phase-contrast optics, 100x.

E14	passage 28	pGT1.8 (OPT/Nuclear) IRESβgeo gene trap clones							
Electroporation date		14.4.97	G418 selection: 200ug/ml in GMEM-ES		LIF: 1000U/ml				
Transfection expt #		7	Total # cols		296	Primary colonies picked			56 D8 22.4.97
secondary assay: +/- G418 300ug/ml, LIF 1000U/ml				X-gal D6					tertiary assay
G418 300ug/ml selection			without selection				GT#	fz date	ES recovery
col	ES	Diff	X-gal blue	ES	Diff	X-gal blue			no LIF,+/-G300
41	xxx	xxx	xx both	xxx	xxx	xx both			
42	xxx	xxx	xx both, P++	xxx	xxx	xx both, P++			
43	xxx	xx	xx both	xx	xxx	xx both, (*ES)			
44	xxx	xxx	xx both	xxx	xxx	xx both			
							21		POSITIVE
46	xxx	xxx	xx both	xxx	xxx	xx both			
47	x	? (only few small cols) no blue		xx	xxx	no blue			
49	xxx	xxx	x/xx both	xxx	xxx	x/xx both			
50	no grth			no grth					
52	xxx	xxx	xx (*ES), P+++	xxx	xxx	xx (*ES), P+++			
53	xxx	xxx	xx both	xxx	xxx	xx/xxx both			
54	xxx	xxx	x both	xxx	xxx	x both			
55	xxx	xxx	xxx both	xxx	xxx	xxx both			
	xxx	xxx	x/xx both, P+++	xxx	xxx	x/xx both, P+++			
	xxx	xxx	xx both	xxx	xxx	xx both, P++			
58	xxx	xx	xx/xxx both	xxx	xxx	xx/xxx both, P++			
59	xxx	few	xxx ES (white diffs)	xxx	xxx	xxx ES only			OKO160 CONTROL
60	xxx	xxx	xxx both	xxx	xxx	xxx all cells			ZIN40 CONTROL
	60 ug DNA plus 49 x 10⁶ ES cells for TF(7)								
	"New" pGT1.8 (OPT/Nuclear) IRESβgeo maxiprep DNA								
	Scored for presence of ES cells (ES) or differentiated (diff) cells and for βgal activity								
	Scale: x (mild), xx (moderate), xxx (marked)								
	P= patchy/heterogeneous lacZ; grth = growth of culture								
	positive secondary; tertiary candidate								
	positive tertiary assay - ES cell-specific gene trap clone								
	non ES cell-specific primary colonies								

60 ug DNA plus 49×10^6 ES cells for TF(7)

Table 3.4: Result sheet for a typical gene trap transfection and screening experiment to identify ES cell-specific integrations. Following electroporation of ES cells with the gene trap vector, primary colonies were screened for ES cell-restricted neomycin resistance by morphological assessment. In a secondary screen, selected primary colonies were expanded in monolayer culture in the presence of LIF, with and without G418, and stained with X-gal to confirm restriction of βgeo expression to the ES cell population. In a tertiary screen, clones were cultured as aggregates in the absence of LIF, with and without G418, and assessed for the selectable recovery of ES cells from the resulting embryoid bodies.

3.5 A LARGE-SCALE SCREEN FOR GENE TRAP INTEGRATIONS IN ES CELL-SPECIFIC GENES

A large-scale study was performed to identify ES cell clones harboring gene trap insertions in candidate ES cell-specific genes. Including the initial three experiments in which the pGT1.8 (OPT) IRESβgeo and the pGT1.8 (OPT/Nuclear) IRESβgeo gene trap vectors were compared (section 3.3, Table 3.1), a total of 21 independent transfection experiments were undertaken. Both E14 and E14Tg2a ES cells (section 2.1.5) were transfected and the OKO160 and ZIN40 control ES cell lines (section 3.4.1) were routinely used as positive and negative controls, respectively, in the secondary and tertiary screens. Gene trap clones deemed positive for ES cell-specific βgeo expression in the secondary screen were kept as frozen stocks and referred to by a gene trap number, GT#. An example of a results sheet for a typical transfection and screening experiment is given in Table 3.4. It should be noted that from the eighth experiment onwards, to save resources, those clones that were clearly not expressing *neo* in an ES cell-restricted manner in the expanded secondary screening cultures, were not stained with X-gal. All selected clones from the secondary assays were screened in at least two independent EB assays, with and without selection, to ensure that a subset of non-differentiating aggregates was not being selected for.

Overall screening results for the 21 transfections are given in Table 3.5. The combined selection efficiency following electroporation of a total 1.03 x 10^9 ES cells was ~1:236,000 (4.2 x 10^{-6}), a yield of 4345 G418 resistant colonies. The primary screen yielded 18% (775) morphologically ES cell-restricted colonies. The secondary screen eliminated the majority of these (93%) as clones expressing βgeo in both ES and differentiating cell types, yielding 54 clones with ES cell-restricted βgeo expression profiles in monolayer culture. The tertiary screen in suspension culture identified 28 of these as ES cell clones harboring gene trap integrations in candidate ES cell-specific genes. This represents less than 4% of the number of colonies selected in the primary screen, and less than 1% of the total number of stably integrated gene trap events.

For the total number of G418 resistant clones that were stained with X-gal in the secondary screen, 85% (387/455) demonstrated some degree of *lacZ* expression regardless of cell type. The variation in expression patterns seen for βgal positive clones is demonstrated in Figure 3.4.
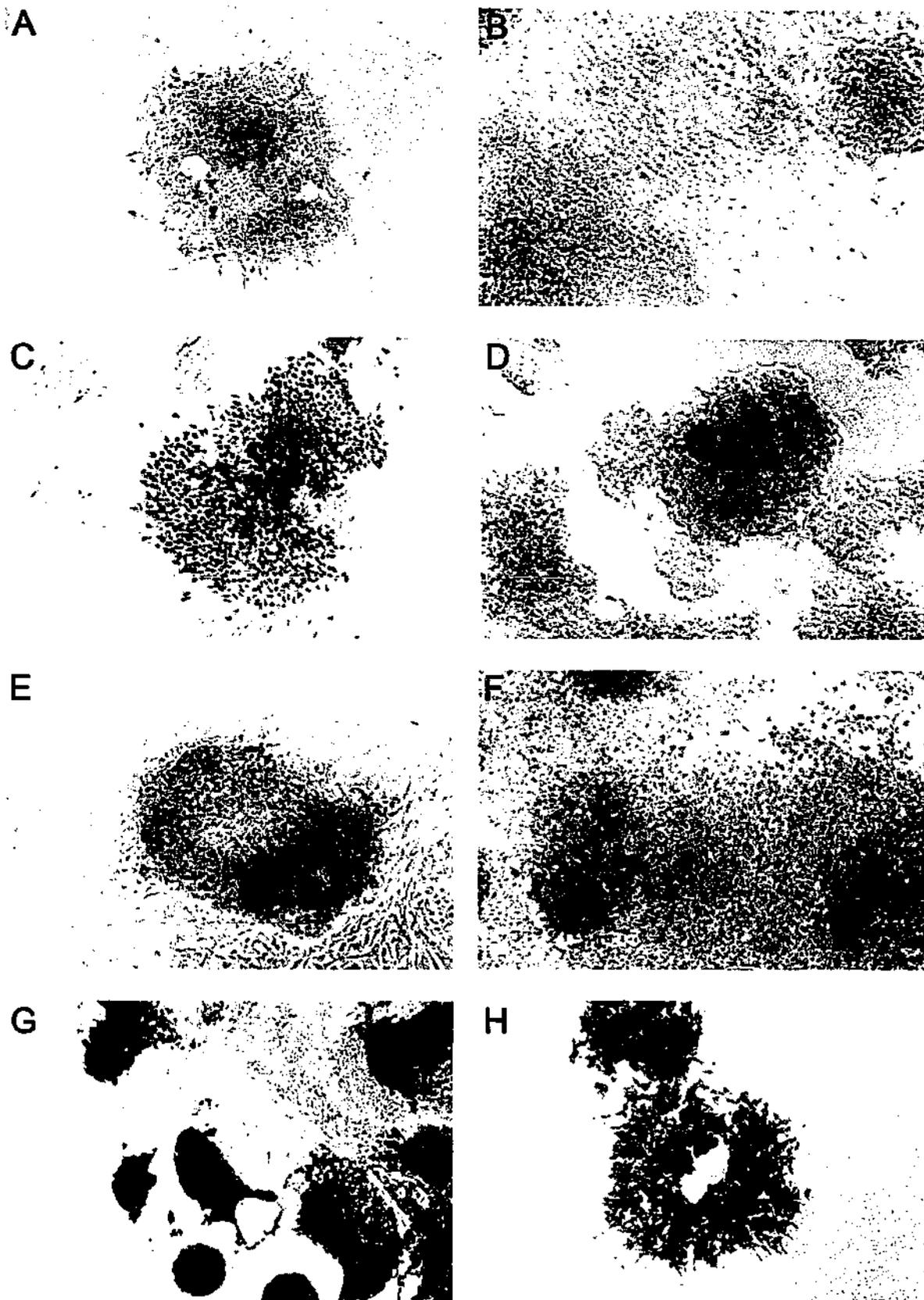
Figure 3.4: Variable βgal expression profiles displayed in ES cell gene trap clones. Following X-gal staining, ES cell gene trap colonies cultured for ~5 days with G418 (A, C, D, H), or without G418 (B, E, F, G), displayed βgal expression levels varying from mild (A, B) to moderate (C, D, E, F), and marked (G, H). In many cases *lacZ* expression was heterogeneous throughout colonies, and variation was seen in growth patterns. For the examples shown, all clones except one (F) were positive following the secondary screen for ES cell-restricted *βgeo* expression. Clones were generated with the pGT1.8 (OPT/Nuclear) IRES*βgeo* (A-G) and the pGT1.8 (OPT) IRES*βgeo* (H) gene trap vectors. Photographed under phase-contrast optics, 100x.

Total #cells electroporated	#G418 resistant colonies	#Primary screen positive	#Secondary screen positive	#Tertiary screen positive
1.03×10^9	4345	775	54	28

Table 3.5: Enrichment for ES cell-specific genes following a large-scale gene trap and screening experiment. A total of 21 electroporation experiments were performed using the optimized IRESβgeo vectors in E14 and E14Tg2a ES cells. Following a three-tiered in vitro screen, less than 1% of G418 resistant colonies were identified as harboring gene trap integrations in candidate ES cell-specific genes.

Results for the 28 ES cell gene trap clones harboring integrations in candidate ES cell-restricted genes, including reporter gene expression profiles and molecular analyses, are given in Chapters four and five.

3.6 DISCUSSION

The results in this chapter demonstrate the successful design and application of an optimized IRESβgeo gene trap vector in a stringent three-tiered screening strategy to identify genes that are specifically expressed in ES cells *in vitro*. The two previously untested gene trap vectors, pGT1.8 (OPT) IRESβgeo and pGT1.8 (OPT/Nuclear) IRESβgeo, were used to provide optimized IRES-mediated translation of a cytoplasmic and a nuclear localized βgeo reporter gene, respectively.

Reading frame constraints for translation of the reporter gene, following integration of the vector into a transcriptionally active gene in ES cells, were overcome by the inclusion of an EMCV IRES element in the constructs. The predominant scanning mechanism of translation initiation in eukaryotic cells is 5' cap-dependent and positions the ribosome at the 5' initiator AUG codon of mRNA transcripts (Kozak, 1989). By contrast, the IRES element of picornaviral mRNA allows cap-independent ribosome binding (Pelletier and Sonenberg, 1988) and effective internal translation initiation in mammalian cells (Jang, et al., 1988, 1989; Molla et al., 1992; reviewed by Mountford and Smith, 1995). Sequences between nt 403 and 811 of the EMCV 5'-nontranslated region are required for efficient translation. They include a stem-loop structure and a stretch of pyrimidine-rich sequences near the initiation codon, both of which are essential for IRES function (Jang and Wimmer, 1990, reviewed by Ghattas et al., 1991). Two nuclear proteins, the nuclear La autoantigen normally involved in termination of

transcription by RNA polymerase III, and the 57 kD polypyrimidine tract-binding protein normally involved in splicing of mRNA precursors, are among the cellular proteins that bind to poliovirus and EMCV RNA IRES structures for efficient usage of the IRES by the cell's translation apparatus (Hellen *et al.*, 1993; Meerovitch *et al.*, 1993).

Insertion of an optimized EMCV IRES element between the splice acceptor and *βgeo* sequences in the pGT1.8 (OPT) IRES*βgeo* and pGT1.8 (OPT/Nuclear) IRES*βgeo* vectors was predicted to increase the number of detectable gene trap events, as integrations need only be in the correct orientation for internal translation of the reporter gene. The first report using an IRES*βgeo* vector in a large-scale gene trap study demonstrated a 3-fold increase over its parental non-IRES vector pSA*βgeo* (Friedrich and Soriano, 1991) in the number of G418 resistant colonies, as well as a substantial increase in the number of *lacZ* positive colonies from 30% to 75% (Chowdhury *et al.*, 1997). The pSA*βgeo* vector, which is the first described *βgeo* fusion gene trap electroporation construct, harbors a point mutation in *neo* that is known to reduce its enzyme activity (Yenofsky *et al.*, 1991), and appears to pre-select for genes expressed at high levels (Skarnes *et al.*, 1995).

The two IRES*βgeo* gene traps used in the current study were derived from the pGT1.8*geo* parental vector in which the *neo* point mutation is corrected, providing a 1.5-fold increase in the number of G418 resistant colonies recoverable over pSA*βgeo* (Skarnes *et al.*, 1995). A recent large-scale gene trap study reported the use of a pGT1.8*geo*-derived IRES*βgeo* vector (Bonaldo P. *et al.*, 1998), with translation of the reporter gene mediated from a mutagenic IRES AUG initiation codon downstream of the native codon (Ghattas *et al.*, 1991). Consistent with the preliminary report of Chowdhury *et al.*, (1997), this resulted in an increase of more than 2.7-fold over the parental pGT1.8*geo* vector in the number of G418 resistant colonies, and a similar substantial increase from 29% to 82% in the proportion of βgal positive colonies (Bonaldo P. *et al.*, 1998). Results of the current study confirm comparable transfection efficiencies between the cytoplasmic and nuclear localized (OPT) IRES*βgeo* vectors (Table 3.1) and an overall combined efficiency in the large-scale gene-trap and screening experiment of ~1:236,000 (Table 3.5). This efficiency is 3.4-fold higher than reported for the parental non-IRES pGT1.8*geo* vector in the previously cited study

117

(Bonaldo P. *et al.*, 1998), though it cannot be excluded that the differences in overall transfection efficiency between the two studies may also contribute to this increase. The yield of G418 resistant colonies that were also *lacZ* positive was also substantially more efficient, at above 70% for each of the optimized vectors (Table 3.1), and at 85% for the total number of colonies stained with X-gal in the large-scale secondary screening experiment (Table 3.5).

As previous reports have suggested that the cytoplasmic expression of βgal activity can be detrimental for developmental processes including gametogenesis in transgenic mice (Kimura *et al.*, 1994), the SV40 nuclear localization signal was incorporated into the preferred vector of this study to maximize the chances for germline transmission of gene trap ES cell clones. It was also anticipated that nuclear localization of βgal activity would improve visualization of the resultant *lacZ* expression profile in the cells and tissues of transgenic mice, and more clearly distinguish βgal activity of the reporter gene from possible non-specific cytoplasmic staining. While the number of colonies and the proportion of βgal positive colonies is slightly higher for the cytoplasmic optimized IRES*βgeo* vector, the inclusion of a nuclear localization signal in the second vector did not have a marked effect on either *lacZ* or *neo* activity (Table 3.1).

Although the use of an IRES element theoretically dictates that every G418 resistant colony should be positive for βgal activity, the failure to detect *lacZ* expression in 15% of neo resistant colonies in this study, may be due to the capture of genes with weak promoter activities, as previously seen for the *Lif* gene (Mountford *et al.*, 1994). This possibility is consistent with the wide variation seen in βgal staining intensities generated in this study (Figure 3.4) and demonstrates both the potent neomycin resistance activity and high efficiency of IRES-mediated translation. The incorporation of an IRES element in the vectors may also have allowed for the detection of some classes of genes where amino terminal fusions with the *βgeo* reporter would cause a partial or complete loss of reporter activity, as previously demonstrated for the secretory trap pGT1.8TM (Skarnes *et al.*, 1995).

Heterogeneous X-gal staining was evident in many clonal ES cell cultures, and often restricted to subsets of undifferentiated ES cells (Figure 3.4). Such variegated expression might be associated with asynchronous growth for a population of ES cells in culture, where activation of an endogenous promoter occurs at a certain time during

118

the 18-24 h cell cycle (Robertson, 1987). Differences in expression of individual cells could also arise if the trapped gene is prone to inactivation via a *cis*-acting epigenetic mechanism such as *de novo* DNA methylation, which is known to occur in ES cells (Szyf *et al.*, 1990).

The three-tiered *in vitro* screening strategy was established to identify those few gene trap events which are representative of ES cell-specific genes in the large number of gene trap clones generated (Table 3.5). Using a series of control ES cell lines which harbor *βgeo* or *lacZ-neo* insertions, in some cases under IRES-mediated translation (section 3.4.1), the three screens were optimized to select for both ES cell-specific neomycin resistance (Table 3.2) and βgal activity in monolayer cultures, and for ES cell-specific neomycin resistance in embryoid body cultures (Table 3.3). The OKO160 cell line, which expresses the *βgeo* reporter at the ES cell-restricted *Oct-4* locus (Mountford *et al.*, 1994, 1998), provided a definitive positive control at each level of screening for the identification of reporter gene insertions in ES cell-specific genes (Figure 3.3).

The initial screen provided a broad morphological assessment for colonies showing ES cell-specific neomycin resistance following electroporation with the gene trap, and subsequent culture in the presence of LIF and G418 (Figure 3.3A). This preliminary screen identified 775 (18%) of the 4345 stable gene trap events generated from the large-scale experiment of 21 gene trap transfections as "pure ES" colonies, apparently displaying ES cell-restricted neomycin resistance (Table 3.5).

The secondary screen confirmed that the selected colonies were able to expand and differentiate normally in monolayer cultures in the presence of LIF, and permitted further selection of clones based on ES cell-restricted *lacZ* expression (colony growth without G418) and ES cell-restricted neo resistance (colony growth with G418), (Figure 3.3B). This 2-dimensional screening assay was highly effective in identifying those primary clones which appeared morphologically to be expressing *neo* at an ES cell-restricted locus on transfection culture plates, but in fact detected lacZ and neo reporter gene products in both ES cells and their differentiating progeny upon expansion. The efficiency of the secondary screen is demonstrated by the elimination of 721 (93%) of the positive primary colonies following X-gal staining of expanded colonies, reducing the number of candidate ES cell-restricted gene trap clones from 775 to 54 (Table 3.5).

The tertiary screening assay is the most stringent of the screens, and provided a further 2-fold enrichment (28/54) over the secondary screen for putative ES cell-specific gene trap integrations (Table 3.5). This final screen confirmed the expression of *neo* at an ES cell-restricted locus, following culture of the ES gene trap cells in the absence of LIF and as aggregates in hanging drop cultures (Figure 3.3C). Pluripotent ES cells in close contact with differentiating cells in the aggregate cultures are effectively eliminated by the efficient induction of differentiation, and possibly by programmed apoptosis associated with EB formation. This inductive process may be promoted by the presence of VE-like cells in EBs and by embryonic VE expression of *Bmp-2* during cavitation (Mummery *et al.*, 1991; Coucouvanis and Martin, 1995; 1999). As previously demonstrated for the known *Oct-4* locus targeted in OKO160 ES cells (Mountford *et al*, 1998), the recovery and expansion of pluripotent ES cells from aggregate cultures is achieved only by the continuous elimination of differentiated cells from the aggregates where the gene trap ES cells express *neo* under the control of an ES cell-specific locus. Culture of aggregates in the presence of G418 in the tertiary screening assay definitively identified 28 gene trap ES cell lines that are expressing *neo* at an ES cell-specific locus, while culture without selection confirmed the capacity for each of these lines to completely differentiate and eliminate pluripotency from EB cultures.

Elimination of over 99% of the stable gene trap integrations generated in this study, confirms the stringent nature of the selection strategy for gene trap events of interest. The established three-tiered *in vitro* pre-screen has identified an anticipated low number of gene trap clones that are specifically expressed in ES cells *in vitro*. While several reports have now described the *in vitro* identification of restricted expression profiles for gene trap ES cell clones (for example, Forrester *et al.*, 1996; Shirai *et al.*, 1996; Baker *et al.*, 1997; Bonaldo P. *et al.*, 1998; Voss *et al.*, 1998a; Xiong *et al.*, 1998), this is the first report of a screening strategy to specifically identify ES cell-restricted genes by gene trap cloning.

In summary, these results demonstrate that the pGT1.8 (OPT) IRES*βgeo* and pGT1.8 (OPT/Nuclear) IRES*βgeo* vectors are effective for trapping genes at high efficiencies, and provide detectable reporter expression for a high proportion of these events. Combined with a stringent three-tiered *in vitro* pre-screen, less than 1% of the clones generated with these vectors have been identified as clones of interest for insertions in

ES cell-restricted genes. Molecular and expression analyses for these 28 gene trap cell lines are discussed in the following chapters.

Chapter Four

MOLECULAR ANALYSIS OF 28 GENE TRAP INTEGRATIONS IN CANDIDATE ES CELL-SPECIFIC GENES

4.1 INTRODUCTION

While several large-scale gene trap studies date have demonstrated the relative ease of generating large numbers of clonal gene trap cell lines for analysis, elucidation of the trapped endogenous gene has not always been straightforward (Wurst *et al.*, 1995; Chowdhury *et al.*, 1997; Hicks *et al.*, 1997; Townley *et al.*, 1997; Voss *et al.*, 1998a, Zambrowicz *et al.*, 1998) and remains a limitation of the overall strategy. Sequence analysis of over 150 gene trap cell lines from one study demonstrated that only 37% of these lines yielded amplifiable 5'-RACE-PCR products and readable single sequences showing correct use of the splice acceptor (Townley *et al.*, 1997). In many cases, the splice site was inefficiently used resulting in recovery of intron-containing sequences. In other cases, *trans*-splicing events yielded more than one sequence 5' of the splice site. The inability to amplify PCR products has been shown to be largely due to deletions in the vector sequence, or may sometimes result from the use of a cryptic splice acceptor site downstream of the constructed splice site in the vector sequence (Townley *et al.*, 1997; J. Brennan & J. Wallis, Centre for Genome Research, Edinburgh, UK; *pers. comm.*). The occurrence of such events before or during integration of the

vector into the genome is, however, likely to be influenced by the design and delivery of the specific gene trap vector in use.

The published limitations of the gene trap cloning strategies suggested a proportion of the 28 gene trap clones derived from the current *in vitro* screening strategy might not yield useful information regarding the endogenous genes. In a previous ES cell gene trap cloning experiment employing an IRES*βgeo* vector, cloning and analysis of the tagged cDNA from 42 gene trap lines identified 19 (45%) novel sequences, 8 (19%) sequences with homology to known ESTs, 11 (26%) sequences with homology to known genes representing all major classes, and 4 (9.5%) containing multiple 5'-RACE products (Chowdhury *et al.*, 1997). The criteria for which these clones were selected for sequence analysis is not clear however, and an overall analysis for the efficiency of generating correctly integrated and spliced events from the hundreds of integrations generated in this study was not provided (Chowdhury *et al.*, 1997; Bonaldo P. *et al.*, 1998). Preliminary experiments by W. Skarnes and co-workers (unpublished) suggested the use of an IRES*βgeo* vector may result in a high incidence of integrations in rRNA transcription units, though this has not been substantiated.

This chapter describes the molecular analysis of the 28 candidate ES cell-restricted genes in gene trap clones identified from the three-tiered pre-screening process.

4.2   NORTHERN BLOT FUSION TRANSCRIPT ANALYSES

To analyse gene trap cell line mRNAs for the presence of inefficiently spliced fusion transcripts, poly A$^+$ mRNA Northern blots were prepared (sections 2.2.2f & 2.2.4) and hybridized with a radiolabelled probe *"En2"*, corresponding to approximately 1.2 kb of *En-2* intron sequence isolated from the pGT1.8 (OPT/Nuclear) IRES*βgeo* construct (sections 2.1.2, 2.2.5a & 2.2.6b). For eleven cell lines analysed (E14 control, ZIN40, GT#'s C3, 4, 14, 15, 18, 24, 26, 50 and 54) the *En2* intron probe detected a non-specific hybridization product migrating high on the gel blots for each sample, while a specific single hybridization band was detected only in the GT#15 sample (data not shown). Although the results obtained suggested a majority of gene trap cell lines had made efficient use of the vector splice site, this data was not considered conclusive as a suitable control preparation of mRNA containing *En-2* intron sequence was not available.

I

ZIN40 ST534 E14 — GTC3 GT4 GT8 GT9 GT10 GT11 GT14 GT15 GT17 GT18

6.6kb→
5.0 →
3.6 →
2.6 →
1.9 →

II

ZIN40 ST534 E14 — GT21 GT24 GT26 GT27 GT28 GT29 GT30 GT32 GT36 GT37

6.6kb→
5.0 →
3.6 →
2.6 →
1.9 →
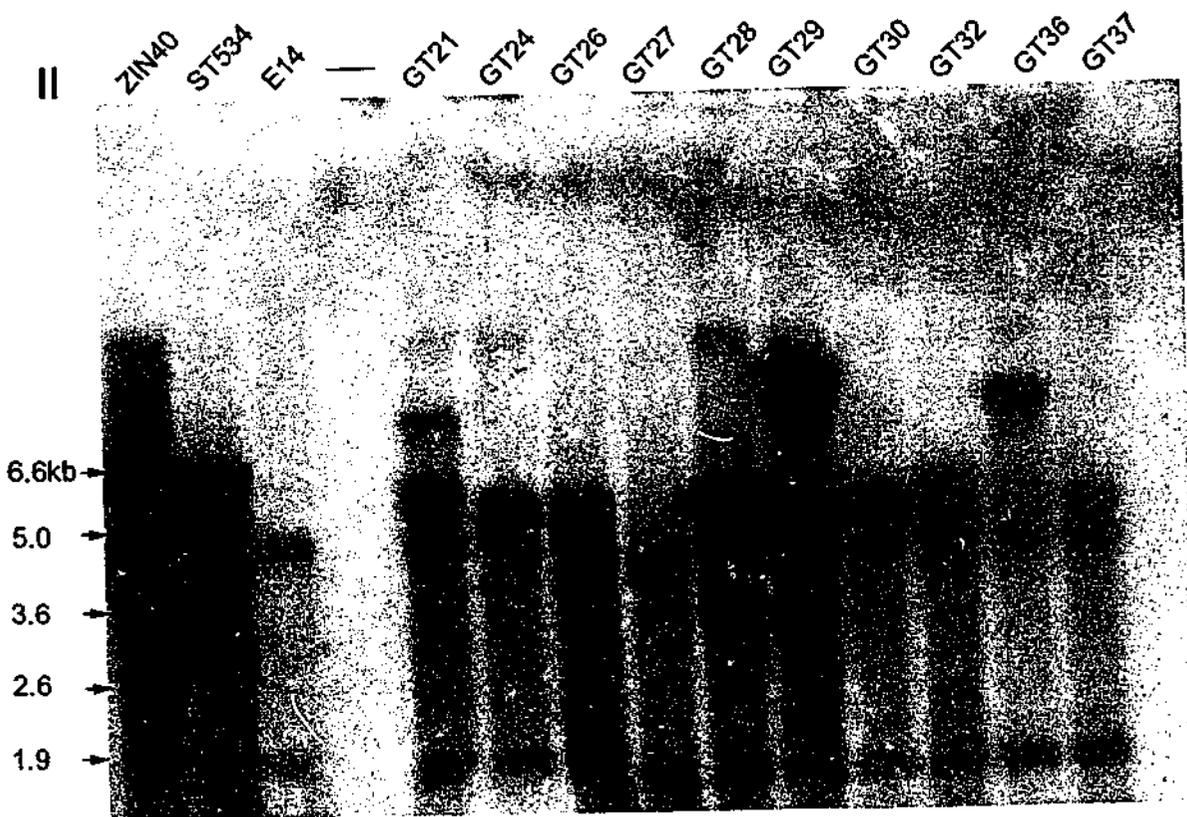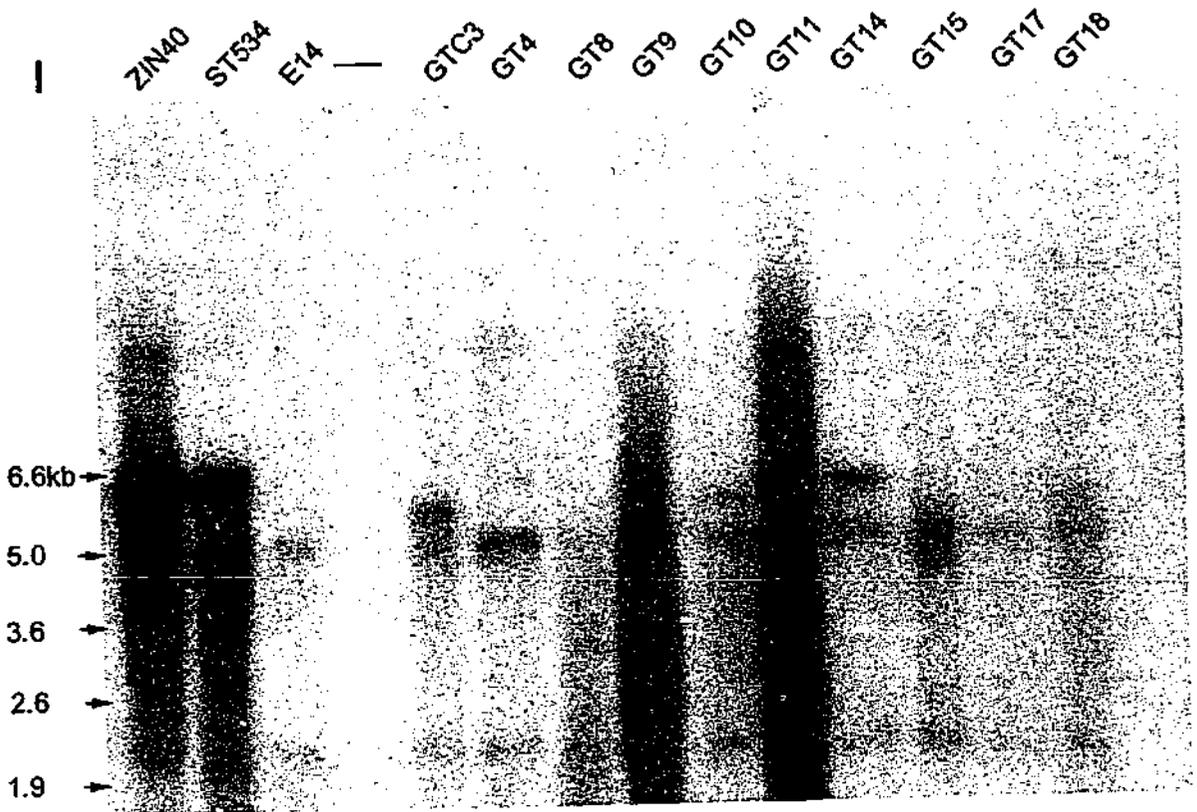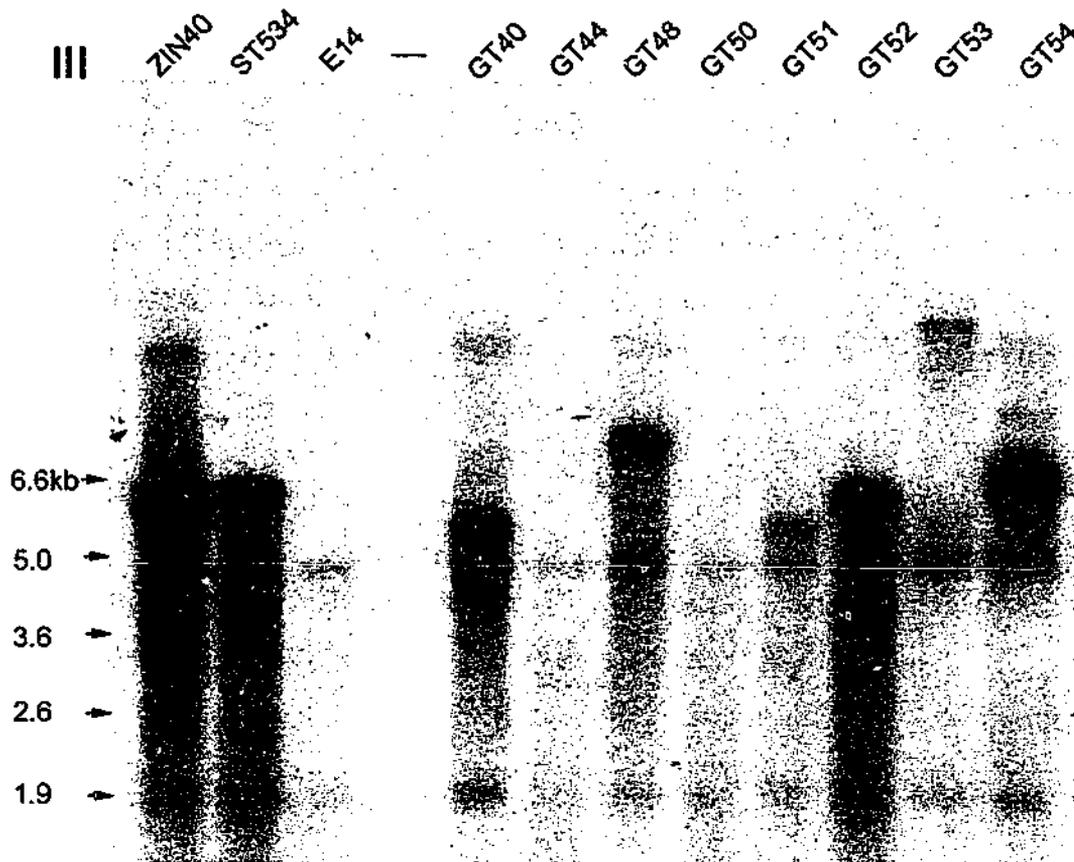
Figure 4.1: The detection of fusion transcripts in 28 gene trap ES cell lines by Northern blot analysis. For each gene trap ES cell line 3 µg of poly A⁺ mRNA was separated on a denaturing gel and analysed by Northern blot hybridization with a *βgeo* probe. Autoradiographic exposure was for 3-5 days at -80°C with intensifying screens. On each blot (I-III), negative (E14) and positive (ZIN40 and ST534) ES cell mRNA controls were included, with the ST534 cell line expected to yield a fusion transcript size of 6.0 kb. A single fusion transcript band of a size greater than the expected vector contribution of 5.0-5.1 kb was detected for 10 gene trap cell lines (GT#'s C3, 9, 10, 17, 18, 26, 30, 37, 51, 52). One or more additional, but weaker, hybridization band(s) was detected for 13 cell lines (GT#'s 4, 14, 15, 21, 24, 28, 29, 32, 36, 40, 48, 53, 54). Fusion transcripts were not visible for 3 lines (GT#'s 8, 44, 50), while a hybridization band of less than 5 kb was detected for two other cell lines (GT#'s 27, 40) and the mRNA for the GT#11 cell line appeared degraded. Residual rRNA bands were evident in all RNA preparations excepting GT#8.

Northern blots prepared from 3 μg of a new poly A$^+$ mRNA sample for each of the 28 cell lines, were hybridized with a radiolabelled probe "βgeo", corresponding to approximately 2.5 kb of *lacZ-neo* sequence isolated from the pGT1.8 (OPT/Nuclear) IRES*βgeo* vector (sections 2.1.2 & 2.2.5a), (Figure 4.1). Poly A$^+$ mRNA was also prepared for the E14 ES cell line as a negative control, and for the ST534 and ZIN40 ES cell lines (section 2.1.5) to serve as positive controls on each blot. The ST534 gene trap cell line was previously reported to generate a fusion transcript of 6.0 kb by Northern blot analysis (Skarnes *et al.*, 1995). This cell line was chosen to provide an indication of upstream fusion transcript sequence length given that the gene trap vector was predicted to contribute 5.0-5.1 kb to the fusion product.

Hybridization with the *βgeo* probe (Figure 4.1) detected residual rRNA 28s and 18s bands for all samples, excepting GT#8 which shows only a very faint signal. Cell lines GT#44 and GT#50 did not show any hybridization bands other than the rRNA bands, indicating the probable event of either a vector deletion event or a very short 5' endogenous sequence, such that the fusion transcript migrates to same distance as the approximately 5 kb 28S band. For the line GT#9, the mRNA appeared to be degraded and subsequently no PCR product was obtainable from this line (Table 4.1). All other lines showed one major fusion transcript band ranging in size from approximately 5.2 kb to above the highest visible marker band of 6.6 kb. One or more additional, although much weaker bands was evident for 13 cell line preparations. Although the presence of multiple bands was suggestive of inefficiently spliced or *trans*-spliced fusion transcripts, direct sequence analysis of mRNA preparations did not support this prospect in all instances (Table 4.1). A band size of less than that expected for vector contribution alone was visible in two instances (GT#27 and GT#40), suggesting a possible vector deletion event. The detection of, and estimated fusion transcript sizes for each of the 28 clones (summarized in Table 4.1) indicated that most transcripts have at least 200 bp of 5' endogenous gene sequence, and were therefore suitable for analysis by 5'-RACE.

4.3   IDENTIFICATION OF GENE TRAP INSERTIONS BY SEQUENCE ANALYSIS OF 5'-RACE PRODUCTS

Fusion transcripts were sequenced by direct analysis of 5'-RACE products (sections 2.2.8a & 2.2.9a). Direct sequencing of cellular mRNA bypasses the need to clone each
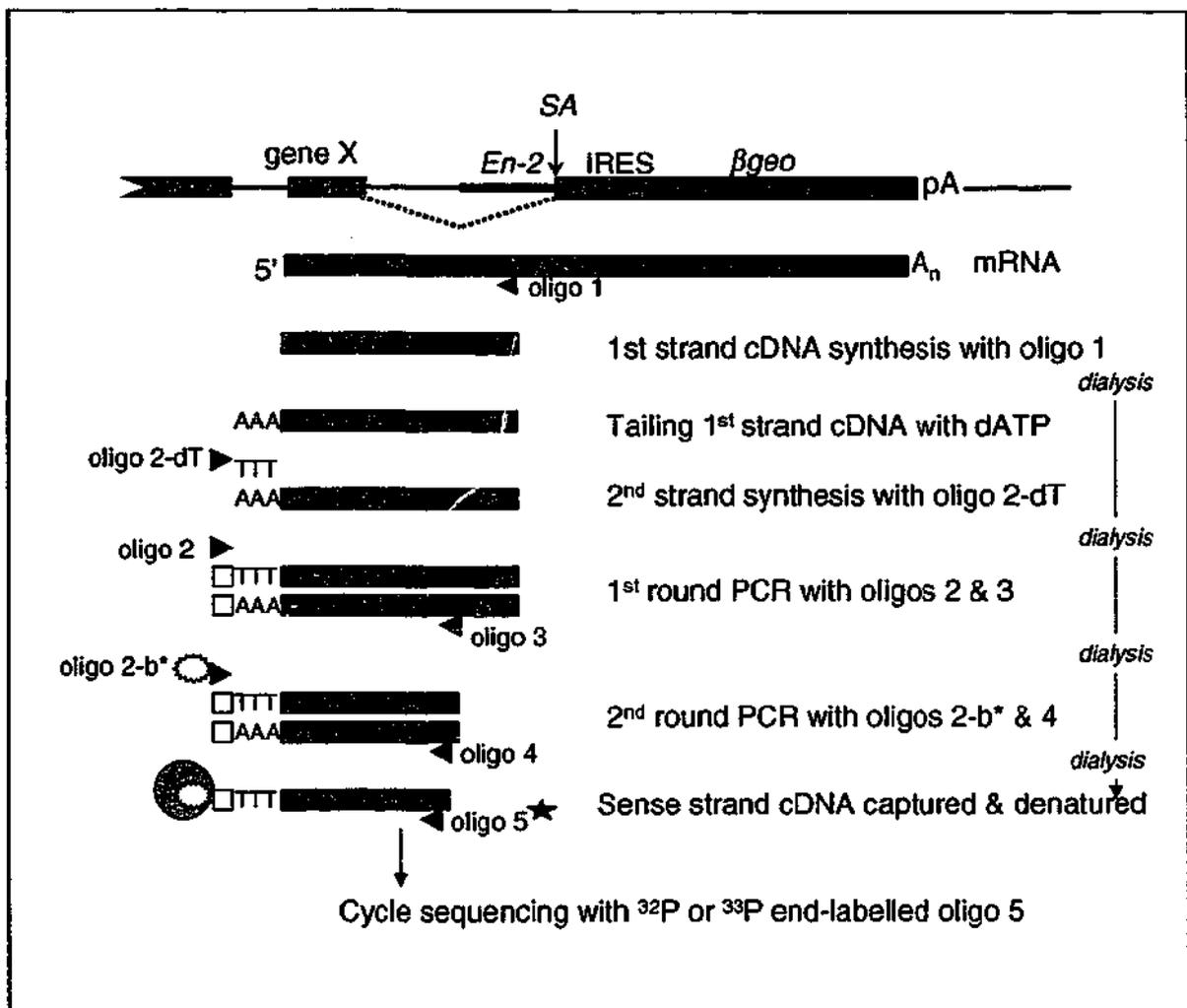
Figure 4.2: A 5'-RACE and direct sequencing strategy to identify pGT1.8 (OPT/Nuclear) IRESβgeo fusion transcripts. 1st strand cDNA was synthesized for 28 mRNA fusion transcripts in RT reactions with an antisense primer to vector IRES sequence (oligo 1), starting 218 nts downstream from the *En-2* splice acceptor site (SA). 1st strand products were 5' A-tailed with dATP and 2nd strand cDNA synthesized with a forward anchor primer including a 5' oligo(dT) stretch (oligo 2-dT). Two rounds of PCR were performed using the 5' anchor primer without the T-tail (oligo 2) and reverse primers nested in the *En-2* exon sequence, starting 129 nts (oligo 3) and 102 nts (oligo 4) 3' of the SA. In the second PCR round the 5' anchor primer was biotinylated (oligo 2-b*), allowing sense strand cDNA to be captured onto streptavidin coated magnetic beads. To purify and size select for cDNA products of at least 300 bp, microdialysis steps were performed between 1st and 2nd strand syntheses and the two rounds of PCR. Captured cDNA strands were denatured and used directly as templates for cycle sequencing with a radioactive end-labelled oligonucleotide primer starting 75 nts 3' of the SA, in *En-2* exon sequence (oligo 5).

PCR product for analysis by standard sequencing procedures. This strategy, outlined in Figure 4.2, was adapted from a method published for the rapid sequence analysis of secretory trap events (Townley *et al.*, 1997), and is specific for the analysis of fusion transcripts incorporating the pGT1.8 (OPT) IRES*βgeo* and pGT1.8 (OPT/Nuclear) IRES*βgeo* vectors.

5'-RACE products were generated and amplified as described in section 2.2.8a, using 3'-end gene trap vector-specific primers and a 5'-end anchor primer. Sequences for these oligonucleotide primers are listed in section 2.1.3 and their positions in the fusion transcript and generated cDNAs, described below, are shown schematically in Figure 4.2.

The 1st strand cDNA was synthesized with an antisense RT primer from the IRES sequence (oligo 1), beginning 218 nts downstream from the *En-2* splice acceptor site. Initially, a primer starting 159 nts 3' of the splice site in the 166 bp *En-2* coding sequence was employed for the RT reaction. However, the subsequent identification of non-vector mouse *En-2* sequence from the amplified products for one clone, GT#53 (data not shown), suggested that RACE products generated from the viral IRES sequence of fusion transcripts would avoid the need to interpret such likely erroneous results. This clone was later confirmed as being an unspliced event containing *En-2* vector intron sequence upstream of the splice site (Table 4.1).

Following 5' A-tailing of 1st strand cDNA products, second strand cDNA was synthesized using a T-tailed anchor primer (oligo 2-dT), containing a 5' stretch of 17 thymidine residues. Two rounds of PCR were performed with reverse primers nested in the *En-2* exon vector sequence, starting 129 nts (oligo 3) and 102 nts (oligo 4) downstream from the splice site, and the forward anchor primer without the 5' oligo(dT) stretch (oligo 2). In the second round of PCR, the anchor primer was biotinylated (oligo 2-b*) to enable the sense strand cDNA products to be captured onto streptavidin coated magnetic beads (section 2.2.9a).

For confirmation of amplifiable 5'-RACE products for each clone, an aliquot of second round PCR products was hybridized (sections 2.2.8a & 2.2.3c) with a ^{32}P end-labelled probe "*En41*", corresponding to the antisense strand of *En-2* exon sequence, for 41 nts 3' of the vector splice site (sections 2.1.2 & 2.2.5b). The 5'-RACE RT primer (oligo 1, 218-198 nts 3' of the splice site) was employed as a negative control for each
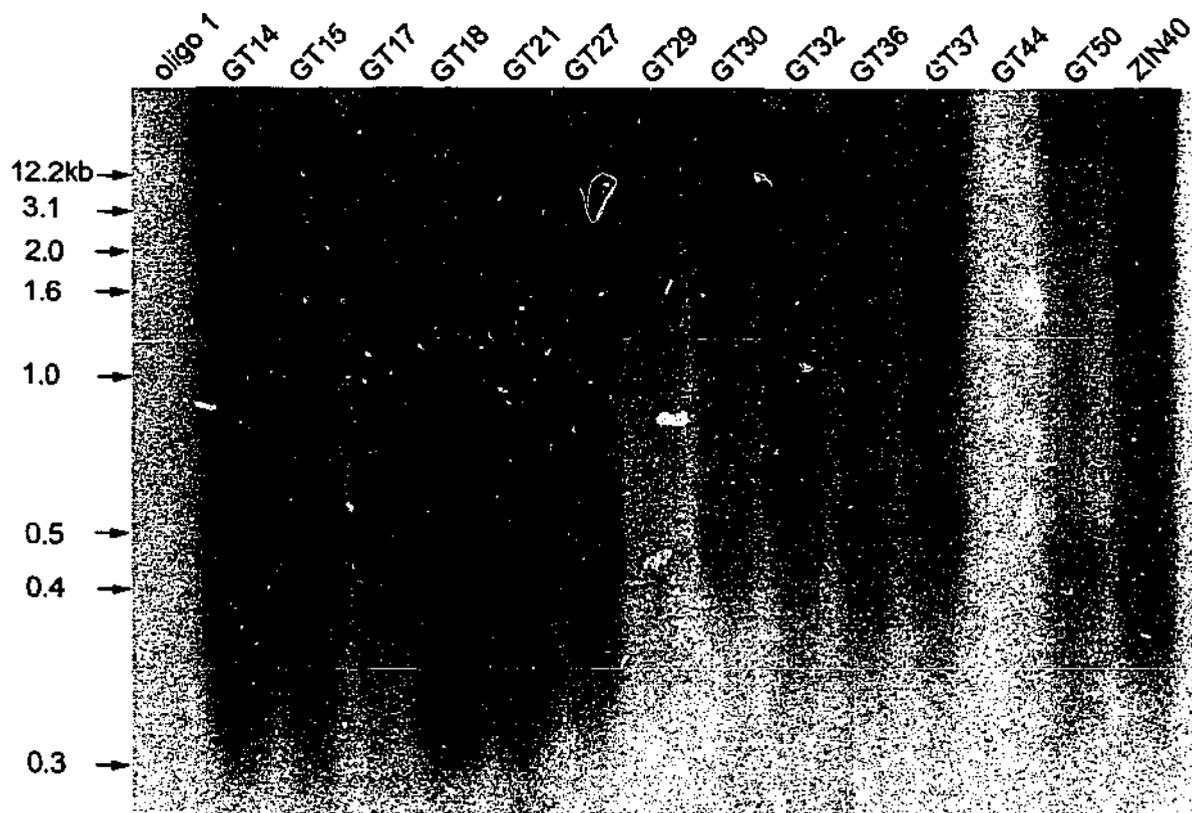
Figure 4.3: The detection of 5'-RACE products from gene trap ES cell lines by Southern blot hybridization. To determine whether 5'-RACE products generated from 28 gene trap clones were amplifiable, 3 µl of each 40 µl 2nd round PCR reaction mix was electrophoresed and DNA transferred to a nylon membrane by a rapid alkaline dry blotting procedure. DNA products were hybridized with the *En41* probe corresponding to the first 41 nts of *En-2* exon sequence, 3' to he splice acceptor site in the pGT1.8 (OPT/Nuclear) IRESβgeo gene trap vector. The 5'-RACE RT primer (oligo 1) to IRES sequence downstream in the vector from the probe sequence, was used as a negative control for hybridization. Autoradiographic exposure was for 3-6 h at -80°C with intensifying screens. For the 14 cell lines shown here, all except one (GT44) are positive for PCR products generated from the gene trap sequence. The smear pattern seen is characteristic for 5'-RACE products, as each 1st strand cDNA length generated is A-tailed and amplified. Products of less than 300 bp were eliminated by dialysis steps.
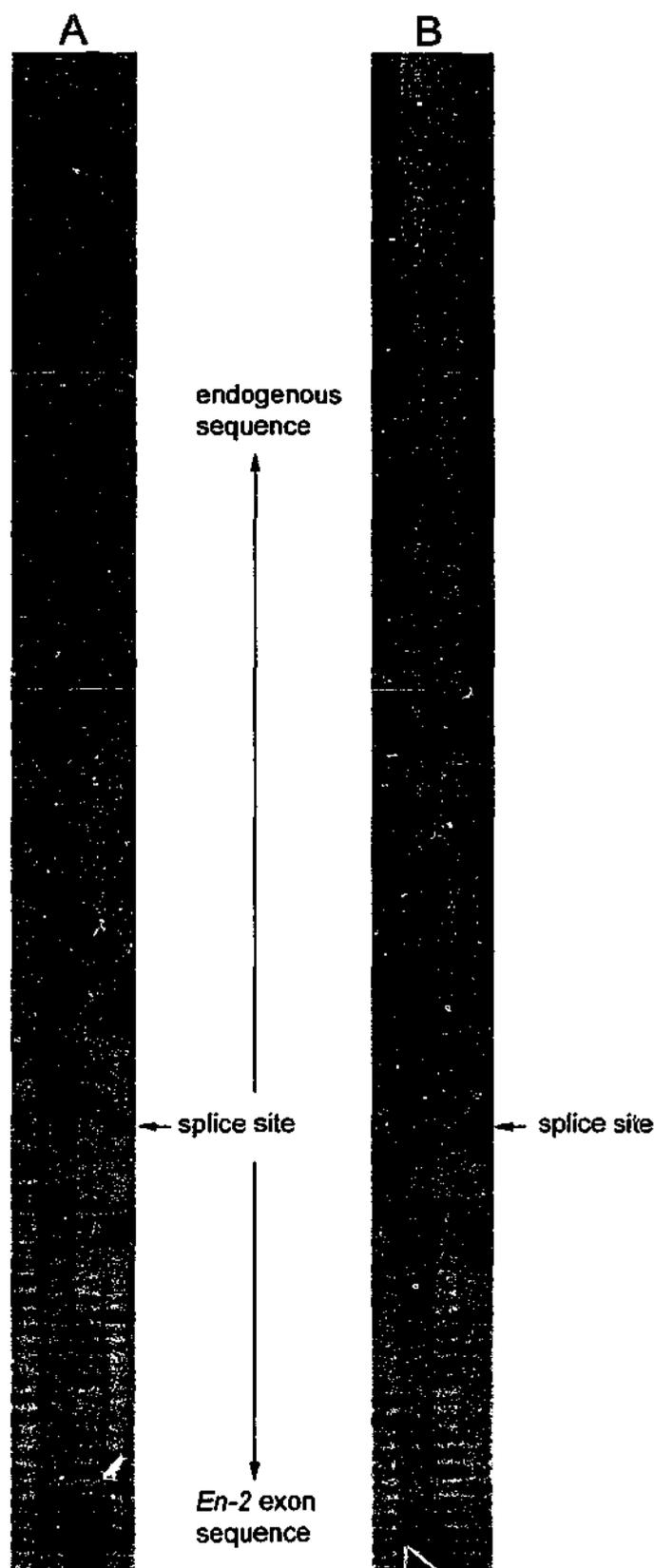
Figure 4.4: *Direct sequence analysis of 5'-RACE products from ES cell lines harboring gene trap fusion transcripts.* Amplified 5'-RACE products were immobilized on magnetic beads and denatured for direct sequencing with an end-labelled primer, nested 75 nts downstream from the splice site in *En-2* exon sequence. Thermal cycle sequence products were separated by electrophoresis on a 6% denaturing polyacrylamide gel and sequence ladders manually analysed following autoradiographic exposure o/n at RT°C, without intensifying screens. (A) Upstream of the *En-2* splice acceptor site, a single 5' endogenous sequence ladder of about 100 nts was provided by electrophoresis of sequence products for 2 h at 50W, followed downstream by about 30 nts of *En-2* exon sequence. (B) For several samples, endogenous sequence upstream of the splice site was difficult to read either as a result of an incorrect splicing event, or superimposed multiple sequence products.

GT #	βgal pattern in ES cells with G418 300µg/ml	Fusion transcript band sizes (~kb) βgeo probed	En41 probed PCR product	Direct sequence (seq) outcome for 5'-RACE products (5'→3')
G3	xx/xxx, P+++	single ~ 5.5	yes xxx	90 bp known seq & SA seq
4	xx in ~3 patches	weak >6.6 & ? ~5.7	no (x3)	no PCR product – no seq
8	xx/xxx	faint – band not clear	yes xxx	96 bp known seq & SA seq
9	xxx	single ~5.5, slight smear	yes xxx	108 bp intron seq & SA seq
10	xx	single ~5.5	yes xx	multiple seqs & SA seq
11	xx/xxx, P+	RNA smeared	no (x3)	no PCR product – no seq
14	xx, P++	~5.7 & faint >6.6	yes xxx	multiple seqs, 39 bp intron & SA seq
15	x/- in some cells	~5.0 smeared & faint >6.6	yes xxx	intron + other seq & SA seq
17	xxx, some cols	single ~5.5	no (x1) yes (x2) x/-	low PCR product - no readable seq
18	xx, P+++	single ~5.4	yes xxx	90 bp known seq & SA seq
21	xx, P++	~5.6 & 2 bands >6.6	yes xxx	intron + other seq & SA seq
24	xx/xxx	~5.3 & faint >6.6	yes xxx	multiple seqs & SA seq
26	xxx, P++	single ~5.4	yes xxx	52 bp known seq & SA seq
27	xx/xxx, P++	2-300 bp < 5 kb	yes xxx	40 bp tandem vector En-2 exon seq + 55 bp intron & SA seq
28	xxx	strong ~5.4 & faint >6.6	yes xxx	101 bp novel seq & SA seq
29	xx/xxx	strong >6.6 & mod ~5.3	yes xxx	intron + other seq & SA seq
30	xx, P++	single ~5.3	yes xx/xxx	multiple seqs & SA seq
32	x/- in a few cells	~5.3 smeared or ?2 bands	yes xx	43 bp tandem vector En-2 exon seq & SA deletion
36	x (xx –G418)	>6.6 & higher faint band	yes xx/xxx	118 bp novel seq & cryptic SD/SA
37	xx, P++	single ~5.2	yes xx/xxx	151 bp novel seq & SA seq
40	xx, P+++	~5.4 & 2 x faint ~6.5 & >6 1 band below 28S	no (x1) yes (x3) x/-	low PCR product – no seq
44	x in some cells	Only rRNA bands visible	no (x4)	no PCR product – no seq
48	xx in some cells	>6.6 & faint higher band	yes xxx	intron + other seq & SA seq
50	xx, P++	Only rRNA bands visible	yes xx/xxx	short unreadable seq & SA seq
51	xx/xxx, P+++	single ~5.5	yes xxx	90 bp known seq & SA seq
52	xxx, P+	single ~6.0	yes xx/xxx	131 bp novel seq & SA seq
53	xx, P+++	2 bands > 6.6	yes xxx	102 bp intron seq & SA seq
54	xx	strong ~6.6, 2 faint higher	yes xxx	126 bp novel seq & SA seq

Table 4.1: Summary of analyses for 28 ES cell-specific gene trap cell lines. For mRNA from gene trap ES cell clones showing ES cell-restricted βgeo expression *in vitro*, Northern blot hybridization with a βgeo probe indicated those clones containing one or more fusion transcript(s) of a size larger than the predicted vector contribution of 5.0-5.1 kb. 5'-RACE products generated using vector-specific nested primers were detected by Southern blot hybridization with the En41 probe to En-2 exon sequence. Direct sequencing of 5'-RACE products identified 10 cell lines (shaded) yielding a single sequence ladder upstream of the En-2 splice acceptor site (SA). The trapped sequence was identical for three different clones (**). For 13 cell lines, errors occurred in the use of the SA or in the vector integration event. x/-=weak; x=mild; xx=moderate; xxx=marked; P=heterogeneous *lacZ*.

hybridization. An example of an *En41* probed DNA blot is shown in Figure 4.3. The smear detected for most samples is characteristic for 5'-RACE products, as generated 1st strand cDNAs of differing lengths are A-tailed and subsequently amplified. However, dialysis steps included in the 5'-RACE protocol (section 2.2.8a) eliminated those products that were less than 300 bp in size. Results of the *En41* hybridization of PCR products for the 28 clones are summarized in Table 4.1. For five cell lines, the hybridization signal was either absent or very weak for repeat analyses, and no sequence information was subsequently obtainable for each of these cell lines (Table 4.1).

Thermal cycle sequence reactions were generated directly from denatured immobilized PCR products (section 2.2.9a), employing a ^{32}P or ^{33}P end-labelled antisense sequencing primer starting 75 nts downstream from the splice site in *En-2* exon sequence (oligo 5, Figure 4.2), (section 2.2.5b). Sequencing products from two or more independent 5'-RACE reactions for each clone were electrophoresed and analysed manually following autoradiographic exposure (section 2.2.9a). Electrophoresis for 2 h at 50 watts enabled visualization of a sequence ladder for 5' endogenous sequence upstream of the *En-2* splice acceptor site and about 30 nts of downstream *En-2* exon sequence (Figure 4.4A). For several clones, although the *En-2* exon sequence was clear, smeared ladders or multiple sequences upstream of the splice site made interpretation of trapped endogenous sequences difficult (Figure 4.4B).

The outcome for direct sequencing of 5'-RACE products from the 28 clones of interest is summarized in Table 4.1. For those clones yielding readable sequences, on average more than 100 bp of sequence was obtained upstream of the splice site, with the longest read being 154 bp. For longer gel run times sequence reads ended at the same place, indicating the limits for the direct sequencing protocol. Several attempts to adapt this protocol for automated sequencing of the captured 5'-RACE products met with limited success.

4.4 SEQUENCE ANALYSIS FOR 28 GENE TRAP INTEGRATIONS

Results in Table 4.1 show that 23 of the 28 gene trap cell lines (82%) provided useful sequence products. For all but one of these, correct *En-2* exon sequence 3' of the splice site was confirmed in sequence ladders, while for the remaining line (GT#32) the first 18 bp of *En-2* exon sequence was absent. Ten of the 28 cell lines (36%), (shaded boxes, Table 4.1) provided a single readable sequence ladder extending 90-154 bp 5' of the

126

splice acceptor site. These sequences were compared to non-redundant databases using the BLASTN and the BLASTX algorithms (Altschul *et al.*, 1990; 1997), and also compared for sequence homology with mouse and human genome databases (section 2.2.9c).

Five of the ES cell gene trap clone sequences corresponded to known cDNAs. One sequence (GT#8) displayed high homology to mouse repetitive LINE elements, another (GT#26) was homologous to *βgeo* sequence suggesting a tandem vector copy, and three (GT#'s C3, 18, 51) showed identical 90 bp sequence reads upstream of the splice site with 100% homology to a processed gene related to a J kappa recombination signal sequence binding protein (*RBP-Jκ*).

Homology searches for the 5' sequences of the five remaining lines (GT#'s 28, 36, 37, 52, 54) did not identify cDNA sequence matches and were considered to represent insertions in novel genes (discussed in section 4.6). None of these five sequences showed homology to gene trap vector sequence, nor to each other. One of these lines (GT#36), however, included 32 bp of most 5' *En-2* intron sequence and apparently made correct use of the predicted *En-2* splice acceptor. This may suggest the use of a cryptic splice donor site in the *En-2* intron sequence and probable integration of the gene trap into the endogenous gene exon.

For the remaining cell lines that yielded sequencing products, errors either in the use of the vector splice acceptor site or in the vector integration event, were observed to have occurred. Examples of these types of events are presented in Figure 4.5. For three of the lines (GT#'s 10, 24, 30), the splice site appeared to have been used efficiently on some occasions, although one or more sequence reads suggested multiple sequencing products 5' of the splice acceptor sequence, making it difficult to interpret results. For seven lines, sequence analysis showed the *En-2* splice acceptor was not used correctly. Sequence reads for these cell lines either contained *En-2* intron sequence alone (GT#'s 9, 53) or *En-2* intron sequence and endogenous sequence (GT#'s 14, 15, 21, 29, 48). Sequence reads for two cell lines (GT#'s 27, 32) suggested the integration of a tandem vector copy, with either the use of alternate splice acceptor and donor sites within the vector sequences, or possible deletion events in the vector sequences. In the case of GT#27, the 5' sequence read matched 40 bp from the 3'-end of the 160 bp *En-2* vector exon sequence, followed by 55 bp of unspliced *En-2* intron sequence upstream of the predicted *En-2* splice acceptor and exon sequence. For GT#32, the 5' sequence read
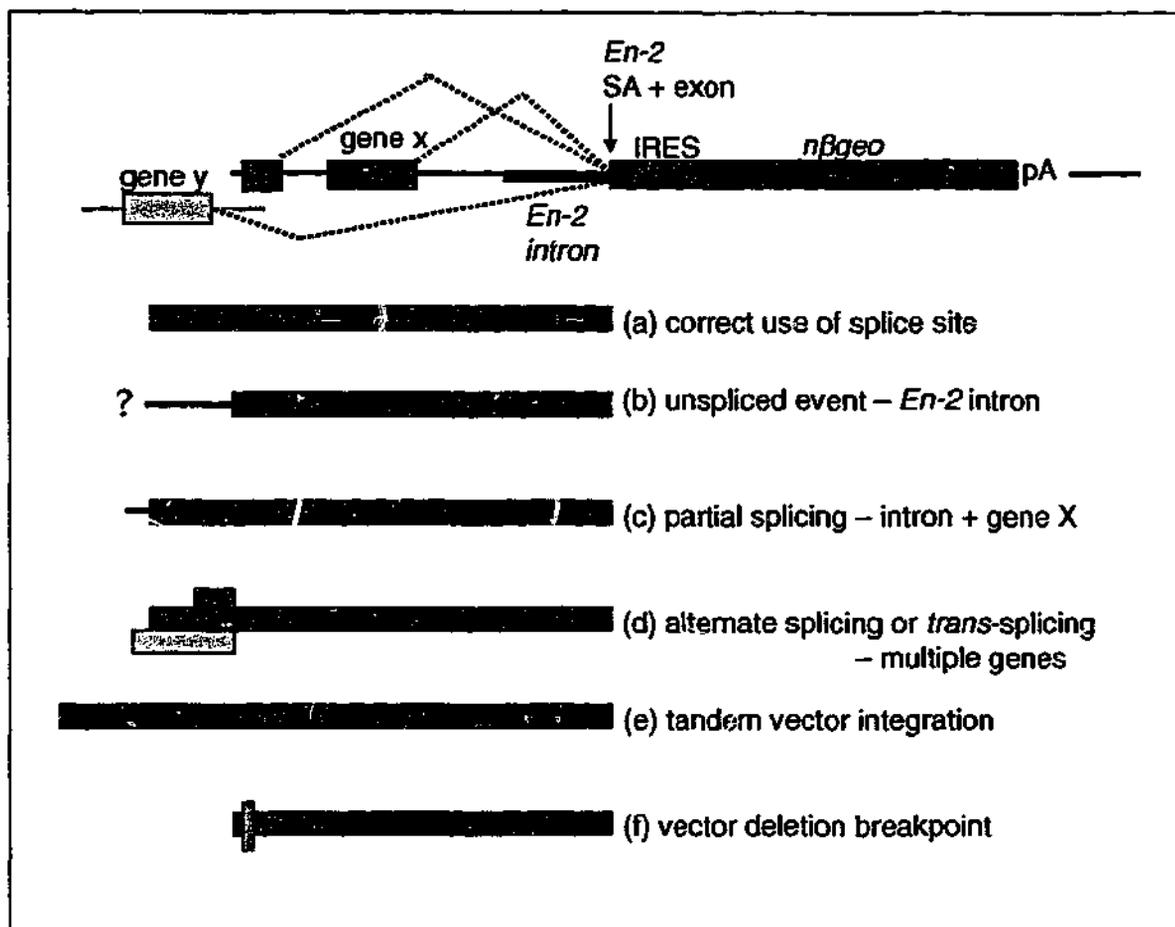
Figure 4.5: Errors associated with gene trap integration events in ES cells. (a) Insertion of a gene trap vector, exemplified here by the pG2 ι.8 (OPT/Nuclear) IRES*βgeo* gene trap vector, into the intron of an endogenous gene X (red line = intron, red boxes = exons) is predicted to generate a nuclear localizing *βgeo* (*nβgeo*) fusion mRNA transcript through the use of the *En-2* splice acceptor site (SA) contained in the vector (thick black line = *En-2* intron, black box = *En-2* exon). (b) Where splicing at the *En-2* splice consensus sequence does not occur, 5'-RACE products will retain vector intron sequence upstream of *En-2* exon sequence, while the endogenous gene coding sequence may lie much further upstream. (c) Where partial splicing occurs, the splice donor from gene X is used, but *En-2* intron sequence remains superimposed on the endogenous sequence. (d) Alternate splicing or *trans*-splicing from the SA may occur across two or more endogenous exons or genes, respectively (yellow box = gene Y), yielding multiple 5' sequence reads. (e) At electroporation, tandem copies of the vector may be inserted into the genome, and as well, may make use of cryptic donor splice sites within the most 5' intronic vector sequence. (f) The occurrence of deletion breakpoints in the gene trap vector (and possibly in tandem vector copies) prior to or during integration, is likely to explain the inability to amplify 5'-RACE products from some gene trap mRNA transcripts.

matched 43 bp of the same 3'-end *En-2* vector exon sequence, followed by 5'-end *En-2* exon sequence excluding the splice acceptor sequence, as previously mentioned. For the remaining cell line (GT#50), a short 5' read of 30 bp with no homology upstream of a smear and the correct splice acceptor sequence was difficult to interpret. This outcome was deemed resultant of either a vector deletion event or possibly an insertion in a very short 5'-untranslated region, considering a fusion transcript size greater than the vector contribution had not been detected for this clone by Northern blot analysis (Figure 4.1).

A summary of outcomes for gene trap integration events, as determined by direct sequence analysis of 5'-RACE products from fusion transcripts generated with the optimized IRES*βgeo* gene trap vectors, is given in Table 4.2.

Outcome for direct sequencing analyses	# of events
Single readable sequence ladder 5' of *En-2* SA and exon sequence	10
Multiple sequence products 5' of *En-2* SA and exon sequence	3
Incorrect use of the splice acceptor site • Unspliced (*En-2* intron sequence 5' of *En-2* SA and exon sequence) • Partially spliced (*En-2* intron & endogenous sequence 5' of *En-2* SA and exon sequence)	7 • 2 • 5
Splicing within vector sequences +/- ? vector deletion events	3
5'-RACE products non-amplifiable	5
Total # of gene trap ES cell lines analysed	28

Table 4.2: Sequence outcome for gene trap integration events generated with an optimized IRESβgeo vector. For 28 gene trap events analysed, mRNA fusion transcripts for 10 (36%) of ES cell lines yielded correctly spliced single 5' sequence reads by direct sequencing of 5'-RACE products. The remaining ES cell lines yielded either multiple sequence products upstream of the vector splice acceptor site (SA), or sequence ladders suggestive of incorrect use of the SA, or errors in vector integration. No sequence information was obtained for five cell lines, for which 5'-RACE products could not be amplified, probably as a result of deletion breakpoints in the fusion transcripts.
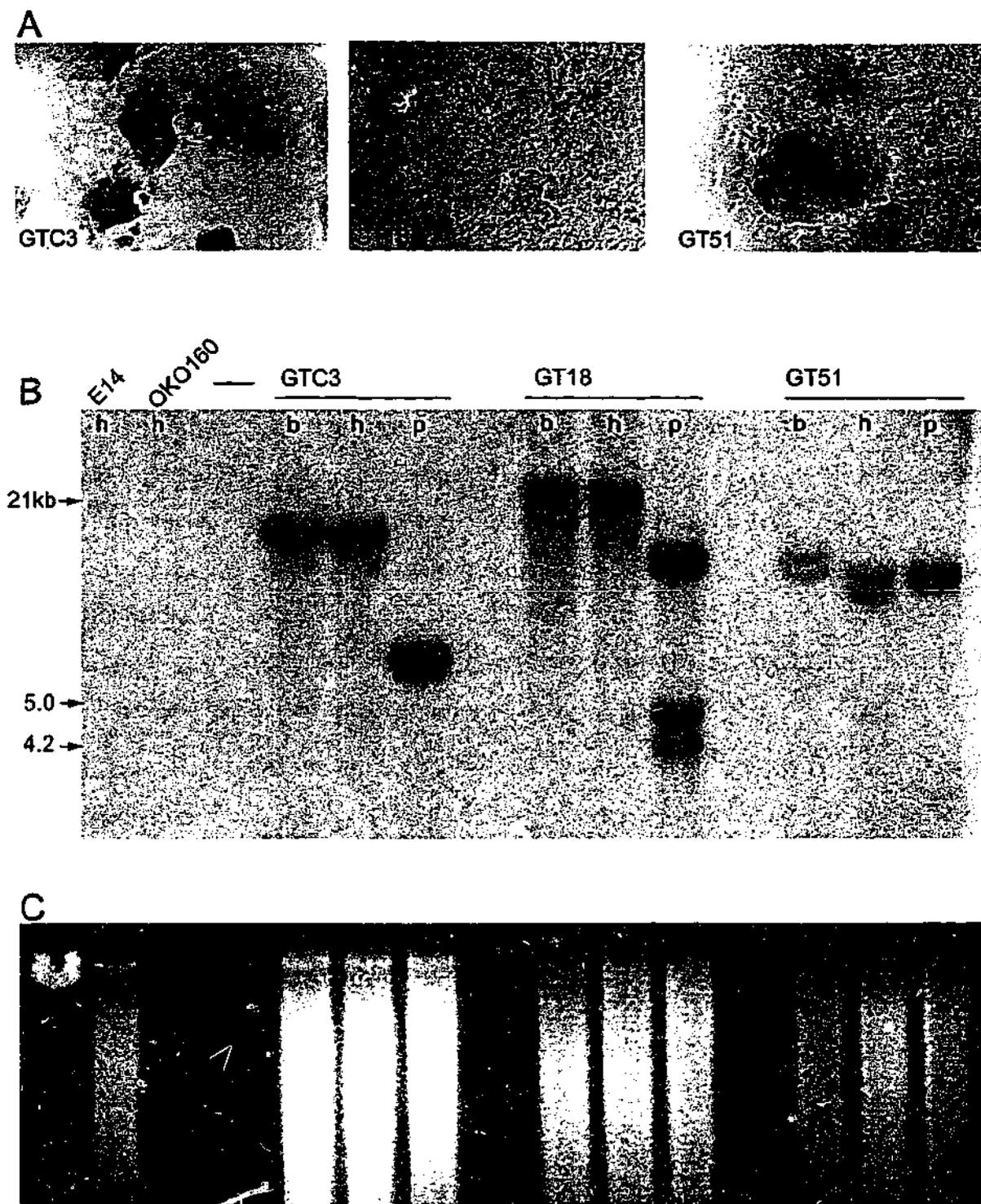
Figure 4.6: Genomic DNA analysis of a potential hotspot for gene trap integration in ES cells. (A) Three gene trap ES clones (GT#'s C3, 18, 21) showing the same ES cell-restricted *lacZ* expression profiles *in vitro* (G418 300 µg/ml; photographed under phase-contrast optics, 100x), yielded 90 bp of identical endogenous sequence 5' of the splice site in mRNA fusion transcripts. (B) For each clone, 15 µg of genomic DNA was digested individually with the enzymes *Bgl* II (b), *Hind* III (h) and *Pst* I (p), and a Southern blot hybridized with the *Neo* probe. *Hind* III digested genomic DNA samples from the E14 and OKO160 cell lines were run as negative and positive controls, respectively. (C) Electrophoresed DNA was stained with ethidium bromide for a loading control. Autoradiographic exposure was for 4 days at -80°C with intensifying screens. No hybridization signal was seen for the OKO160 control due to a gel loading error (see Figure 5.5 for positive *Neo* hybridization of a *Pst* I restriction fragment for the same OKO160 DNA). Hybridization results confirm that for each clone, the gene trap has integrated at a different genomic locus for the same *RBP-Jκ* pseudogene sequence.

4.5 GENOMIC ANALYSIS FOR A POTENTIAL GENE TRAPPING HOTSPOT IN ES CELLS

Three ES cell lines (GT#'s C3, 18, 21) arising from independent transfections showed the same ES cell-restricted *in vitro* expression profiles (Figure 4.6A), and yielded identical sequences upstream of the *En-2* splice site in mRNA fusion transcripts (section 4.4), suggesting a possible "hotspot" for gene trap integration in ES cells. To investigate if the three independent gene trap integration events were in identical genomic insertion sites or alternate sites within a common intron, genomic DNA prepared from each of the three COB gene trap ES cell lines (section 2.2.2c) was digested individually with the enzymes *Bgl* II, *Hind* III and *Pst* I, all of which cut the gene trap vector (Figure 3.2), while control E14 and OKO160 ES cell genomic DNA (section 2.1.5) was digested with *Hind* III (sections 2.2.1c & 2.2.3b). A Southern blot was hybridized (sections 2.2.3b & 2.2.6a) with a radiolabelled probe *"Neo"*, corresponding to approximately 800 bp of *neo* sequence isolated from the pGT1.8 (OPT/Nuclear) IRESβgeo vector (sections 2.1.2 & 2.2.5a).

Hybridization results shown in Figure 4.6B demonstrate clearly that hybridizing restriction fragments are not the same for each clone, confirming that the gene trap vector has integrated into a different genomic locus in each gene trap cell line. A more detailed analysis would be required, however, to confirm the anticipated integration of the gene trap vector within a common intron. A hybridization signal for the OKO160 genomic DNA is not seen on this blot due to a gel loading error (Figure 4.6C), although *Neo* hybridization of a *Pst* I restriction fragment for the same DNA sample (see Figure 5.5) confirmed a single copy βgeo insertion in the OKO160 cell line (B. Zevnik, Centre for Genome Research, Edinburgh, UK; *pers. comm.*). As expected, the E14 control did not show a hybridization signal.

4.6 DISCUSSION

Various methods based on 5'-RACE (Frohman *et al.*, 1988) have been previously used to identify entrapped genes via DNA sequencing of cloned cDNA sequences (Skarnes *et al.*, 1992; Takeuchi *et al.*, 1995; Chen Z., 1996) and more recently, direct solid-phase sequencing of 5'-RACE products (Skarnes *et al.*, 1995; Townley *et al.*, 1997) arising from the gene trap fusion transcripts. This chapter describes results obtained for 28 gene trap cell lines analysed by a direct sequencing approach (Townley *et al.*, 1997),

adapted for the generation of 5'-RACE products from the pGT1.8 (OPT) IRES*βgeo* and pGT1.8 (OPT/Nuclear) IRES*βgeo* gene trap vectors employed in this study (Figure 4.2). Use of this strategy has efficiently eliminated clones that represent erroneous gene trap events, and identified 10 ES cell lines (36% of those analysed) yielding single readable sequences upstream of correctly used vector splice acceptor sequence. This result is consistent with a 37% yield following direct sequence analysis of 153 gene trap lines generated with a secretory trap vector (Townley *et al.*, 1997) and indicates that the inclusion of an IRES and nuclear localization signal in the vectors used here, has not had a detrimental effect on the number of clones providing useful sequence information.

Database analysis of cDNA sequences for each of the 10 ES cell lines identified 5 gene trap integrations in previously uncharacterized genes. Three of the cDNA sequences for these integrations did, however, show some homologies for ESTs identified in embryonic tissues. Interestingly, for 5 sequences identified as gene trap integrations in known genes, 3 separate clones provided sequence information suggesting integration within the same gene.

The cell line GT#54 was of most interest for homology of the 5' trapped gene sequence with ESTs identified in preimplantation stage mouse embryos. A 126 bp sequence upstream of the splice site showed 100% homology for an EST [Genbank accession number (gb) AA465972] generated from a mouse 2-cell cDNA library (Rothstein *et al.*, 1992), indicating that the trapped endogenous gene is expressed at a time point during preimplantation development that is earlier than that for the ICM-derived ES cells. The gene trap insertion occurs 298 bp downstream in this 560 bp EST sequence, and the 5' and 3' ends of the GT#54 sequence also overlapped with complete homology for two other ESTs from the same 2-cell library (gb AA794905, AA414462). The latter 3' EST and the trapped EST also aligned with a 767 bp theoretical clone (TC38526) in the TIGR mouse gene index database. A 92% homology was also identified for 84 bp of the GT#54 sequence with an EST (gbAA683807) from a mouse blastocyst cDNA library (Rothstein *et al.*, 1992).

The GT#37 line yielded 151 bp of novel sequence with 93% homology for three ESTs (gb AA637056, AI594135, AI503178) from a mouse blastocyst cDNA library (Rothstein *et al.*, 1992), and for an EST (gb C89031) for an early mouse blastocyst cDNA clone (Sasaki *et al.*, 1998). The sequence also showed lesser homologies (85-

91%) with ESTs (gb AA221852, AI466714, AA052459, W11513) for total mouse embryo 12.5 dpc and 19.5 dpc cDNA clones (Soares *et al.*, 1994; Bonaldo M. *et al.*, 1996). Interestingly, the first 60 bp of the GT#37 sequence upstream of the splice site identified a 93% homology with the 5'-end of a mouse putative REX-2 mRNA (gb AF051348), and might suggest that the remaining trapped sequence is in the 5'-untranslated region for this gene. That this putative REX-2 mRNA shows reduced expression with retinoic acid induction of differentation in F9 EC cells (Faria *et al.*, 1998) would correlate well with the ES cell-restricted expression pattern identified for the trapped gene.

Other EST homologies of 80-90% for the GT#37 sequence indicated that the gene, or related genes, may be expressed elsewhere in both embryonic and adult tissues. These homologies included a lymph node EST (gb AA288451), male thymus ESTs (gb AA199488, AA183755), and mammary gland ESTs (gb AA475046, AI181274, AA636887, AA981461) from a mouse 4 week cDNA library (Soares *et al.*, 1994; Bonaldo M. *et al.*, 1996); a thymus EST (gb AW61762) from a mouse 3 week cDNA library (X. Ren and L. Stubbs, unpublished); a gene trap cDNA clone (gb X98138) for a mouse Bruce 4 ES cell line (Holzschu *et al.*, 1997); a 7.5 dpc EST (gb AA474825) from a mouse embryonic library (Harrison *et al.*, 1995), and two kidney ESTs (gb AI789447, AI317531) from mouse adult cDNA libraries (Suzuki *et al.*, 1997). Several of these ESTs are reported to resemble a mouse zinc finger protein Zfp-51 mRNA (gb X74855), and also align with a number of TIGR database theoretical mouse clones – TC113816 (427 bp), TC113817 (294 bp), TC134038 (681 bp), TC152798 (237 bp), TC29768 (586 bp) and TC118201 (276 bp).

Sequence analysis for the GT#36 line indicated the use of a cryptic splice donor site within the gene trap vector sequence, and suggests possible integration into an exon rather than an intron in this case. Splicing has occurred at position 32 in the vector *En-2* intron buffer sequence, and has been previously observed to occur at this same site when the *En-2* splice acceptor sequence has been in the context of other gene trap vectors and the apparent lack of an endogenous splice donor site has suggested insertion into an exon (McClive *et al.*, 1998). Analysis of the 118 bp of novel exon sequence upstream of this 32 bp intron sequence revealed highest homologies with fetal mouse 13.5-14.5 dpc ESTs (89-90%, gb AA041682, AA030393) and 19.5 dpc ESTs (90-97%, gb W07943, W30639, AA036361), as well as a mammary gland EST (92%, gb

AI173310), (Soares *et al.*, 1994; Bonaldo M. *et al.*, 1996). All of these ESTs overlapped and aligned with an 836 bp theoretical clone (TC145714) in the TIGR mouse gene index database.

The novel 131 bp sequence read upstream of the splice site in the GT#52 cell line does not show any homology matches for EST sequences. Of interest, however, is that the first 102 bp of the sequence showed 82% homology with a gene for the ABCC4 protein (gb NM_005845), which belongs to the human ATP-binding cassette (ABC) superfamily of proteins that transport various molecules across extra- and intra-cellular membranes (Allikmets *et al.*, 1996). The ABCC4 protein is a ubiquitously expressed member of the multidrug resistance-associated proteins (MRP) subfamily, and may play a role as an organic anion pump in cellular detoxification (Lee *et al.*, 1998). However, as the 102nd bp does not represent the 5'-end of the ABCC4 gene, the significance of this homology for the trapped gene is unclear.

For the remaining uncharacterized gene trapped in the GT#28 cell line, recent EST database entries identified 100% homology for the entire 101 bp of 5' sequence with two mouse germinal B-cell ESTs (gb AW826115, AW825750), which correspond to a hypothetical 39.8 kD protein (Soares *et al.*, 1994; Bonaldo M. *et al.*, 1996). The sequence also shows complete homology with a mouse adult liver EST (gb AW260613) and mouse adult kidney ESTs (gb AW044996, AI527892), (Suzuki *et al.*, 1997). All of these EST clones overlap with a TIGR mouse database 1071 bp theoretical clone (TC108176). An EST derived from *Bos Taurus* fetal ovary (gb AV618319), (Takeda *et al.*, 1998), and from a human lung tumor cDNA clone (gb BF110702), (NCI - Cancer Genome Anatomy Project), also show homologies of 92% and 89%, repectively, for 98 bp of the GT#28 sequence. Prior to the above identities, the GT#28 sequence had only shown 70% homology with two overlapping ESTs (gb AI110152, AA942367) from *Drosophila* adult head and embryo cDNA libraries (D. Harvey, L. Hong, M. Evans-Holm *et al.*, unpublished).

From the sequence ladders for the five correctly spliced gene trap fusion transcripts identified as known genes, it is interesting to note that three of the gene trap cell lines (GT#C3, 18 and 51) appear to harbor gene trap integrations in the same gene. This result is suggestive of a possible gene trapping hotspot within the 5'-flanking sequence of a processed pseudogene for the transcription factor *RBP-Jκ*. Each cell line resulted from independent transfection experiments, for both the cytoplasmic and nuclear

132

localizing optimized IRES*βgeo* constructs, and each was confirmed as a true clone by differential Southern blot hybridization (Figure 4.6).

The fusion sequences show 100% homology with nucleotides 159-248 of the *RBP-Jκ* pseudogene (gb M81871). This sequence is at the 5'-end of the reported pseudogene, which has identical cDNA coding sequence for exons 2-11 of the functional gene for the RBP-Jκ binding protein (*RBP-Jκ*, gb X17459), with exon 2 commencing at nt 248 in the related pseudogene (Matsunami *et al.*, 1989; Kawaichi *et al.*, 1992). Incidentally, the functional *RBP-Jκ* gene generates a major and minor transcript from different transcription initiation sites, and by different useage of splice signals in the same exon 1 sequence (gb X58337, Kawaichi *et al.*, 1992). The 5'-end sequence upstream of exon 2 in the pseudogene shows no homology to intron 2 of the functional gene, nor to any other flanking sequences of the functional and other related genes. There are two other reported pseudogenes, termed *Pseudo1* and *Pseudo2*, which have high homology with exons 1-7 and 1-6 of the functional gene, respectively, but are considered to be non-functional due to the presence of many scattered stop codons (Kawaichi *et al.*, 1992). It is of interest that the first 158 bp of reported sequence for the *RBP-Jκ* pseudogene shows high homology to mouse LINE-1 (L1) elements (see following discussion), and may explain the variation in genomic organization for the *RBP-Jκ* gene.

While the sequence repeatedly trapped in this study may represent an alternate exon 1 for a functional gene, the authors of the original report describing the genomic analysis of the *RBP-Jκ* gene concluded that this processed type gene was most likely to represent a non-functional pseudogene, as they did not detect any clones without the *RBP-Jκ* 1st exon following intensive screening of libraries (Kawaichi *et al.*, 1992). Curiously, sequence homologous to nucleotides 203-248 for the *RBP-Jκ* pseudogene has been previously identified following transfection of CGR-8 ES cells with the pGT1.8TM secretory trap vector (Skarnes *et al.*, 1995), and analysis for 13 lines of interest in a screen for novel secretory molecules involved in embryonic development (J. Brennan, Centre for Genome Research, Edinburgh, UK; *pers. comm.*). No further information regarding the 5'-flanking regions of the *RBP-Jκ* gene family has presently come to light (T. Honjo, Kyoto University, Japan; *pers. comm.*).

The incidence for trapping this sequence with two vector constructs in independent experiments for this study, and with a different construct again in another laboratory,

does suggest that this 5'-flanking *RBP-Jκ* sequence represents a genomic hotspot for gene trapping in ES cells. The RNA transcript sizes and *in vitro* expression data for the three clones are indistinguishable (Table 4.1), while Southern analysis has revealed gene trap insertions at different genomic loci, presumably within a common intron (Figure 4.6). While promoter elements are usually missing from pseudogenes, clearly the sequence in this 5' region is being transcribed and in this study represents the repeated trapping of a gene that is under the control of an ES cell-specific promoter. In this context, it is interesting that immunostaining for the RBP-Jκ protein is localized in the nuclei of ES and F9 EC cells, and is lost upon differentiation (Sakai *et al.*, 1995). Although the protein is widely expressed in mouse tissues and cells, mostly in the nuclei, it appears to exist in a free form and a chromatin-bound form, with an apparent loss of the free form in differentiated cells (Sakai *et al.*, 1995; reviewed by Honjo, 1996).

The *suppressor of hairless* [*Su(H)*] *Drosophila* homolog for the highly conserved *RBP-Jκ* transcription factor has been shown to play a role downstream of the *Notch-Delta* signal transduction pathway in early peripheral nervous system development (Furukawa *et al.*, 1994; reviewed by Honjo, 1996). The RBP-Jκ protein, which does not possess any known DNA binding motif, has also been identified as the cellular protein with which the Epstein-Barr virus nuclear antigen-2 (EBNA-2) binds to target its DNA responsive elements in the process of B cell immortalization (Grossman *et al.*, 1994; Henkel *et al.*, 1994; Waltzer *et al.*, 1994; Zimber-Strobl *et al.*, 1994). Mutation analysis for *RBP-Jκ* has demonstrated that homozygote null mouse ES cells do not contain the encoded protein but are capable of normal growth, while null mutant mice show embryonic lethality before 10.5 dpc, with severe growth retardation seen at 8.5 dpc (Oka *et al.*, 1995). The mutants show defects in somatogenesis similar to those observed for *Notch 1* (Conlon *et al.*, 1995), and *Delta* homolog *Dll 1* (Hrabe de Angelis *et al.*, 1997) mutant mice. It is now known that the RBP-Jκ/Su(H) protein interacts directly with the intracellular domain of the transmembrane Notch receptor to provide a signal transduction pathway between the cell surface and the nucleus, although the precise mechanisms for processing of the Notch receptor presently remain unclear (reviewed by Honjo, 1996; Lendahl, 1998). Given the role for the Notch family of receptors in mediating cell-cell communication and inducing lateral specification of equivalent cells (reviewed by Greenwald, 1998), an interesting role in developmental

134

cell fate determination thus remains a possibility for the ES cell-restricted *RBP-Jκ-*related sequence repeatedly trapped in this study.

For the remaining two clones, sequence analysis showed that cell line GT#26 had 126 bp of *lacZ* sequence upstream of the splice site, suggestive of tandem vector insertion and use of a splice donor site within the *lacZ* sequence. The identification of upstream endogenous sequence for this clone will therefore require genomic cloning and screening of cDNA libraries. The remaining cell line, GT#8, provided a 96 bp sequence upstream of the splice site which shows high homology to many mouse ESTs containing L1 repetetive elements.

L1 elements constitute a large superfamily of mammalian retrotransposons that have been replicating and evolving in mammals for more than 100 million years, and now compose 20% or more of the DNA of some mammals. These autonomously replicating elements, which lack the long terminal repeated (LTR) sequences typical of their retroviral counterparts, contain regulatory signals and encode for two proteins – one an RNA-binding protein, and the other a protein with endonuclease and reverse transcriptase activities, which likely functions as an integrase-replicase (reviewed by Furano, 2000). The human and mouse X chromosomes are now known to be rich in L1 elements, compared with the autosomes, with the greatest density occurring at the region containing the X inactivation center (Bailey *et al.*, 2000; reviewed by Lyon, 2000). Apart from presumed evolutionary and structural roles in the genome, a recent hypothesis proposes that L1 repeat sequences act as booster elements to aid the onset of spread for *Xist* RNA at the time when X chromosome inactivation first occurs in the embryo, and the accumulation of stable *Xist* RNA begins (Lyon, 2000). If L1 elements are indeed part of the mechanism for X chromosome inactivation, trapping of these repeat sequences in the GT#8 clone is a feasible finding, given that at the time of ES cell isolation both X chromosomes are still functioning in the epiblast (Monk and Harper, 1979), and may even be related to the repeated incidence of gene trap integration in the *RBP-J$_k$* pseudogene sequence. It is perhaps worth noting that L1 sequences have been previously isolated from 2 of 42 clones generated with an IRES*βgeo* vector (Chowdhury *et al.*, 1997).

The direct sequencing strategy used in this study has enabled cDNA sequencing of mRNA fusion transcript sequences 5' to the gene trap transcribed sequence. Analysis of the fusion transcript sequence has enabled the identification of the trapped gene and the

elimination of those gene trap insertions which show errors in use of the splice acceptor site or vector integration (Tables 4.1 & 4.2). The different types of events that have been observed to occur for gene trapping insertions are illustrated in Figure 4.5. 5'-RACE products yielding sequence information were obtained for 23 (82%) of the 28 lines, with those for the remaining 5 lines failing to hybridize with vector *En-2* exon sequence nested 5' of the downstream gene-specific PCR primers (Figure 4.3). The inability to amplify these lines is a likely result of deletion breakpoints at positions downstream of the primer sites used in the original 5'-RACE reactions (Townley *et al.*, 1997; J. Brennan, Centre for Genome Research, Edinburgh, UK; *pers. comm.*), (Figure 4.5f). For 3 of the gene trap cell lines generated in this study, deletion events have possibly occurred upstream of the PCR primer site, as sequence product was obtained but appeared to represent splicing or deletions within tandem vector integrations (Figure 4.5e). The incidence of such events is likely to be particular to the vector used, and may also be a consequence of degradation of DNA during electroporation (Niwa *et al.*, 1993).

The results presented here show that 7 of the 28 (25%) gene trap insertions analysed have resulted in fusion transcripts containing vector-specific intron sequence, and as such have failed to make correct use of the vector's splice acceptor sequence. Two of these insertions contained intron sequence alone (Figure 4.5b), while five partially spliced events contained intron sequence superimposed on another distinct sequence (Figure 4.5c). The significant presence of unspliced fusion transcripts has been previously observed in three gene trap studies using the pGT1.8TM secretory trap vector. For fusion transcripts analysed in these studies 5/11 (46%), (Skarnes *et al.*, 1995), 44/153 (29%), (Townley *et al.*, 1997) and 7/24 (29%), (J. Brennan, Centre for Genome Research, Edinburgh, UK; *pers. comm.*) hybridized with intron sequences of the vector. While inefficient splicing of the vector to upstream exons could be an inherent property of some endogenous loci, it is also predicted that integration of the vector into exons of genes or into non-polymerase II transcription units would generate intron-containing fusion transcripts (Townley *et al.*, 1997). Insertion into an exon of a polymerase II gene is expected to create competition between the vector splice acceptor and that of the disrupted exon. Where the vector inserts into rRNA genes, the resulting uncapped RNA polymerase I transcripts are thought to act as poor substrates for *cis*-splicing, and as such would leave the vector splice acceptor free to interact with splice

donors of other gene products which may be utilized in *trans* (J. Brennan, Centre for Genome Research, Edinburgh, UK; *pers. comm.*). While the incidence of intron-containing events generated with the vectors in this study compares favorably against the reports cited above, these results again highlight the advantage of first assessing hybridization of fusion transcripts with vector intron sequence prior to 5'-RACE and sequence analyses.

For 3 (11%) of the 28 insertions, the vector splice acceptor had been properly used, but yielded multiple sequence reads upstream of the splice site. This type of event may be the result of either alternate splicing of the gene trap splice acceptor to multiple upstream splice donors of the trapped gene, or even *trans*-splicing of the gene trap splice acceptor to entirely different genes (Figure 4.5d). As discussed above, insertion into rRNA genes may favour the occurrence of *trans*-splicing events. It is less likely, but possible, that the vector has inserted into a site upstream of the first splice donor of the endogenous gene. The resulting fusion transcript, in which the splicing element most 5' would be an acceptor, might be unstable and lead to *cis*- or *trans*-splicing with other splice donors (Chowdhury *et al.*, 1997). It also remains possible that the multiple products generated for the three lines in my study have arisen from the selection of mixed colonies, although the tertiary *in vitro* screen (section 3.4.5) is likely to have eliminated this possibility.

Multiple sequence products were also identified in the large-scale analysis for secretory gene trap events (Townley *et al.*, 1997), for 14/153 (9%) of the insertions analysed. Another large gene trap study has reported 5/55 (9%) such events, 4 of these arising from 42 insertions (9.5%) generated by an IRESβgeo vector (Chowdhury *et al.*, 1997). Analysis of 28 gene trap events generated with the pGT1.8geo vector (Skarnes *et al.*, 1995) revealed 2 multiple sequence events (7%), one for which two gene trap insertions segregated upon breeding, and the other apparently representative of splicing to several unrelated splice donors (Voss *et al.*, 1998a). The frequency of multiple sequence events following gene trapping is likely to depend on individual constructs, and does not appear to have seriously affected the usefulness of the vectors used in the above studies. Results for the optimized IRESβgeo vectors presented here are comparable with those cited elsewhere and due to time constraints, it was not considered worthwhile to pursue identification of the trapped sequences for the three clones showing multiple sequence reads.

In summary, my results show that the inclusion of an IRES element and a nuclear localization signal in the gene trap vectors used in this study, has allowed the efficient generation of gene trap events yielding useful sequence information regarding the endogenous genes. Analysis of gene trap clones by a direct sequencing strategy has enabled the identification of 10 ES cell lines yielding correctly spliced single mRNA fusion transcripts for putative ES cell-restricted genes, including a possible gene trapping hotspot in an *RBP-Jκ* -related gene. Five ES cell lines harboring gene trap integrations in previously uncharacterized genes were identified. These cell lines have subsequently been denoted as COB28, COB36, COB37, COB52 and COB54 for further characterization of the endogenous genes.

Chapter Five

DETAILED ANALYSIS OF GENE TRAP INSERTIONS WITHIN FIVE NOVEL ES CELL-SPECIFIC GENES

5.1 INTRODUCTION

The *in vitro* screening strategy from which the gene trap clones of interest were identified (Chapter three), made use of the inherent ability for blastocyst-derived mouse ES cells to differentiate spontaneously *in vitro*, and in both 2-dimensional and aggregate cultures to give rise to cell types reflecting the lineage-specific differentiation programs found in the embryo (Doetschman *et al.*, 1985; Keller, 1995; Nishikawa *et al.*, 1998; Kawasaki *et al.*, 2000; Yamashita *et al.*, 2000). *In vivo*, mouse ES cells are able to differentiate readily into multiple tissue types in teratocarcinomas formed following injection to vascularised ectopic sites in adult mice (Evans and Kaufman, 1981; Martin 1981), and when reintroduced to host blastocysts can combine with the ICM to form chimaeras of both ES cell- and host-derived cell lineages (Bradley *et al.*, 1984). Providing a normal karyotype and pluripotential status has been maintained in culture, ES cells can participate fully in normal embryogenesis, colonizing all tissues including the germ line in a developing host embryo (Bradley *et al.*, 1984; Beddington and Robertson, 1989). Such ES cell colonization is principally of the fetus and extraembryonic mesoderm, but is also seen at a lower frequency in the primitive endodermal and trophoblastic extraembryonic tissues, resembling the behaviour of early

139

ICM cells rather than embryonic epiblast (Gardner and Rossant, 1979; Beddington and Roberston, 1989). Supporting this is the common observation of cells akin to extraembryonic endoderm and large trophoblastic giant cells in the outgrowths of clonally derived ES cell cultures (Doetschman et al., 1985; Robertson, 1987). While prolonged passage in culture may diminish the proportion of ES cells retaining full potential, completely ES cell-derived viable mice have been produced following aggregation of ES cells cultured up to passage 24 with a developmentally compromised tetraploid embryo (Nagy et al., 1993). For this normal development of the embryo proper to occur, the ES cells are required to have faithfully propagated a normal status for imprinted genes in culture. The tetraploid embryo is only able to contribute to the extraembryonic tissues (Dean et al., 1998).

The gene trap and in vitro screening approach used in this study pre-selected five gene trap ES cell lines of interest - COB28, COB36, COB37, COB52 and COB54 – based on their in vitro ES cell-restricted expression profile, abating the need to generate and screen large numbers of chimaeric embryos for restricted in vivo patterns of reporter gene expression. While in some cases the developmental lacZ expression pattern has appeared more restricted in its distribution than for the expression of the endogenous gene (Skarnes et al., 1992; Deng and Behringer, 1995; Voss et al., 1998a), the results of several gene trap screening studies suggest that in most cases reporter gene expression reflects the activity of the trapped gene (for example, Skarnes et al., 1992, 1995; Takeuchi et al, 1995; Chen J. et al., 1996; Torres et al., 1997; Voss et al., 1998a). Although not all cell types are highly represented in vitro, a strong enrichment for restricted in vivo patterns following in vitro pre-selection protocols has been previously reported (Forrester et al., 1996; Baker et al., 1997; Bonaldo P. et al., 1998; Voss et al., 1998a). The in vivo expression profiles for the five ES cell-restricted genes identified in this study are therefore of much interest.

This chapter describes results of extended molecular analyses for the five cell lines, and the introduction of each cell line into host blastocysts for the generation of chimaeric animals. A detailed analysis and discussion is provided for the endogenous gene integration site of the COB54 cell line.
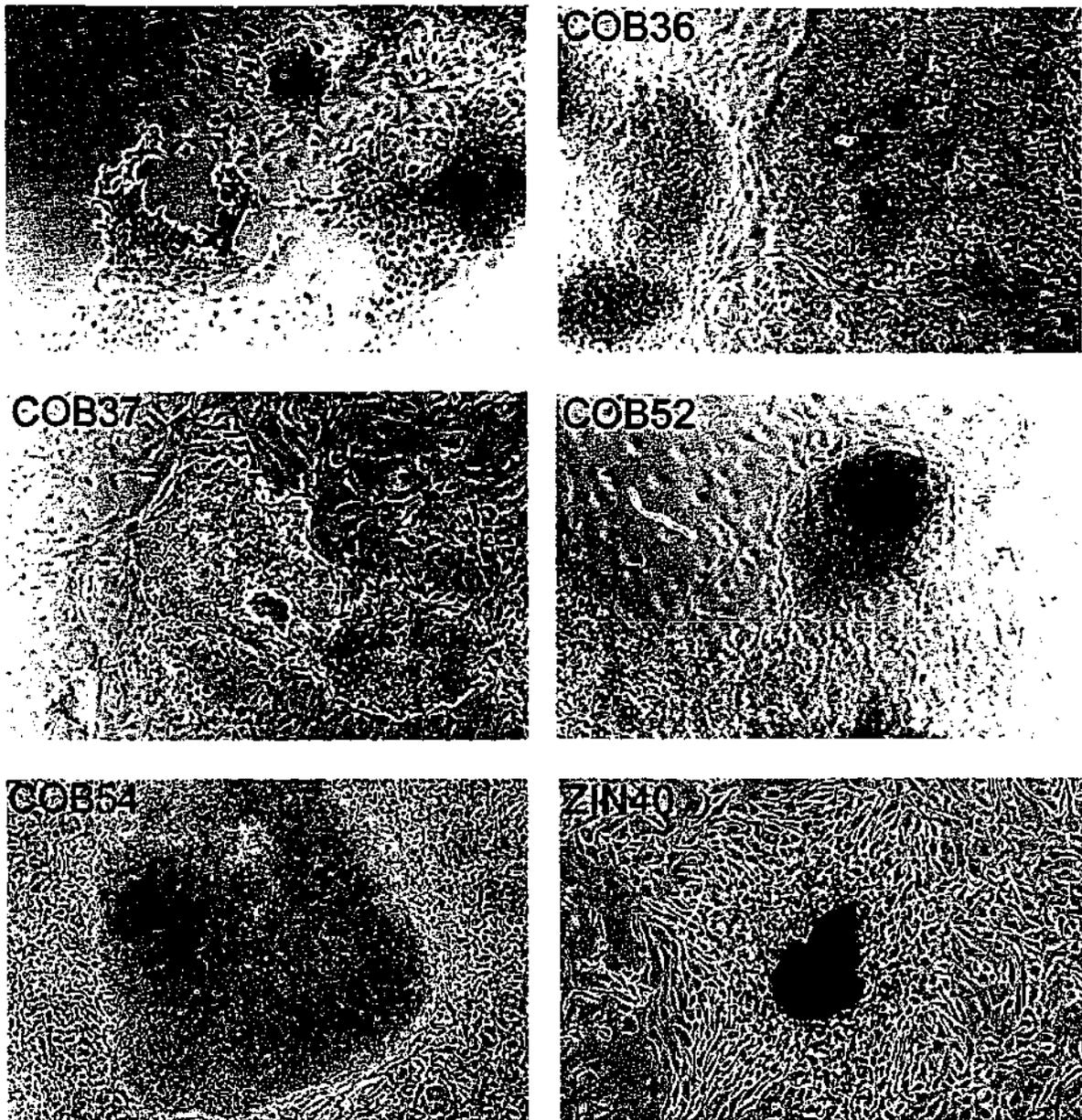
Figure 5.1: ES cell-restricted lacZ expression profiles for five novel gene trap insertions. COB28, COB36, COB37, COB52 and COB54 gene trap ES cell lines were grown at clonal density for 4-6 days in the presence of LIF, and without selection. Following X-gal staining, βgal activity is detected only in the undifferentiated ES cell population and not in differentiating cell types, for each of the lines. The pattern of detectable ES cell βgal activity varies for each cell line with the lowest expression seen for the COB28 cell line, and the strongest for the COB52 cell line. Excepting the COB28 cell line, *lacZ* expression is also observed to be heterogeneous across the ES cell population. For the ZIN40 control gene trap ES cell line grown under the same conditions, marked and homogeneous *lacZ* expression is seen in both ES and differentiating cells. Photographed under phase-contrast optics, 100x.
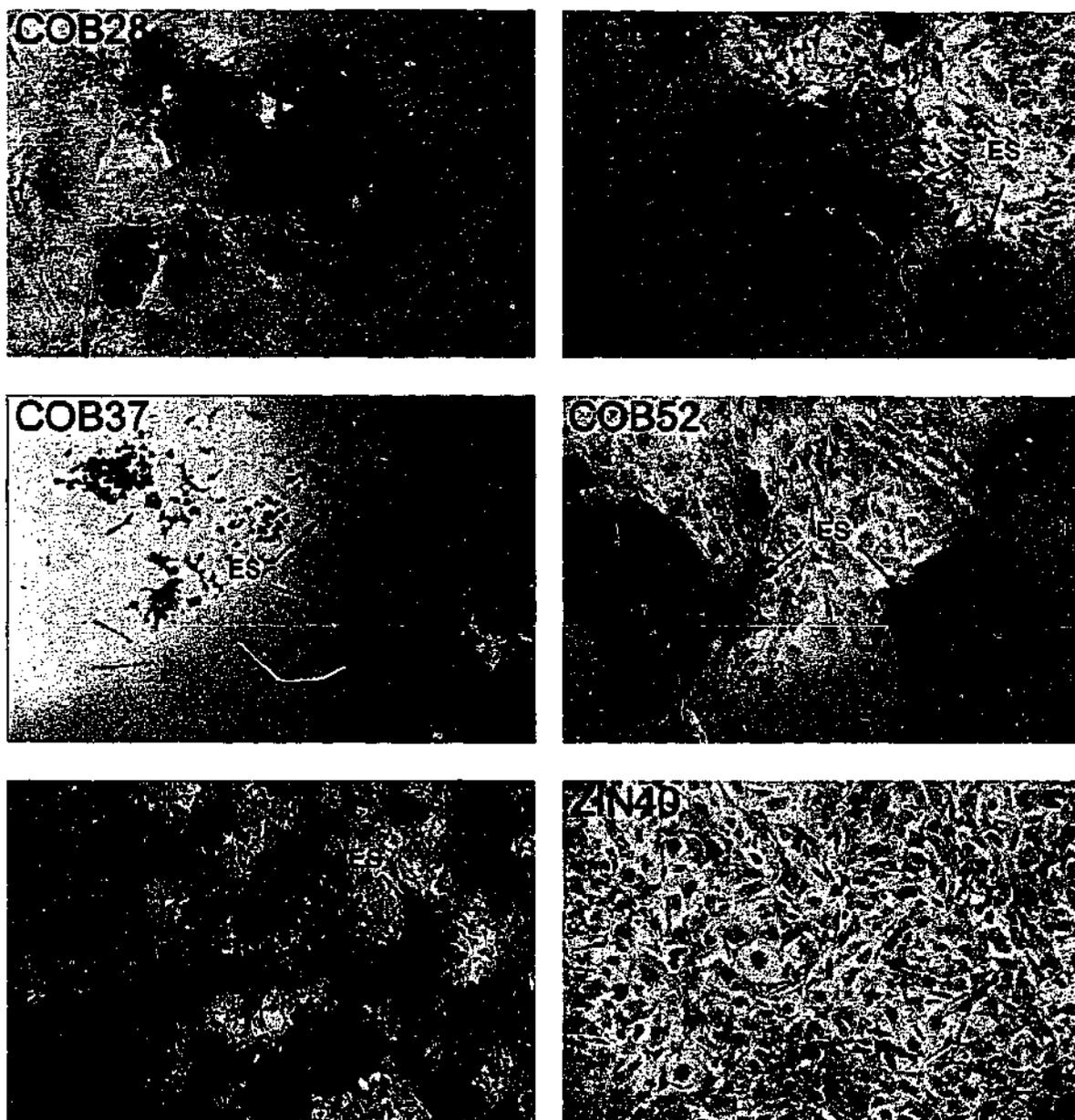
Figure 5.2: Morphology of cells recovered from aggregate culture by selection for neo expression at ES cell-restricted loci. For the gene trap cell lines COB28, COB36, COB37, COB52 and COB54, ES cell cultures were recovered from embryoid bodies grown in the absence of LIF and in the presence of G418 at 300 µg/ml for 10 days. Embryoid bodies were trypsinized and replated cells are shown following Leishman's staining after a further 2 days culture in the presence of LIF and without selection. The COB54 cell line displays the most robust recovery of undifferentiated ES cells (ES, arrows) under these conditions, while the COB37 line displays the recovery of ES cells at a much lower density, and the remaining three lines show equivalent and strong recovery of ES cell cultures. There is no recovery of ES cells from aggregate selection cultures grown for the negative control ZIN40 cell line. Photographed under phase-contrast optics, 100x.

5.2 IN VITRO EXPRESSION PROFILE FOR FIVE NOVEL ES CELL-SPECIFIC mRNAs

Analysis of *in vitro lacZ* and *neo* expression profiles revealed variations for the five cell lines harboring gene trap insertions in ES cell-restricted genes (Figures 5.1 & 5.2). In monolayer cultures grown without selection (section 3.4.4), the mildest intensity for βgal activity following X-gal staining was observed for the COB36 line, while the COB52 line showed the most intense staining pattern (Figure 5.1). The COB54 cell line displayed a particularly sharp downregulation in *lacZ* expression at the borders of ES cell colonies. In the COB28 line, the *lacZ* expression pattern was observed to be homogeneous, while the other lines displayed varying degrees of heterogeneity across subsets of undifferentiated cells in ES cell colonies (Figure 5.1). As a general observation the COB28 cell line showed less robust growth than the other lines. In the embryoid body cultures selecting for *neo* expression at an ES cell-restricted locus (section 3.4.5), the recovery of ES cell cultures in the absence of LIF was most marked for the COB54 cell line (Figure 5.2). The EBs for this line displayed strong growth and yielded larger numbers of ES cells at replating compared with the other lines. This result indicated strong expression of the trapped gene in undifferentiated ES cells and a lack of expression in differentiated progeny. By comparison, recovery of lower numbers of ES cells from smaller EBs for the COB37 line may represent a gene with low level expression in ES cells (Figure 5.2).

Time constraints of this study limited the number of cell lines which could be analysed in detail. COB54, which shows strong ES cell-restricted expression and complete homology of the trapped endogenous sequence with an EST from a mouse 2-cell cDNA library (section 4.6) was selected for detailed analysis.

5.3 EXTENDED SEQUENCE ANALYSES

5.3.1 Cloning 5'-RACE products from five gene trap fusion transcripts

For the purpose of generating DNA probes specific to the five novel ES cell-specific genes and to extend the sequence information obtained by the direct sequencing protocol (Table 4.1), non-biotinylated 5'-RACE products generated from the gene trap vector for the five cell lines (section 2.2.8a) were end-polished with *Pfu* DNA polymerase and blunt end ligated into the vector *pCR-BLUNT* for subsequent transformation of *One shot* Top 10 *E. coli* cells (sections 2.1.1 & 2.2.1), and automated

141

sequence analyses of cloned products (section 2.2.9b). Bacterial colony blots (section 2.2.3a) were hybridized with the *En41* probe (sections 2.1.2, & 2.2.6a), and 7 independent 5'-RACE clones positive for *En-2* exon sequence were picked for each gene trap ES cell line. Thermal cycle sequencing reactions were performed on DNA from at least 4 clones for each line (section 2.2.9b) using the *En-2* exon sequencing primer previously used for direct sequencing (oligo 5, Figure 4.2), (section 2.1.3). In a few cases, sequencing reactions were also run using a primer corresponding to the T7 promoter sequence of the cloning vector.

Cloning of 5'-RACE products was successful for four of the gene trap ES cell lines, and extended direct sequence reads by 32 bp (COB37) to 616 bp (COB54), (Table 5.1). In each case the 5' anchor primer sequence was readable from sequence reactions, indicating the limit of the 5'-RACE product.

Sequencing products obtained from 7 plasmid DNA preparations for the COB28 cell line either contained very short 5'-RACE products, or spurious results from poor automated sequencing read-outs. As all of these clones had shown the presence of recombinant DNA products following gel electrophoresis (data not shown), this cell line possibly harbors a gene trap insertion in the 5' region of an endogenous gene.

Cell line (GT#)	Direct sequence (bp)	#Clones analysed	Cloned sequence (bp)
COB28	101	7	N/A
COB36	118	7	263
COB37	151	4	183
COB52	131	7	232
COB54	126	6	742

Table 5.1: Sequence extension of cloned 5'-RACE products for five gene trap fusion transcripts. 5'-RACE products generated from the gene trap insertions for five ES cell lines were cloned into the *pCR BLUNT* vector for automated analysis of sequencing reaction products. All clones excepting those from COB28 products yielded extra sequence information 5' to the vector splice site when compared with the direct sequencing protocol. The longest clear sequence read for each cell line is given, with the anchor sequence at the 5' end being identified in each case.

Cloned sequences for the four cell lines provided confirmation of novel cDNAs and the EST homologies previously identified by direct sequence analysis (section 4.6). The extra sequence obtained from the COB37 and COB52 cell lines did not identify any further significant database homologies.

For the COB36 cell line, the extra 145 bp of 5' sequence identified homologies of 97 and 98% with two ESTs (gb AI551549, AA437965) from a mouse embryo 11.5 dpc cDNA library (M.S. Ko and X. Wang, unpublished). These ESTs overlapped to form a 431 bp theoretical clone (TC122152) in the TIGR mouse gene index database. The extended COB36 sequence also identified EST homologies of 93 and 96% for two mouse mammary tumor cDNAs (gb BE308393, BE377112), (NIH – Mammalian Gene Collection), and 86-88% for a *Bos taurus* placental EST (gb BF044543), (Soares *et al.*, 1994; Bonaldo M. *et al.*, 1996), a *Bos taurus* adipocyte cell line EST (gb AV667030), (Takeda *et al.*, 1998), a human colon EST (Dias *et al.*, 2000), a human meningioma EST (gb AI460004), (M. Schiller, unpublished), and a *Sus scrofa* embryo EST (gb BE013969), (Rohrer *et al.*, 1996).

The 616 bp 5' extension for the COB54 entrapped gene, sequenced from both ends of the cloned insert, confirmed 100% sequence homologies for the entire EST harboring the gene trap insertion (gb AA465972) and its 5' overlapping EST (gb AA794905), both from a mouse 2-cell library (Rothstein *et al.*, 1992). The cloned sequence also bridged these EST sequences to a further 5' EST from the same library (gb AA465973), as well as a 5' overlapping EST (gb AA571680) from a mouse blastocyst cDNA library (Rothstein *et al.*, 1992), with complete homologies except at one nt in the former EST. Database analysis identified 100% overlapping homology for an EST (gb AA426752) downstream of the trapped EST, and generated from a mouse 8-cell cDNA library (Rothstein *et al.*, 1992). This EST sequence information allowed the database construction of a COB54 contig cDNA of approximately 1.7 kb, with the gene trap inserted approximately 1.2 kb downstream of the 5' cap (Figure 5.3). The 1.2 kb of 5' endogenous gene sequence was consistent with Northern blot results (Figure 4.1) which indicated that the predicted COB54 cDNA sequence encompassed the majority of the endogenous mRNA sequence 5' to the insertion site.

```
   1 cttacttgcc tgtccaagat ctgttggaat ctgcttctac agaagaccag ctgaaacaaa
  61 tagcttcgtg ggactgagca caactactag attcttggac ttccgttcac agctgccaat
 121 tgttgggagt acaataatgg aggagtcgga attggagatt tttagaagta agtttgttag
 181 aggctcatct gtcacgaagc agcatgcctg gcgaaaccag cacagcgaga agcgttgctc
 241 ttcctccatc agttctatat ccctggacag aatgccatcg gaaatcttgg tgaagatact
 301 ttcttacttg gatgcggtga ccttggtgtg cattggatgt gtgagcagac gcttttatca
 361 tttggctgat gacaatctta tttgggtcag gaagtacgca gctgcattTa gatcaaaaag
 421 atcacgttgg aaagctactt cagtggagga aacagccaca agtctgagct tgctgtcagt
 481 ttgggataaa gaagatggat actggaagaa agaatatatt acaaagcaga tctcatctgt
 541 gaaagcagcc ctcaccaaca gcctcagtcc tgtcaaacgc cgcacaagcc ttccttcgaa
 601 aaccaaagag tccctcagaa tatctggctt aggttggaca atcatcttaa gagaagccag
 661 tggCaaagaa cacatcatgc agcattcgaa tctttccgta aatgacaact ctgtcactgt
 721 tttttggcat gacaaaaatt ggccacatgt agacacgttg tccaccctgg atttgtatgg
 781 tgccacacca atttttatgg agcagtataa aggccctaac acaagttgtc cacgatggct
 841 gtctttaatt gaaaagtacg atctgagtaa tttacgcaag tctgctatga ttggctgcga
 901 cagacatgtt cgggtattct gtgtaaatcc tggcctcctg gtggggctgt ggcaggagaa
 961 tggtggacta gcttttgtca tggcaaatat tcattcccat ggccttttcg agagaagcat
1021 aatgggctca gacactattc cctatacatt gcctcccgac actacatttg tggataacta
1081 cccagactca atgacctttt atggagataa aggctttcag ctgcatatcg acattcatgg
1141 cagtaagact tacttcctgt gtagcacctt ccacaatctc ttctgcagga gagcgggcat
1201 taacaatgga tatgtgaagt tcttgatgat aaacttaaaa aataacagag aacacctacc
1261 tcttgttgga aaagttggcc ttgaatggag aactgactgt ttaaatggcc gtattgagag
1321 ttgcattgta gtggatatga ccttgctgga tgaggacaag aagcccatct ggtatgtgag
1381 ttctccagtg tgcttgagat ctgcctgcct tcctgatttc ccgcagccgg cttactcttt
1441 cgagtacatg gacagcgtag gaggagtgtg cgcagaccta gggtggtttg aaaataccga
1501 tgaatacttc attgtcagac tggacattta cctcagtgta gcaaaattac aacaatggtt
1561 tgggaggcaa taaatgctga gttagcagta gggagtcttg ttattagtaa gctgtttgtt
1621 ttttacaact ttgtttttat tgaaagttaa aataaagcat atttgtggt
```

Figure 5.3: cDNA sequence for the native COB54 gene. The 1669 nt COB54 contig sequence is derived from a 5'-RACE clone generated from the gene trap fusion transcript, and overlapping database EST sequences of 99-100% homologies – gb AA571680 (nts 1-542), AA45973 (nts 234-782), AA794905 (nts 829-1110), AA465972 (nts 896-1455), AA426752 (nts 1251-1662). The site of the gene trap insertion is marked with a solid arrow. Forward primers [COB54F1, COB54F2, COB54F3] and reverse primers [COB54RT, COB54R1, COB54R2, COB54R3] were used for 5'- and 3'-RACE and PCR analyses, confirming endogenous sequence from nt 197 to the 3' end. Differences to the predicted contig sequence were identified for a cytosine substitution at nt 664 and an additional thymidine at nt 409 (bold capitals). The *Cob54Race* probe isolated from a 5'-RACE clone extends 730 bp upstream from nt 1188, while the 817 bp *Cob54Pcr* probe (see section 6.3.9) is derived from PCR amplification between the COB54F1 and COB54R1 primers.

5.3.2 5'- and 3'-RACE cloning of the endogenous COB54 gene

To establish a full length cDNA sequence for the gene identified by the COB54 gene trap insertion, 5'- and 3'-RACE products were generated from wild type E14 ES cell poly A$^+$ mRNA using primers specific to the COB54 sequence. Sequences for these primers, described below, are listed in section 2.1.3, and their positions in the COB54 contig sequence are shown in Figure 5.3.

5'-RACE products were generated as previously described (section 2.2.8a, Figure 4.2) except that 1st strand cDNA was synthesized with a different RT primer, COB54RT, beginning at 1065 bp in the contig sequence. Two rounds of PCR were performed using the 5'-RACE oligo 2 anchor primer (Figure 4.2) and two reverse primers, COB54R1 and COB54R2, nested at 1029 and 941 bp in the contig sequence, respectively. 3'-RACE products were generated using the 5'-RACE oligo 2-dT anchor primer (Figure 4.2) to initiate first strand synthesis from the native poly A tail of the COB54 message (section 2.2.8b). PCR amplifications were performed using the 5'-RACE anchor primer (oligo 2, Figure 4.2) and two forward primers, COB54F2 and COB54F3, nested 1046 and 1112 bp in the contig sequence, respectively. All PCR products were ligated with the T-overhang *pGEM-T* cloning vector for subsequent transformation of JM109 *E. coli* cells (sections 2.1.1 & 2.2.1). Bacterial colony blots (section 2.2.3) were hybridized with end-labelled gene-specific primers corresponding to the second round of PCR in each case (sections 2.2.5b & 2.2.6a). Thermal cycle sequencing reactions were performed on DNA from at least 2 positive clones picked for each analysis, using the SP6 and T7 *pGEM-T* promoter primers to sequence in both directions (section 2.2.9b).

Automated sequence analyses (section 2.2.9b) confirmed the COB54 contig sequence from 197 bp through to the 3' end, with the insertion of an extra thymidine residue at 409 bp, and 7 extra nucleotides following the 3'-end EST preceding a poly A tail. One nucleotide substitution was identified at 664 bp in the EST sequence (gb AA465973), (Figure 5.3). An attempt to complete confirmation of the most 5' EST (gb AA571680) by 5'-RACE from an RT primer beginning at 232 bp in the sequence, unfortunately resulted in erroneous amplification between anchor primer sequences.

Conceptual translation of the corrected 1669 bp contig sequence identified a 524 amino acid protein, with an open reading frame of 479 residues extending from the first

Fbx15

```
cttacttgcctgtccaagatctgttggaatctgcttctacagaagaccagctgaaacaaat
  L  L  A  C  P  R  S  V  G  I  C  F  Y  R  R  P  A  E  T  N    20
agcttcgtgggactgagcacaactactagattcttggacttccgttcacagctgccaatt
  S  F  V  G  L  S  T  T  T  R  F  L  D  F  R  S  Q  L  P  I    40
gttgggagtacaataatggaggagtcggaattggagattttttagaagtaagtttgttaga
  V  G  S  T  I  M  E  E  S  E  L  E  I  F  R  S  K  F  V  R    60
ggctcatctgtcacgaagcagcatgcctggcgaaaccagcacagcgagaagcgttgctct
  G  S  S  V  T  K  Q  H  A  W  R  N  Q  H  S  E  K  R  C  S    80
tcctccatcagttctatatccctggacagaatgccatcggaaatcttggtgaagatactt
  S  S  I  S  S  I  S  L  D  R  M  P  S  E  I  L  V  K  I  L   100
tcttacttggatgcggtgaccttggtgtgcattggatgtgtgagcagacgcttttatcat
  S  Y  L  D  A  V  T  L  V  C  I  G  C  V  S  R  R  F  Y  H   120
ttggctgatgacaatcttatttggtcaggaagtacgcagctgcatttagatcaaaaaga
  L  A  D  D  N  L  I  W  V  R  K  Y  A  A  A  F  R  S  K  R   140
tcacgttggaaagctacttcagtggaggaaacagccacaagtctgagcttgctgtcagtt
  S  R  W  K  A  T  S  V  E  E  T  A  T  S  L  S  L  L  S  V   160
tgggataaagaagatggatactggaagaaagaatatattacaaagcagatctcatctgtg
  W  D  K  E  D  G  Y  W  K  K  E  Y  I  T  K  Q  I  S  S  V   180
aaagcagccctcaccaacagcctcagtcctgtcaaacgccgcacaagccttccttcgaaa
  K  A  A  L  T  N  S  L  S  P  V  K  R  R  T  S  L  P  S  K   200
accaaagagtccctcagaatatctggcttaggttggacaatcatcttaagagaagccagt
  T  K  E  S  L  R  I  S  G  L  G  W  T  I  I  L  R  E  A  S   220
ggcaaagaacacatcatgcagcattcgaatctttccgtaaatgacaactctgtcactgtt
  G  K  E  H  I  M  Q  H  S  N  L  S  V  N  D  N  S  V  T  V   240
ttttggcatgacaaaaattggccacatgtagacacgttgtccaccciggatttgtatggt
  F  W  H  D  K  N  W  P  H  V  D  T  L  S  T  L  D  L  Y  G   260
gccacaccaattttttatggagcagtataaaggccctaacacaagttgtccacgatggctg
  A  T  P  I  F  M  E  Q  Y  K  G  P  N  T  S  C  P  R  W  L   280
tctttaattgaaaagtacgatctgagtaatttacgcaagtctgctatgattggctgcgac
  S  L  I  E  K  Y  D  L  S  N  I  R  K  S  A  M  I  G  C  D   300
agacatgttcgggtattctgtgtaaatcctggcctcctggtggggctgtggcaggagaat
  R  H  V  R  V  F  C  V  N  P  G  L  L  V  G  L  W  Q  E  N   320
ggtggactagcttttgtcatggcaaatattcattcccatggccttttcgagagaagcata
  G  G  L  A  F  V  M  A  N  I  H  S  H  G  L  F  E  R  S  I   340
atgggctcagacactattccctatacattgcctcccgacactacatttgtggataactac
  M  G  S  D  T  I  P  Y  T  L  P  P  D  T  T  F  V  D  N  Y   360
ccagactcaatgacctttttatggagataaaggctttcagctgcatatcgacattcatggc
  P  D  S  M  T  F  Y  G  D  K  G  F  Q  L  H  I  D  I  H  G   380
agtaagacttacttcctgtgtagcaccttccacaatctcttctgcaggagacggggcatt
  S  K  T  Y  F  L  C  S  T  F  H  N  L  F  C  R  R  A  G  I   400
aacaatggatatgtgaagttcttgatgataaacttaaaaaataacagagaacacctacct
  N  N  G  Y  V  K  F  L  M  I  N  L  K  N  N  R  E  H  L  P   420
cttgttggaaaagttggccttgaatggagaactgactgtttaaatggccgtattgagagt
  L  V  G  K  V  G  L  E  W  R  T  D  C  L  N  G  R  I  E  S   440
tgcattgtagtggatatgaccttgctggatgaggacaagaagcccatctggtatgtgagt
  C  I  V  V  D  M  T  L  L  D  E  D  K  K  P  I  W  Y  V  S   460
tctccagtgtgcttgagatctgcctgccttcctgatttcccgcagccggcttactctttc
  S  P  V  C  L  R  S  A  C  L  P  D  F  P  Q  P  A  Y  S  F   480
gagtacatggacagcgtaggaggagtgtgcgtagacctagggtggtttgaaaataccgat
  E  Y  M  D  S  V  G  G  V  C  A  D  L  G  W  F  E  N  T  D   500
gaatacttcattgtcagactggacatttacctcagtgtagcaaaattacaacaatggttt
  E  Y  F  I  V  R  L  D  I  Y  L  S  V  A  K  L  Q  Q  W  F   520
gggaggcaataatgctgagttagcagtagggagtcttgttattagtaagctgtttgttt
  G  R  Q  -  M  L  S  -  Q  -  G  V  L  L  L  V  S  C  L  F   540
tttacaactttgttttttattgaaagttaaaaataaagcatatttgtggt
  F  T  T  L  F  L  L  K  V  K  I  K  H  I  C  G
```

Figure 5.4: Conceptual translation of the native COB54 F-box mRNA. Conceptual translation of the 1669 bp COB54 cDNA sequence (upper line) predicts a 524 amino acid protein (lower line), with an open reading frame of 479 residues extending from the first methionine codon (red box). The gene trap insertion site (blue arrow), stop codon (black box) and polyadenylation signal (black underline) are shown. The protein contains a 38 amino acid F-box motif (green box) commencing at the second methionine codon. The recently identified *Fbx15* partial cDNA sequence commences at nt 21 in the COB54 contig sequence (*Fbx15* black arrow) and differs at one residue in the open reading frame, #181 (in bold), by substitution of the lysine codon (aaa) with an arginine codon (aga).

methionine codon at 137 bp in the contig to a stop codon and poly A signal (Figure 5.4). Analysis of the predicted amino acid sequence (section 2.2.9c) identified a 38 amino acid F-box motif starting at the second methionine codon at 272 bp (Figure 5.4). More recently, a 1654 bp partial coding sequence has been reported (Winston *et al.*, 1999a), with complete homology for the COB54 contig open reading frame except for the substitution of a lysine codon (K) at the 181st residue with an arginine (R) codon (Figure 5.4). Termed *Fbxl5* (gb AF176530), this mouse cDNA belongs to a novel family of mammalian F-box proteins which are involved in targeting specific substrate proteins for rapid degradation by the ubiquitin proteolysis pathway (Cenciarelli *et al.*, 1999; Winston *et al.*, 1999a).

5.3.3 Confirmation of a gene trap insertion in the *Fbx15* gene

The COB54 gene trap insertion site was further confirmed by PCR analysis of wild type and COB54 gene trap ES cell line RNA. Briefly, E14 and COB36 ES cell control poly A$^+$ mRNA samples were reverse transcribed from the COB54RT primer, and a PCR product generated between the reverse COB54R1 primer and a new upstream primer COB54F1, beginning at 213 bp in the contig sequence (Figure 5.3). COB54 ES cell poly A$^+$ mRNA was reverse transcribed from the IRES oligo 1 primer (Figure 4.2), and a PCR product generated between the *En-2* oligo 3 primer (Figure 4.2) and the same forward COB54F1 primer (sections 2.2.7a & b; T$_a$ = 56°C). The expected PCR product sizes of 0.8 kb for wild type *Fbx15* cDNA and 1.1 kb for the COB54 fusion transcript were observed by gel electrophoresis (data not shown). Sequence analysis for cloned E14 and COB54 products (as for section 5.3.2), confirmed the RACE generated sequence and insertion of the gene trap in the same sequence. As shown in Figure 5.4, the gene trap insertion disrupts translation of the 470 amino acid open reading frame at the 353rd residue.

5.4 SOUTHERN ANALYSIS OF GENE TRAP INSERTIONS IN FIVE ES CELL-SPECIFIC GENES

Genomic DNA prepared from each of the five COB ES cell lines (section 2.2.2c) was separately digested with the enzymes *Bgl* II, *Hind* III and *Pst* I, while control E14 and OKO160 ES cell DNA (section 2.1.5) was digested with *Pst* I (sections 2.2.1c & 2.2.3b). A Southern blot was hybridized (sections 2.2.3b & 2.2.6a) with the *Neo* probe (section 2.1.2) to analyse copy number of the insert DNA (Figure 3.2). Hybridization
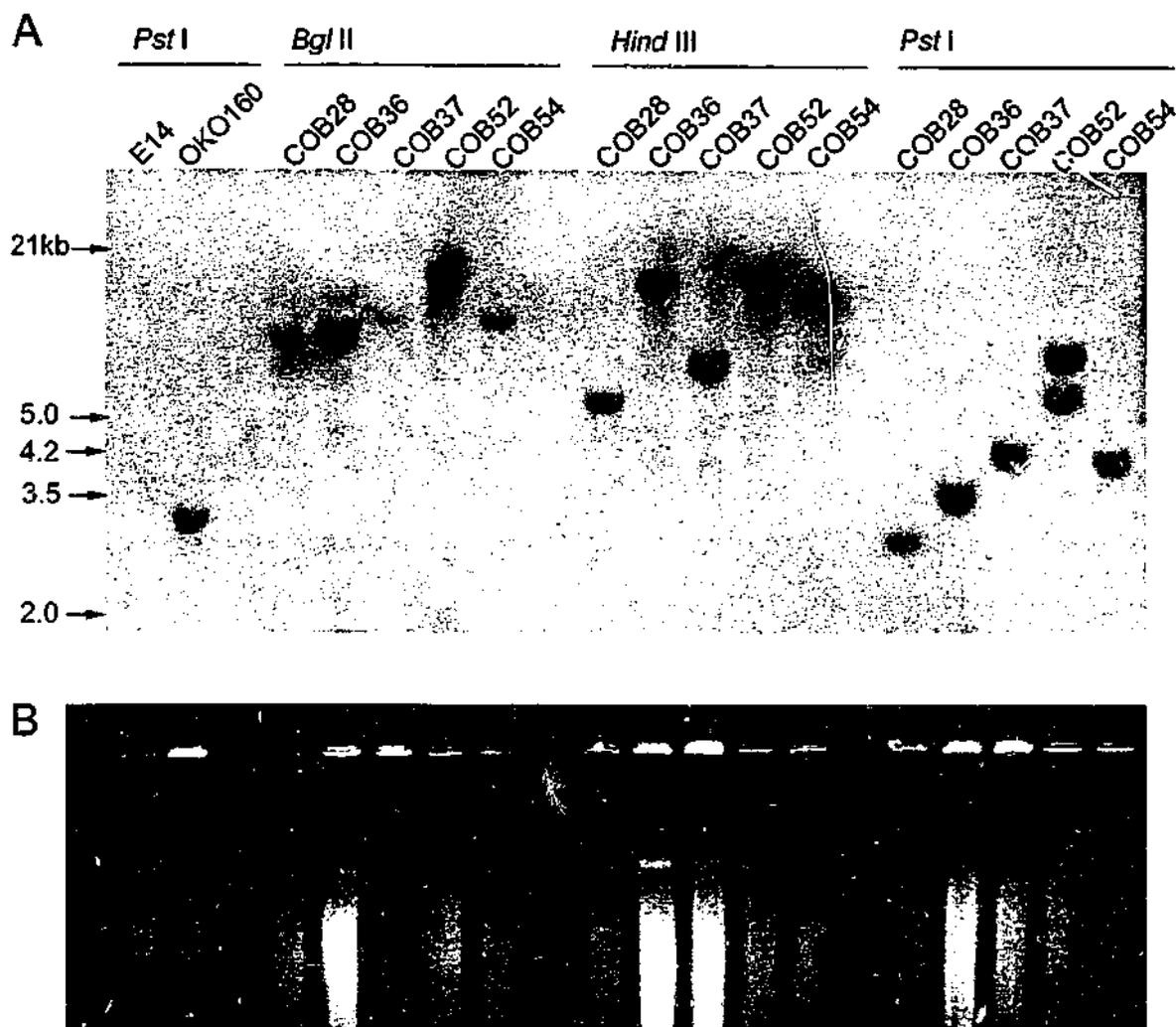
Figure 5.5: *Genomic DNA analysis of gene trap insertions in five novel ES cell-specific genes.*
(A) 15 μg of genomic DNA for each of the five gene trap ES cell lines (COB28, COB36, COB37, COB52, COB54) was digested individually with the enzymes *Bgl* II, *Hind* III and *Pst* I, and a Southern blot hybridized with the *Neo* probe. *Pst* I digested genomic DNA samples from the E14 and OKO160 cell lines were run as negative and positive controls, respectively. (B) Electrophoresed samples were stained with ethidium bromide for a DNA loading control. Autoradiographic exposure was for 4 days at -80°C with intensifying screens. Hybridization was positive for a known single copy *neo* insertion in the OKO160 control line, and at different genomic loci for the five gene trap insertions. Signal intensities comparative to gel loadings suggests a possible multiple copy insertion for the COB52 cell line, and a single copy insertion event for all other COB ES cell lines.

results shown in Figure 5.5 confirm a gene trap integration into different endogenous loci for the five cell lines. The OKO160 control confirms a single positive *neo* band (B. Zevnik, Centre for Genome Research, Edinburgh, UK; *pers. comm.*) and the E14 control DNA is negative as expected. Results for the five COB ES cell lines suggest a single copy insertion event in each cell line except for COB52, which may harbor a multiple insertion event.

5.5 CHIMAERA GENERATION

To test the ability of gene trap COB ES cell lines to contribute to chimaera formation *in vivo*, ES cells were injected into C57Bl/6J blastocysts (sections 2.5.1 & 2.5.2) and transferred to pseudopregnant recipient females (section 2.5.3). Early passage frozen gene trap ES cell lines were thawed for culture (sections 2.3.1 & 2.3.3) and used within an average of 7 passages for injections, with a fresh change of media 2 h prior to trypsinizing cells (section 2.3.1). All cell lines tested negative for mycoplasma by PCR analysis (data not shown). Transferred embryos for each line were allowed to develop to term with the resulting number of viable offspring varying from 0-26.5% of the number of injected blastocysts transferred (Table 5.2).

GT cell line	#embryo transfers	#blastocysts transferred	#pups born	#pups weaned (% of b/c)	#chimaeras (male/female)
COB28	6	69	0	0 (0%)	0
COB36	4	48	6	5 (10.4%)	2 male
COB37	3	40	5	4 (10.0%)	1 male/1 female
COB52	7	91	?5 or 6	4 (4.4%)	2 male (1 x FD)
COB54	3	34	11	9 (26.5%)	6 male/1 female

Table 5.2: Chimaera generation for ES cell lines harboring gene trap integrations in ES cell-restricted genes. 129/Ola-derived ES cells for five gene trap cell lines were injected into C57Bl/6J host blastocysts and transferred to pseudopregnant recipients from C57Bl/6J x CBA matings. Pregnancies were allowed to continue to term and viable offspring weaned at 3 weeks of age. No offspring were born following transfers for COB28 injected embryos. ES cell contribution to chimaeric animals was assessed by the presence of agouti, white or sandy hairs amongst black hairs of the host, and in all cases well exceeded 50% (see Figure 5.6). One COB52 male chimaera, small in size and with a high degree of coat colour chimaerism, was found dead (FD) around 5 weeks of age. For chimaeras generated from all cell lines 85% displayed a male phenotype.

Figure 5.6: Chimaeras generated by blastocyst injection of gene trap ES cells. For five gene trap 129/Ola-derived ES cell lines, ES cells were injected into host C57Bl/6J blastocysts, and chimaeric offspring identified by the presence of light coat colour regions (sandy, agouti, white) amongst the black hairs contributed by the host. Viable offspring and chimaeric males resulted for the lines COB36, COB37, COB52 and COB54, but not COB28. All chimaeras showed a very high degree of ES cell contribution as judged by coat colour.

Despite the higher number of COB28 injected embryo transfers, no offspring were generated for this cell line. One resorption site was observed following examination of a recipient mother which had failed to produced a pregnancy following the transfer of 10 COB28 injected blastocysts, suggesting the possible inability of these embryos to implant in the uterus.

At least one viable male chimaera was generated from each of the remaining four cell lines (Table 5.2) and in each case showed a very high degree of ES cell contribution as judged by coat colour (section 2.5.4), (Figure 5.6). This is generally considered to be a good indicator for potential colonization of the germ cells, as is the proportion of phenotypic males born (Bradley et al., 1984). A high contribution of XY ES cells will convert an XX host blastocyst to a male phenotype, and as XX germ cells do not form viable sperm, all sperm from these males will be ES cell-derived and result in the complete transmission of the ES cell genome. The parental E14 and E14Tg2a ES cell lines used in this study are of a male karyotype (section 2.1.5), and indeed the sex ratio is skewed to the male phenotype for 11/13 (85%) chimaeric animals generated from all gene trap cell lines (Table 5.2). Of two male chimaeras generated for the COB52 line, one was found dead at about 5 weeks of age, was small in size and showed a high degree of coat colour chimaerism.

5.6 GERMLINE TRANSMISSION OF THE COB54 INSERTION

Adult coat colour chimaeras were test bred with C57Bl/6J mice to assess for germline transmission. An agouti coat colour in offspring was indicative of ES cell germline transmission, whereas offspring derived from host C57Bl/6J germ cells displayed a black coat colour (section 2.5.4). Table 5.3 summarizes results for the test breeding of each adult chimaeric animal.

Judging by the agouti coat colour of offspring, germline transmission was achieved only from male chimaeras for the COB54 gene trap ES cell clone, and in 82% (123/150) of offspring derived from 21 litters. Three male chimaeras for this line produced 100% (95/95) ES cell-derived agouti offspring, while one yielded 74% (23/31) and the other a lesser contribution of 21% (5/24). The 6th COB54 male chimaera did not mate successfully. This animal was small in size and became unwell, displaying an enlarged thymus upon examination. As anticipated, the female chimaera for this line yielded a small litter of black coat colour pups and breeding was discontinued.

For the COB36 line, the 1st male chimaera produced 14 pups from 3 litters, with 4 of these displaying a black coat colour and the remaining 10 found dead prior to coat development. The 2nd COB36 male chimaera, which displayed an occlusion, achieved only one successful mating over a four month breeding period, yielding a single pup which again died prior to coat development. The single male COB37 chimaera although displaying a genital prolapse, produced 32 pups from 6 litters. Despite the high degree of coat colour chimaerism for this male (Figure 5.6), all offspring were derived from the host germ cells. The female COB37 chimaera showed poor fertility, but yielded 11 pups from 2 litters, all with a black coat colour. The surviving COB52 male chimaera (section 5.5) displayed a very high degree of coat colour chimaerism and failed to produce any offspring over a four month breeding period.

GT line	chimaera	#litters	#pups	#agouti	germline	comments
COB36	male 1	3	14	?	no	11 FD/? coat colour
	male 2	1	1	?	no	malocclusion/poor breeding
COB37	female 1	2	11	0	no	poor breeding
	male 1	6	32	0	no	prolapse
COB52	male 1	0	-	-	no	poor breeding
COB54	female 1	1	3	0	no	
	male 1	4	24	5	yes	
	male 2	5	38	38	yes	
	male 3	6	31	23	yes	
	male 4	3	22	22	yes	
	male 5	3	35	35	yes	
	male 6	0	-	-	no	enlarged thymus

Table 5.3: Test breeding chimaeras for germline transmission of gene trap ES cells. An agouti coat colour in offspring indicated ES cell colonization of the germ cells, while offspring derived from host strain germ cells displayed a black coat colour. For offspring generated from the COB36 male chimaeras, 11 pups were found dead (FD) prior to development of the coat. Three chimaeras showed poor breeding success over a four month period (COB36 male 2, COB37 female 1, COB52 male 1). COB54 was the only ES cell line to contribute to the germline in chimaeras.
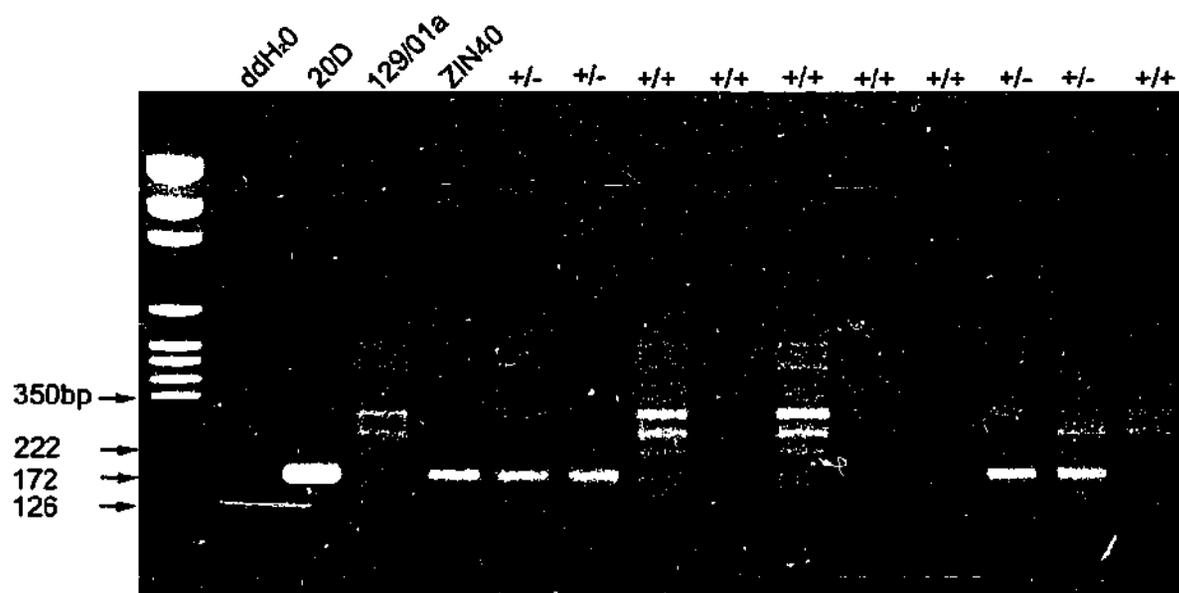
Figure 5.7: Analysis of gene trap transmission by PCR amplification of βgeo sequence. Unextracted genomic DNA was prepared from ear punches for three week old agouti offspring generated from COB54 male chimaera and C57Bl/6J matings. Offspring heterozygote for the gene trap allele (+/-) were detected by amplification of a 172 bp product corresponding to vector-specific βgeo sequence. Positive control reactions were performed for genomic DNA from the transgenic ZIN40 mouse strain as well as for gene trap vector DNA, while genomic DNA from non-transgenic 129/Ola mice and sterile ddH₂0 were used as negative control templates.

5.7 GENOTYPIC ANALYSIS FOR TRANSMISSION OF THE COB54 GENE TRAP ALLELE

5.7.1 PCR analysis for *βgeo* sequence

For rapid detection of the COB54 gene trap allele in germline ES cell-derived offspring, crude genomic DNA samples were prepared from weaning age agouti pup ear punches (section 2.2.2d) and used directly in a PCR analysis for vector-specific sequence. Amplification reactions were prepared in a final volume of 30 μl (section 2.2.7a), with 20 pmol each of the BGEO5' sense and BGEO3' antisense primers corresponding to *βgeo* sequence (section 2.1.3). Crude genomic DNA prepared from ear punches for the 129/Ola and ZIN40 mouse strains (section 2.1.6), as well as pGT1.8 (OPT/Nuclear) IRES*βgeo* plasmid DNA (section 3.3) were used as control templates. Reactions were amplified for 1 cycle of 95°C/6 min (adding the *Taq* polymerase in the final minute), 52°C/30 s, 72°C/30 s, then for 30 cycles of 95°C/30 s, 52°C/30 s, 72°C/ 30 s, and a final cycle of 95°C/30 s, 50°C/30 s, 72°C/5 min. Detection by gel electrophoresis for a PCR product corresponding to 172 bp of *βgeo* sequence was indicative of animals positive for the gene trap allele (Figure 5.7).

Results of the rapid PCR screening of ear punch genomic DNA from 123 agouti pups indicated that the COB54 gene trap allele was transmitted to 52% of offspring, with no significant bias towards male or female heterozygotes (data not shown).

5.7.2 Southern analysis of the COB54 transgenic allele

Heterozygosity for the transgenic COB54 allele was confirmed by Southern blot hybridization (sections 2.2.3b & 2.2.6a) of extracted weaning age tail tip genomic DNA (section 2.2.2d), using a DNA probe isolated from cloned 5'-RACE products for the COB54 fusion transcript (section 5.3.1). This 730 bp probe, "*Cob54Race*", is specific to COB54 endogenous sequence immediately upstream of the gene trap insertion site (section 2.1.2, Figure 5.3).

To firstly identify a genomic restriction fragment length unique to the COB54 gene trap allele, genomic DNA prepared from four COB gene trap ES cell lines and from control E14 and OKO160 ES cell lines (sections 2.1.5 & 2.2.2c) was separately digested with a panel of restriction enzymes including *Xba* I, *Sph* I, *Eco*R I, *Bam*H I, *Bgl* II, *Hind* III and *Pst* I (sections 2.2.1c & 2.2.3b). Southern blots were hybridized (sections 2.2.3b &
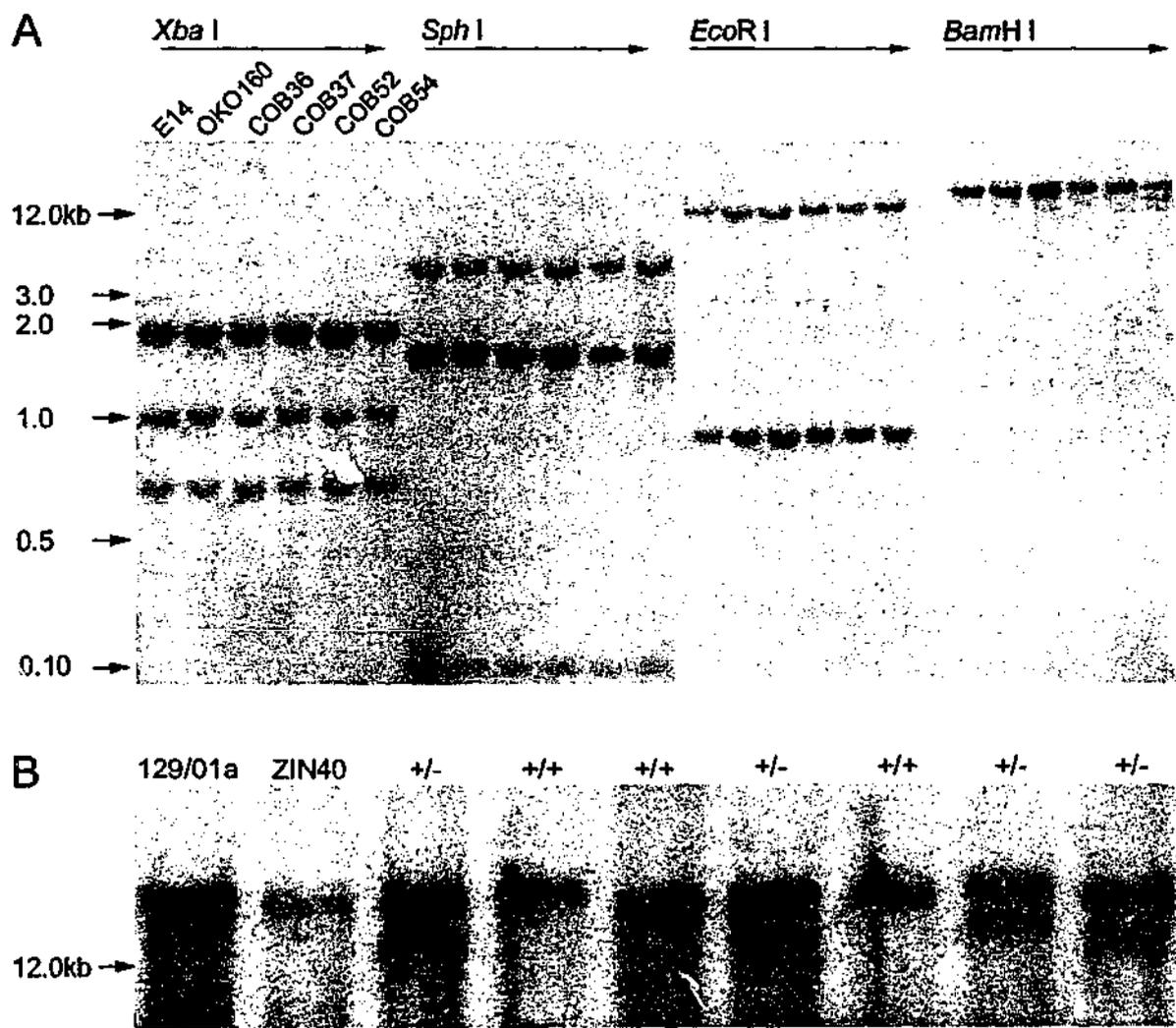
Figure 5.8: Southern blot analysis to identify the COB54 transgenic allele. (A) 15 µg of genomic DNA for each gene trap ES cell line (COB36, COB37, COB52, COB54), and for control ES cell lines (E14, OKO160), was digested with a panel of restriction enzymes including *Xba* I, *Sph* I, *EcoR* I and *BamH* I. Southern blots were hybridized with the *Cob54Race* probe corresponding to 730 bp of endogenous sequence 5' of the gene trap insertion site in the COB54 ES cell line. The 5'-RACE-derived probe hybridized with the wild type COB54 allele in all reactions, and also identified a smaller restriction fragment unique to the COB54 heterozygote gene trap allele (+/-) following digestion with *BamH* I. (B) 15 µg of tail tip genomic DNA for weaning age COB54 germline agouti offspring, and for ZIN40 and 129/Ola adult mice, was digested with *BamH* I and Southern blots hybridized with the *Cob54Race* probe. For the samples shown, a heterozygote (+/-) or wild type (+/+) genotype for the COB54 gene confirmed the genotype identified by PCR analysis for *βgeo*. Autoradiographic exposure was for 5-6 days at -80°C.

2.2.6a) with the *Cob54Race* DNA probe. Results, shown for four restriction enzyme digests (Figure 5.8A), confirmed hybridization of the 5'-RACE-derived sequence with a wild type allele in all ES cell lines and with a smaller *Bam*H I restriction fragment length unique to the heterozygote COB54 ES cell line. Both alleles were seen to migrate high on the gel above the highest marker of 12 kb for the *1 kb Plus* DNA ladder used (section 2.2.1d).

Tail tip genomic DNA samples for COB54 agouti offspring, and for control ZIN40 and 129/Ola mice (section 2.1.6), were subsequently digested with *Bam*H I and Southern blots hybridized with the *Cob54Race* probe (sections 2.1.2, 2.2.3b & 2.2.6a). Hybridization results confirmed detection of the restriction fragment for the endogenous COB54 allele in germline offspring, as well as that for the transgenic allele in heterozygotes (Figure 5.8B). This analysis identified a false positive error rate of 4.8% in the previous rapid *βgeo* PCR analysis for heterozygote offspring (section 5.7.1) and confirmed a transmission frequency of 47% for the gene trap allele, with no significant bias in the sex ratio (Table 5.4).

# Agouti pups	Sex	Wild type genotype	Heterozygote genotype
56	male	29	27
67	female	36	31

*Table 5.4: COB54 gene trap allele transmission frequency as determined by Southern blot analysis. For 123 agouti offspring derived from COB54 male chimaeras, tail tip genomic DNA was digested with Bam*H *I and Southern blots hybridized with the Cob54Race probe. This analysis detected the endogenous COB54 allele as well as the COB54 gene trap allele in 47% of offspring, without any significant bias in the sex ratio.*

5.8   BREEDING THE COB54 GENE TRAP ALLELE TO HOMOZYGOSITY

Intercross natural matings were established between adult agouti heterozygote mice to observe if offspring homozygote for the COB54 gene trap insertion would display any phenotypic difference to wild type and heterozygote litter mates. As the gene trap allele harbors an interruption in a gene that is expressed specifically in ES cells *in vitro*, it was a considered possibility that homozygote embryos might not develop to term.

5.8.1 Southern analysis of COB54 heterozygote intercross offspring

Offspring derived from heterozygote intercross matings were initially genotyped by Southern hybridization analysis for the COB54 gene trap and wild type alleles as described in section 5.7.2, using genomic DNA prepared from 10.5 dpc midgestation embryos and from weaning age neonate tail tips (section 2.2.2d). Results for the Southern blot analysis of 74 postimplantation stage embryos and 73 weaned neonates are shown in Table 5.5. The expected mendelian ratio was observed for the wild type genotype in both developmental stages but was skewed significantly above that expected for heterozygotes, with no offspring homozygote for the gene trap allele. For weaned offspring, no bias was observed in the sex ratio either amongst the heterozygotes, or overall (data not shown).

#Intercross litters	Age at analysis	#Embryos/ Offspring	Wild type genotype	Heterozygote genotype	Homozygote genotype
8	10.5 dpc	74	20 (27%)	54 (73%)	0 (0%)
14	3 weeks post-partum	73	19 (26%)	54 (74%)	0 (0%)

Table 5.5: Southern blot analysis of COB54 genotypes for heterozygote intercross midgestation and liveborn offspring. Genomic DNA prepared from 10.5 dpc fetuses and 3 week old offspring was digested with BamH I and Southern blots hybridized with the Cob54Race probe. For both developmental stages the percentage of wild type genotypes observed for the total number of offspring analysed (%) is as expected for a normal mendelian ratio, whereas the transgenic genotype is skewed significantly above the predicted ratio for heterozygotes. No homozygotes were identified in either group.

The high ratio observed for heterozygote to wild type genotypes (2.7:1 for embryos and 2.8:1 for neonates) was further confirmed by PCR analysis for βgeo sequence on all neonate DNA samples (section 5.7.1, data not shown). This suggested that the established Southern analysis was not accurate in determining the homozygote genotype and that animals homozygous for the transgenic allele had been included with the number of heterozygote animals. Southern gels were initially electrophoresed for about 20 h at 30-40 volts (section 2.2.3b), but this was later extended to 24-25 h to provide improved separation of hybridization signals for the high Mw restriction fragments
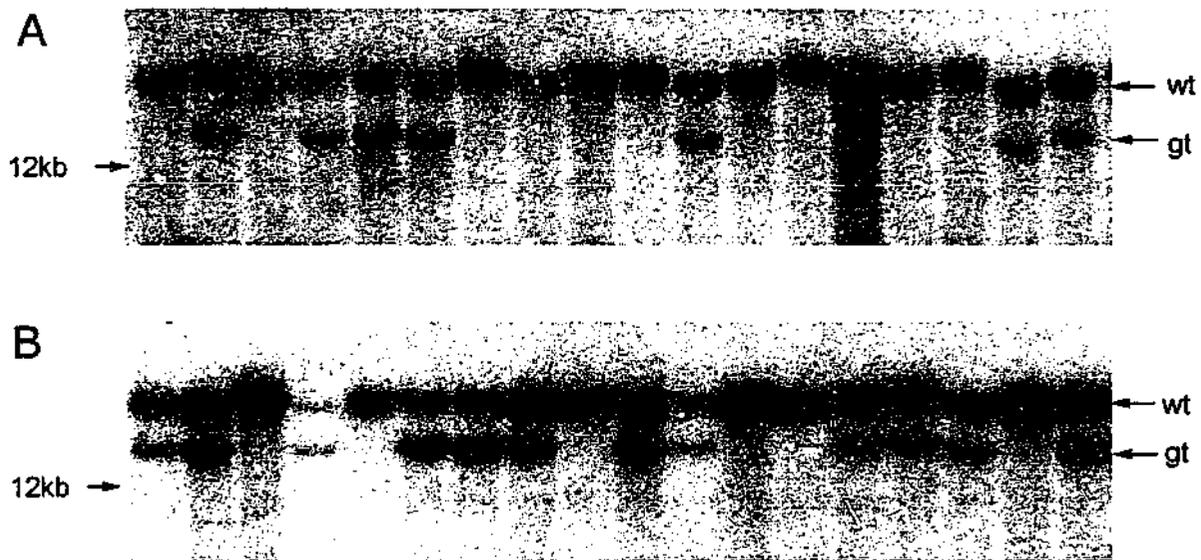
Figure 5.9: Southern blot analysis of COB54 heterozygote intercross offspring. Genomic DNA from (A) 3 week old neonate tail tips and (B) 10.5 dpc midgestation embryos derived from COB54 heterozygote intercross matings, was digested with BamH I and analysed by Southern blot hybridization with the *Cob54Race* probe. While this analysis detected a unique restriction fragment for the COB54 gene trap allele (gt), it did not accurately decipher the homozgyote and heterozygote genotypes in intercross offspring. Electrophoresis of genomic DNA for over 24 h at 30-40 volts, and a shorter autoradiographic exposure of 2-3 days at -80°C with intensifying screens, identified a double band at the position of the restriction fragment for the wild type allele (wt). This doublet included a cross-hybridizing band of unknown affiliation and a smaller band associated with the endogenous COB54 allele.

corresponding to wild type and transgenic COB54 alleles (section 5.7.2). Increased separation coupled with shorter autoradiographic exposure times (usually 3-4 days), clearly confirmed a double band at the wild type restriction fragment position. The doublet included a cross-hybridizing band of unknown affiliation and a smaller band associated with the endogenous COB54 allele. Examples of this for both midgestation embryo and neonatal analyses are shown in Figure 5.9. The slight variation seen in mobilities for hybridization bands in each sample made deciphering a homozygote from a heterozygote genotype difficult.

5.8.2 Genotyping by multiplex RT-PCR analysis

To decipher the heterozygote and homozygote COB54 genotypes, a multiplex RT-PCR assay was developed for specific amplification of the wild type and disrupted COB54 alleles. The forward primer COB54F2, previously used for 3'-RACE (section 5.3.2), corresponds to COB54 sequence starting 147 bp upstream of the gene trap insertion (Figure 5.3), and should be common to both alleles. A new reverse primer COB54R3, corresponding to endogenous sequence starting 23 bp downstream of the gene trap insertion site (Figure 5.3), was predicted to amplify a 170 bp cDNA fragment for the wild type allele. In the disrupted allele, where the COB54R3 sequence would be absent, annealing with the previously used 5'-RACE oligo 3 primer corresponding to *En-2* gene trap sequence 129 bp downstream of the splice site (Figure 4.2), was predicted to amplify a 276 bp cDNA fragment. To test the three primers and confirm the predicted product sizes, E14 (wild type) and COB54 ES cell poly A^+ mRNA templates (section 2.2.2f) were reverse transcribed (section 2.2.7b) from the COB54R3 and oligo 3 primers, respectively. Each of the RT products was amplified individually with their respective reverse primers and the forward COB54F2 primer, using 2.5 units of *Taq* DNA polymerase and a T_a of 54°C (section 2.2.7a). Gel electrophoresis of products confirmed the expected band sizes for each allele (data not shown). All primer sequences are given in section 2.1.3.

As the gene trap vector was designed to insert into an intron, it is presumed that the insertion site in the cDNA sequence represents the junction between two exon sequences. Attempts to establish a genomic PCR analysis using the above primers and several heterozygote intercross tail tip genomic DNA templates (section 5.8.1) were not successful, and were suggestive of a large intronic region between the 5' and 3' exon sequences. However, as early expression studies for the COB54 gene trap allele had
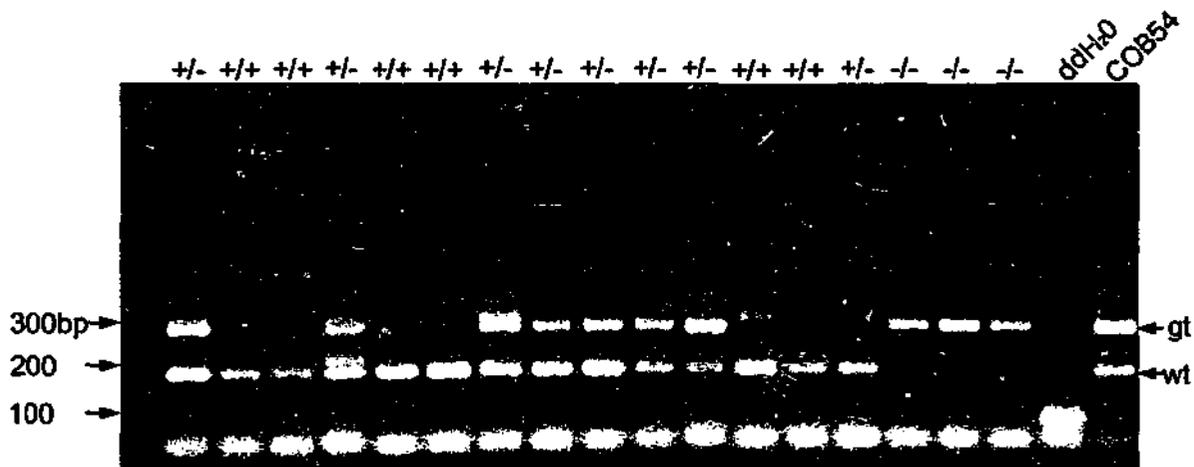
Figure 5.10: Identification of COB54 homozygote embryos by a multiplex RT-PCR analysis.
Embryonic cDNAs were generated for single 3.5 dpc blastocysts from COB54 heterozygote
intercross matings with random primer RT. In a multiplex PCR reaction, a 276 bp product
corresponding to sequence for the gene trap allele (gt) was amplified between *En-2* gene trap
vector sequence and an upstream primer 5' of the insertion in the endogenous sequence. A 170 bp
product corresponding to sequence for the wild type allele (wt) was simultaneously amplified
between the same endogenous upstream primer, and a downstream primer in endogenous
sequence 3' to where the gene trap inserts in the transgenic allele. COB54 ES cell mRNA (COB54)
was used as a positive control for each allele, with sterile ddH$_2$0 as a negative control template.
Detection of either one or both alleles in single embryos identified each of the expected genotypes at
this stage of development.

identified strong expression in preimplantation heterozygote embryos (discussed in Chapter six), a multiplex RT-PCR assay was able to be established for the analysis of RNA from single intercross blastocysts.

Heterozygote intercross superovulated matings were established for genotyped COB54 mice, and 3.5 dpc blastocysts flushed from the uterine horns of plugged females (section 2.4.1). Using random primers, RNA was reverse transcribed directly from lysates of single blastocysts and PCR reactions amplified according to the protocol described in section 2.2.7c. A mixture of all three primers in the same reaction allowed the simultaneous amplification of both the wild type and/or gene trap alleles, and also served as an internal control for amplification of RNA in each reaction. COB54 ES cell mRNA was run as a positive heterozygous control (Figure 5.10). As embryonic cDNA was amplified in 50 cycles some amplification of non-specific products was seen, but this was largely minimized by a reduction in cycle times and by increasing the T_a to 56°C (section 2.2.7c).

From a pool of 136 intercross blastocysts from 6 litters, RT-PCR analysis was performed on 106 embryos. Results in Table 5.6 indicate that at the preimplantation stage, the distribution of COB54 genotypes is approximately as expected for a normal mendelian inheritance. In most cases the PCR result was unambiguous, with only 5 samples requiring repeat analysis. For 8 samples, no amplifiable product was detected, most probably due to a lack of, or poor quality RNA template in the lysate.

#Blastocysts	Genotype +/+	Genotype +/-	Genotype -/-	Not amplified
106	26 (26.5%)	52 (53.1%)	20 (20.4%)	8

Table 5.6: Multiplex RT-PCR analysis of blastocysts from COB54 heterozygote intercross matings. Single 3.5 dpc blastocysts were genotyped by a multiplex RT-PCR reaction to detect both the gene trap and wild type COB54 alleles. Genotype distributions are calculated as a percentage of those blastocysts yielding detection of one or both alleles. Results confirm the presence of homozygous genotypes for the preimplantation stage of development, at a ratio reflecting Mendelian inheritance.

Results for the RT-PCR genotyping of blastocyst stage embryos, and the Southern analyses for midgestation stage embryos and weaned offspring provided strong evidence to suggest that homozygotes for the COB54 gene trap allele are viable. Subsequent to these analyses, viable and fertile adult homozygote animals were

confirmed by 100% transmission of the gene trap allele to offspring (discussed in Chapter six), and do not display any overt phenotype. A normal, healthy COB54 homozygote mouse line has since been maintained for four successive generations.

5.9 DISCUSSION

The results in this chapter describe the compilation of a 1.7 kb cDNA sequence for the entrapped gene in the COB54 ES cell line. This cDNA sequences encodes a 38 amino acid F-box motif and is homologous but for one nucleotide (Figure 5.4) with a recently reported gene, *Fbx15*, belonging to a new family of approximately 50 mammalian F-box proteins (Cenciarelli *et al.*, 1999; Winston *et al.*, 1999a). Named for conservation of the characteristic motif originally identified at the SKP1 binding site in a human Cyclin F-interacting complex (Bai *et al.*, 1996), the F-box proteins belong to a growing family of adaptor subunits that are thought to specifically recruit substrates for ubiquitin-mediated degradation and are linked by their common motif to a core ubiquitination complex. In most instances, substrate-level phosphorylation drives capture of the substrate by a protein-protein interaction domain in the F-box protein. Although a number of F-box proteins have been identified in budding yeast *S. cerevisiae* and in the nematode *C. elegans*, prior to the reporting of this new family only a handful of mammalian F-box proteins – Cyclin F, SKP2, β-TRCP, Elongin A and NFB42 – had been identified (Winston *et al.*, 1999a; reviewed by Tyers and Jorgenson, 2000). While functions have not yet been ascribed to these novel proteins, nor for a partially overlapping group of 5 novel *Xenopus* F-box proteins (Regan-Reimann *et al.*, 1999), the following discussion outlines the important role that is emerging for this expanding family of proteins in cell regulation (for reviews see Patton *et al.*, 1998a; Craig and Tyers, 1999; Tyers and Jorgenson, 2000).

Ubiquitin-mediated protein degradation is now recognized as a principal regulatory mechanism of the cell, controlling many cellular processes including cell cycle progression, morphogenesis, and signal transduction (Hochstrasser, 1996). Major cell cycle transitions, for example, require the precise destruction of a number of proteins including inhibitors of the cyclin-dependent kinases (Cdks) that drive DNA replication, mitotic cyclins and anaphase inhibitors (reviewed by King *et al.*, 1996).

In marking proteins for protease destruction, the highly conserved 76 amino acid ubiquitin protein is covalently attached to lysine residues of substrate proteins through

the coordinated action of three enzymes: The E1 ubiquitin-activating enzymes, the E2 ubiquitin-conjugating enzymes, and the E3 ubiquitin ligases. Whereas the E1 and E2 enzymes are primarily involved in activating and transferring ubiquitin through formation of high-energy thioester bonds directly with the carboxyl terminus of ubiquitin, E3 enzymes mediate the critical step of substrate-specific recognition and promoting polyubiquitin ligation to that substrate (Hochstrasser, 1996; Hershko, 1997). Assembly of a polyubiquitin chain on a target substrate quickly leads to its capture and degradation by the 26S proteasome (Baumeister et al., 1998), (reviewed by Hershko and Ciechanover, 1998; Laney and Hochstrasser, 1999).

At least two main multiprotein complexes that function as E3 ubiquitin ligases have been identified from extensive studies of many cell cycle regulators. The anaphase-promoting complex (APC; also called the cyclosome), is active from the beginning of anaphase until the end of G1 phase, a window in which it eliminates anaphase inhibitors, mitotic cyclins and components of the mitotic spindle, and even elements of its own activation pathway (reviewed by Zachariae and Nasmyth, 1999). The APC targets proteins containing a short primary sequence determinant called the destruction box (Hershko, 1997) and has been shown in several organisms to utilize subfamilies of WD40 repeat proteins as cofactors in the degradation of specific substrates (reviewed by Patton et al., 1998a). The second multiprotein complex, the Skp1-Cdc53/cullin-F-box protein (SCF) complex, has been discovered through extensive analysis of cyclin and Cdk inhibitor degradation pathways in the yeast cell cycle, and is now known to catalyse the phosphorylation-dependent ubiquitination of proteins in a variety of signaling pathways ranging from nutrient sensing in yeast, to conserved developmental pathways in plants and animals (reviewed by Patton et al., 1998a; Craig and Tyers, 1999; Tyers and Jorgenson, 2000).

The SCF complexes (Figure 5.11) recruit their phoshphorylated substrates via the variable F-box protein interaction domains, while the conserved approximately 40 amino acid F-box motif links these adaptor proteins to Skp1 in the core E3 ligase (Bai et al., 1996). In addition to binding the F-box protein, Skp1 also binds directly to an N-terminal region of yeast Cdc53p, or its homolog from the cullin family of proteins in other species (Kipreos et al., 1996; Krek, 1998). Within the SCF complex, Cdc53p provides an independent binding site for the E2 ubiquitin-conjugating enzyme Cdc34p, and thus acts as a scaffold protein for the E2-Skp1-cullin core ubiquitination complex
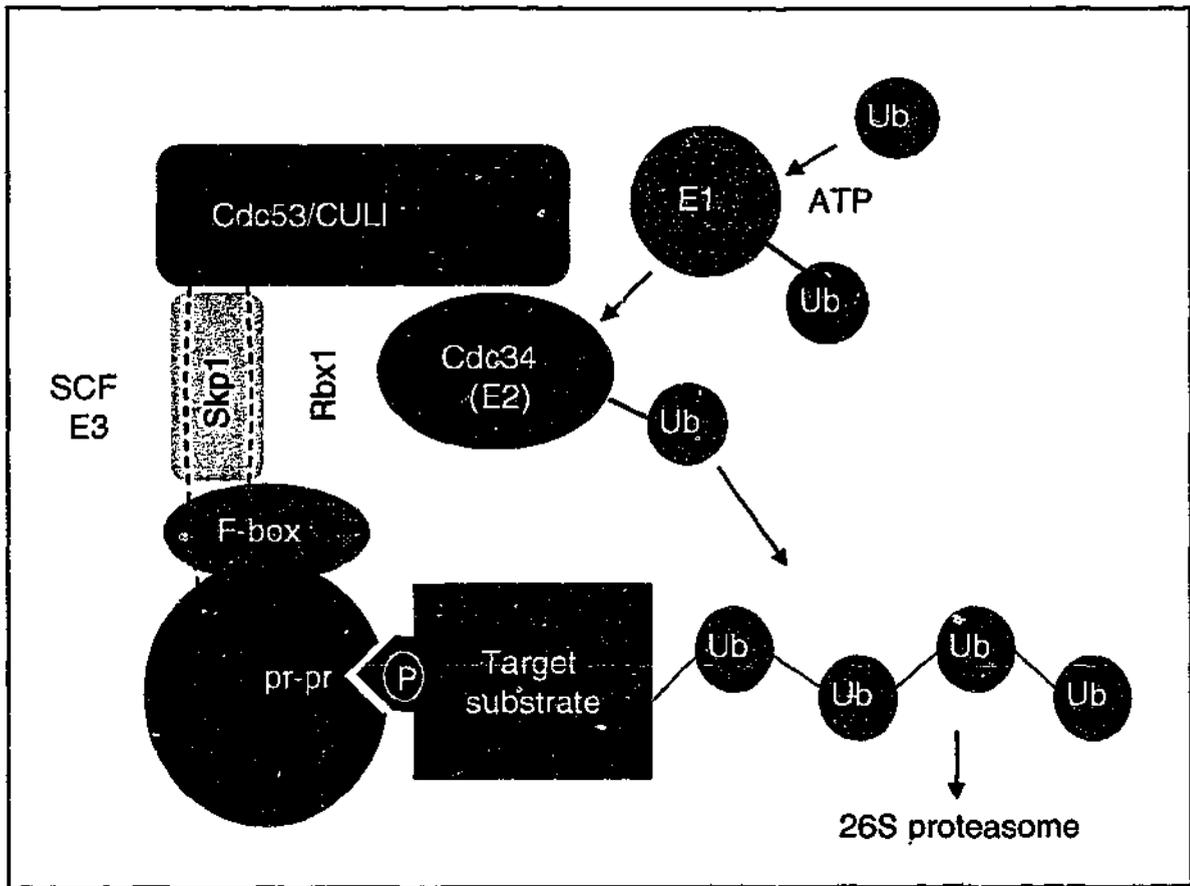
Figure 5.11 A role for F-box proteins in ubiquitin-mediated proteolysis. In the ubiquitin system for cell regulation, substrate proteins are marked for protein degradation by the coordinated action of three enzymes. The E1 enzyme activates ubiquitin (Ub) in a thioester linkage, which is transesterefied to an E2 conjugating enzyme, Cdc34. This is turn transfers ubiquitin to a substrate lysine residue in association with an E3 ubiquitin ligase complex. Polyubiquitinated substrates are rapidly recognized and degraded in the 26S proteasome. In the Skp1-Cdc53/cullin-F-box protein (SCF) E3 complexes, F-box proteins serve as adaptor units that specifically recruit phosphorylated (P) target substrates for ubiquitin degradation. F-box proteins are linked by their conserved F-box motif (F-box) to Skp1 in the core E3 ligase, and by a variable protein interaction domain (pr-pr) to the phosphorylated target substrate. Yeast Cdc53 or its cullin family homolog (CUL1) acts a scaffold protein for a Skp1-cullin-E2 ubiquitination complex. The RING finger protein Rbx1 interacts independently with Cdc53 and Cdc34, as well as with many F-box proteins, to promote E2 ubiquitination of the F-box recruited substrate.

(Skowyra *et al.*, 1997; Feldman *et al.*, 1997; Patton *et al.*, 1998b). Rbx1, a RING finger protein containing a small zinc-binding domain (also called Hrt1 or Roc1), has been recently identified as a further integral component at the hub of SCF complexes. Rbx1 appears to stabilize the E2/E3 complex by interacting independently with Cdc34p, Cdc53p, and multiple F-box proteins, to promote E2 ubiquitination (Kamura *et al.*, 1999; Ohta *et al.*, 1999; Seol *et al.*, 1999; Skowyra *et al.*, 1999). Rbx1 is also a component of the von Hippel-Lindau (VHL) tumor suppressor complex, and is strikingly similar to the Apc11 subunit of the APC (Skowyra *et al.*, 1999). Several recent reports now suggest that proteins containing a RING finger domain not only stimulate ubiquitination by specifically interacting with E2 conjugating enzymes, but may in some cases act as E3 ubiquitin ligases themselves (reviewed by Freemont, 2000).

It is of much recent interest that homologs of the cullin proteins are now implicated in other modular E3 ubiquitin ligase systems. In mammalian cells CUL2 appears to form a family of complexes based on a different family of adaptor subunits, the SOCS-box proteins, which act as intracellular inhibitors of several cytokine signal transduction pathways (Kamura *et al.*, 1998; Hilton *et al.*, 1998). These proteins, of which there are currently 20, contain a BC-box submotif within a C-terminal SOCS-box that interacts with Elongin C, a Skp1-related protein. Elongin C in turn interacts with a ubiquitin-like protein Elongin B, and also with CUL2. SOCS-box proteins, like F-box proteins frequently contain additional protein-protein interaction motifs such as WD40 repeats, ankyrin repeats, and SH2 domains, that may serve to link them to possible substrates (Hilton *et al.*, 1998; Kamura *et al.*, 1998; Zhang J. *et al.*, 1999). Interestingly, Elongin A is a SOCS-box protein that interacts with the Elongin C/B heterodimer, and also contains an F-box like sequence (Aso *et al.*, 1996; Bai *et al.*, 1996).

The VHL tumor suppressor protein contains a SOCS-box in complex with Elongins C/B (the VCB complex), and is now itself implicated as an E3 ubiquitin ligase (Stebbins *et al.*, 1999). *In vitro*, a VCB-Cul2-Rbx1 complex displays a ubiquitin ligase activity which may target the important hypoxia inducible transcription factor HIF1-α for degradation (Iwai *et al.*, 1999; Lisztwan *et al.*, 1999; Maxwell *et al.*, 1999). That at least five cullin family members and a Cdc53p homolog in the APC called Apc2, have now been found to interact with the RING finger Rbx1/Apc11 family of proteins, lends further evidence for different substrate receptors in distinct cullin ubiquitin ligase

complexes, and may underlie the diversity of regulated ubiquitination (Kamura *et al.*, 1999; Ohta *et al.*, 1999; Zachariae and Nasmyth, 1999), (reviewed by Craig and Tyers, 1999; Tyers and Jorgenson, 2000).

Characterization in yeast of the first SCF pathway identified the F-box protein Cdc4p as a regulator of the G_1/S-phase cell cycle transition, by directing ubiquitination of the Clb-Cdc28 inhibitor Sic1p. A WD40 repeat domain in Cdc4p binds phosphorylated Sic1p, while the F-box domain binds Skp1 in the SCF complex (SCF^{Cdc4}) for ubiquitination by Cdc34p (Feldman *et al.*, 1997; Skowyra *et al.*, 1997). In a similar manner, other key substrates recruited to SCF^{Cdc4} for ubiquitination include a second Cdk inhibitor Far1p, a replication protein Cdc6p and a transcription factor Gcn4p (reviewed by Patton *et al.*, 1998a). Grr1p, a different yeast F-box protein containing leucine rich repeats, mediates G1 cyclin ubiquitination following autophosphorylation of the Cln1/2-Cdc28 kinase complex (Deshaies *et al.*, 1995; Willems *et al.*, 1996; Skowyra *et al.*, 1997). SCF^{Grr1p} also eliminates activators of polarized growth called Gic1p and Gic2p (Jaquenoud *et al.*, 1998), as well as an unknown target in the glucose induction pathway (Li and Johnston, 1997). Met4, a transcription factor for methionine biosynthesis genes, is recruited for degradation by yet another yeast F-box protein called Met30p (Patton *et al.*, 1998b; Rouillon *et al.*, 2000). SCF^{Met4} also targets the Cdk inhibitory kinase Swe1 (Wee1 in other species) for degradation in the G_2/M phase, restricting the window in which Swe1 can phosphorylate and inhibit Cdc28 (Kaiser *et al.*, 1998).

The SCF pathway has proven to be evolutionarily conserved, since all the subunits of yeast SCF complex have orthologs in mammals. SKP1 was in fact originally identified in a complex with Cyclin A/Cdk2 and the F-box protein SKP2 in human cells (Zhang H. *et al.*, 1995). In mammalian cells, the best studied SCF^{Skp2} complex for an F-box with leucine rich repeats includes the invariant core consisting of SKP1 and CUL1 associated with Cdc34p, and has arisen as a key player in the regulation of Cyclin E and the G_1/S phase transition (reviewed by Tyers and Jorgenson, 2000). Recently, SKP2 has been implicated in ubiquitination of the Cdk inhibitor $p27^{Kip1}$ and the transcriptional control protein E2F-1 (Carrano *et al.*, 1999; Marti *et al.*, 1999; Sutterluty *et al.*, 1999; Tsvetkov *et al.*, 1999).

Currently F-box proteins are the largest known class of E3 ubiquitin ligase receptors, 15 in budding yeast, more than 60 in *C. elegans* (Bai *et al.*, 1996; Patton *et al.*, 1998a), and

now over 50 in mammalian and vertebrate cells (Cenciarelli *et al.,* 1999; Regan-Reimann, 1999; Winston *et al.,* 1999a), suggesting that the SCF pathway will control ubiquitination of a large number of proteins. However, the finding that the F-box protein Ctf13p interacts with Skp1 as a structural component of the CBF3 kinetochore complex in yeast, which does not contain Cdc53, Cdc34 or Rbx1 and does not appear to be a ubiquitin ligase (Kaplan *et al.,* 1997), cautions that some F-box proteins will have roles beyond substrate ubiquitination (reviewed by Patton *et al.,* 1998a; Tyers and Jorgenson, 2000).

It is not known how cells regulate the level of the different SCF complexes and how these levels are altered through the cell cycle, and in reponse to extracellular signals. The human F-box protein SKP2 is expressed in a cell cycle-dependent manner, suggesting that regulation of SCF complexes may mainly affect F-box proteins (Lisztwan *et al.,* 1998). The yeast F-box proteins Grr1p, Cdc4p and Met30p have been recently shown to be unstable components of the SCF, and are themselves degraded in a ubiquitin- and proteasome-dependent manner *in vivo.* This suggests that ubiquitination occurs within the SCF complex by an autocatalytic mechanism, which may allow rapid switching among multiple SCF complexes (Zhou and Howley, 1998; Galan and Peter, 1999). Such a mechanism would provide for quick cellular adaptation to changed physiological conditions, and progression through different phases of the cell cycle. This mechanism may also facilitate the change in timing of cell cycle phases which occurs upon ES cell differentiation (Savatier *et al.,* 1996).

While F-boxes have been found prominently within a number of cell cycle regulatory proteins, evidence is beginning to emerge for F-box function in other gene regulatory processes, as exemplified by the following reports. The *Drosophila* gene *Slimb* encodes a conserved F-box/WD40 repeat protein related to yeast Cdc4p and is implicated in negative regulation of the *Hedgehog (Hh)* and *Wnt/Wingless (Wg)* signaling pathways by targeted degradation of their respective protein products, *cubitus interruptus* and *armardillo* (Jiang and Struhl, 1998). A Cdc4p ortholog in *C. elegans,* SEL-10, negatively regulates LIN-12/Notch signaling by either inhibiting release of and/or by stimulating proteolysis of the Notch intracellular domain (Hubbard *et al.,* 1997). In *Arabidopsis,* an SCF complex is required for the plant hormone auxin response that regulates diverse aspects of plant growth and development, and contains an F-box protein TIRI which is related to human SKP2 and yeast Grr1p (Ruegger *et al.,* 1998;

158

Gray *et al.*, 1999). The mammalian β-Trcp F-box protein targets phosphorylated IκBκ in inflammatory, immune and stress response pathways, and is also thought to act as a negative regulator in the *Wnt/β-catenin* signaling pathway responsible for dorsal axis formation in Xenopus embryos (Yaron *et al.*, 1998; Marikawa and Elinson, 1998; Winston *et al.*, 1999b). In the one spectacular example to date for viral exploitation of an SCF pathway, the Vpu protein of HIV has been shown to downregulate expression of the CD4 membrane receptor in infected T-helper cells, by redirecting the specificity of human β-TRCP to capture and degrade CD4 (Margottin *et al.*, 1998).

At present, the full functional significance for the expanding family of F-box proteins is only beginning to be understood, with the identity of the relevant substrates for the myriad of presumed SCF complexes remaining the key outstanding issue. It is therefore of much interest, that the novel mouse *Fbx15* cDNA (Winston *et al.*, 1999a) has been identified in this study on the basis of an ES cell-restricted *in vitro* expression profile for the entrapped gene. In light of the preceding discussion, a regulatory role is implied for this protein in ES cell growth and self-renewal by targeting a transcriptional activator or repressor substrate for ubiquitin-mediated degradation. The marked downregulation of this F-box protein upon *in vitro* ES cell differentiation, suggests that rapid upregulation of the target substrate may be an early requirement for an ES cell to enter into a different cell cycle or differentiation pathway. As discussed earlier, the Fbx15 protein might itself be degraded in a ubiquitin-dependent manner in order to elicite an even quicker cellular response to facilitate this progression. Considering the evidence to date for F-box protein regulation of cyclin expression in cell cycle control, the induction of Cyclin D/Cdk4 expression and the reinstatement of G1 cell cycle control mechanisms associated with *in vitro* differentiation of ES cells (Savatier *et al.*, 1996) may be regulated by the level of *Fbx15* expression in ES cells. Although no known protein interaction domain has been identified for the novel Fbx15 protein (Winston *et al.*, 1999a), the elucidation of its target substrate(s) and associated function(s) may eventually throw new light on critical pathways for ES cell regulation.

For the five 129/Ola-derived gene trap ES cell lines introduced to host C57Bl/6J blastocysts in this study, only the COB54 line for the *Fbx15* interruption was able to contribute to the germline in chimaeras. This cell line gave rise to the most viable offspring, and yielded the highest proportion of male chimaeras from the least number of blastocysts transferred (Table 5.2), suggesting a greater permissibility for COB54 ES

cells to contribute to chimaeric embryos. *In vitro*, the very strong selectable recovery of ES cell cultures for the COB54 cell line from suspension cultures grown in the absence of LIF (Figure 5.2) also implies that pluripotentiality has not been compromised by the associated gene trap insertion in an ES cell-restricted gene. The six COB54 male chimaeras showed a high degree of ES cell contribution as judged by predominant light coat colouration (Figure 5.6) and with one exception, all were viable and fertile. After initial identification of germline-derived offspring by agouti coat colouration, (Table 5.3), transmission of the gene trap allele was assessed by a rapid PCR screen for the *βgeo* gene (Figure 5.7). Southern blot analysis for the COB54 endogenous and gene trap alleles subsequently confirmed heterozygosity for the COB54 gene, in approximately 50% of progeny (Figure 5.8; Table 5.4) and identified a 4.8% false positive error rate in the PCR analysis.

Having trapped genes that are preferentially expressed in the pluripotential ES cells *in vitro*, the failure to generate any germline chimaeras for the four other gene trap cell ES cell lines is in itself, an interesting result and is possibly suggestive of a disruption in an endogenous gene effecting pluripotency of the transgenic ES cells. A modification in gene function or in the level of the ES cell-restricted protein produced by the entrapped allele, may lead to an associated effect on the pluripotential cells of the preimplantation embryo. Such a change in expression might increase the likelihood for terminal differentiation *in vivo*, and consequently reduce fitness to form totipotent germ cells. This has been previously suggested in the case of heterozygosity at the *Oct-4* locus (Nichols *et al*, 1998), where targeted ES cells contained about one half of the wild type levels of Oct-4 DNA binding activity and generated low absolute numbers of germline transmitting chimaeric mice (4/30 chimaeras from 2 injected clones), (B. Zevnik, Centre for Genome Research, Edinburgh, UK; *pers. comm.*).

For the COB28 cell line, which was derived from a lower passage of the same E14Tg2a parental cell line that gave rise to the COB54 cell line, no offspring were born from 6 embryo transfers of a total 69 blastocysts (Table 5.2). This result suggests that even a minor contribution of the gene trap COB28 ES cells to a host blastocyst was detrimental and points to a possible dominant effect of the ES cells due to a single gene trap allele dosage effect, or possibly a null phenotype due to a gene disruption in a single X or Y chromosome. The observation that the COB28 cell line was generally less capable of

robust growth during routine ES cell culture also supports a possible detrimental effect of the gene trap interruption in the COB28 cell line.

While two chimaeras were obtained for each of the COB36, COB37 and COB52 cell lines, none of these were germline transmitting despite a high ES cell contribution as judged by coat colour (Figure 5.6; Table 5.3). The two male COB36 chimaeras displayed low breeding success and a high mortality rate in newborn offspring, although it is not possible to conclude whether this can be attributed in any way to colonization of the germ cells in COB36 ES cell chimaeras. For the COB37 chimaeras, one of which was female, all offspring were derived only from the C57Bl/6J host germ cells. That the COB36 and COB37 cell lines were derived from a higher passage parental ES cell line than the other three lines might is also a consideration regarding their germline competence. For the two COB52 male chimaeras, one failed to produce any pregnancies over a four month breeding period, while the other died of an unknown cause around 5 weeks of age. Whether the failure of these cell lines to contribute to the germline is the result of a heterozygous dosage effect in gene trap ES cells, or perhaps the result of an interruption in a gene on a single X or Y chromosome, and even a high ES contribution to germ cells where the mutated gene is required for spermatogenesis, is purely speculative. To date, the COB28, COB36, COB37 and COB52 cell lines represent interruptions in unknown genes, and remain of much interest for further studies as candidate regulators of ES cell self-renewal.

The COB54 gene trap mutation was bred to homozygosity to observe whether any developmental phenotype would result from a disruption in both alleles of the ES cell-specific gene. While the established Southern analysis for the COB54 gene trap allele identified the expected Mendelian ratio for transgenic to wild type genotypes, in both 3 week old offspring and 10.5 dpc intercross embryos (Table 5.5), it did not accurately decipher the homozygote from heterozygote genotypes (Figure 5.9). RT-PCR analysis for the wild type and disrupted COB54 alleles subsequently confirmed the presence of blastocyst stage embryos homozygous for the gene trap allele, at approximately the expected Mendelian frequency (Figure 5.10; Table 5.6). A viable and fertile homozygote COB54 mouse line has since been established.

The apparent lack of any overt phenotype in animals homozygote for the COB54 insertion may underscore the reason why this line was the only one to successfully be passed to the germline in chimaeras. While it remains unknown whether the novel

Fbx15 gene is developmentally important, this prospect warrants further investigation given the emerging roles for F-box proteins in cell regulatory proteolysis and the tightly regulated expression of this gene in ES cells.

There are a number of possible reasons for an apparent lack of phenotype in COB54 homozygous animals including the carboxy terminal insertion of the gene trap in the open reading frame for the Fbx15 protein (Figure 5.4). The 38 residue F-box motif, which is likely to bind an E3 ubiquitin core ligase complex, is at the 5' end of the open reading frame and its function may not be disturbed by the downstream insertion. It would seem that the insertion has also not interrupted the function of any as yet unidentified protein-interaction domains for substrate binding, downstream of the F-box motif. While it is not possible to predict the position or draw conclusions about potential disruptions to these domains, analysis of other mammalian F-box proteins with known downstream domains (Cenciarelli *et al.*, 1999; Winston *et al.*, 1999a) would suggest a carboxy terminal gene trap insertion, would in many cases be unlikely to disrupt upstream protein function. It is also possible, as previously discussed in Chapter one (section 1.6.5), that a mutant gene trap effect may in some cases be compensated for by the presence of low levels of wild type transcript or protein arising as a result of splicing around the gene trap event, or a failure to make use of the gene trap construct's polyadenylation signal (Faisst and Gruss, 1998; McClive *et al.*, 1998; Sam *et al.*, 1998; Voss *et al.*, 1998a). Finally, it remains possible that a redundancy in *Fbx15* gene function accounts for the lack of phenotype observed for COB54 homozygotes.

In summary, the results in this chapter describe the identification of a gene trap insertion for a recently identified gene, *Fbx15*, which shows a strong ES cell-restricted expression pattern *in vitro*, and belongs to a new mammalian family of F-box proteins with an emerging role in cell regulatory processes.

Chapter Six

A DETAILED EXPRESSION PROFILE FOR THE COB54 GENE TRAP INSERTION IN ES CELLS AND TRANSGENIC MICE

6.1 INTRODUCTION

Results presented in Chapter five confirm a clearly ES cell-restricted expression profile for the COB54 reporter gene insertion in a new member of the F-box gene family. This tightly restricted *in vitro* expression profile (Figures 5.1 & 5.2) and a potential regulatory role for the endogenous gene in controlling ES cell growth and self-renewal via substrate-specific ubiquitin-mediated proteolysis (section 5.9), makes this cell line and the entrapped endogenous gene a particularly interesting candidate for further investigation.

While the COB54 gene trap insertion in the *Fbx15* gene has not caused an obvious homozygote phenotype, the presence of the *βgeo* reporter gene under the control of the endogenous promoter in transgenic animals does provide a valuable histochemical reporter for simplified visualization of the entrapped allele's expression pattern. Although ES cells resemble early ICM cells in their behaviour (Beddington and Robertson, 1989), the dramatic downregulation of transgene expression upon differentiation of COB54 ES cells was not sufficient to accurately predict the loss of endogenous gene expression during embryonic development, or to exclude the possibility of renewed expression in more highly differentiated cell types within the complex *in vivo* environment. An expression profile for the embryonic stages preceding blastocyst and ICM formation was also of interest given the homology of the *Fbx15*

163

gene with an EST sequence previously identified from a mouse 2-cell cDNA library (Rothstein *et al.*, 1992). This chapter verifies native and fusion transcripts for the *Fbx15* cDNA, and presents a detailed and interesting expression profile for the COB54 gene trap allele.

6.2 NORTHERN ANALYSIS OF COB54 NATIVE AND GENE TRAP FUSION TRANSCRIPTS IN ES CELLS

To confirm that the *Fbx15* sequence derived by 5'-RACE from the COB54 gene trap clone represented the interrupted gene in transgenic mice generated from this cell line, poly A$^+$ mRNA prepared from the E14, COB36 and COB54 ES cell lines (sections 2.2.2f) was hybridized on a Northern blot (sections 2.2.4 & 2.2.6b) with the *Cob54Race* probe (sections 2.1.2 & 5.7.2, Figure 5.3). This *Fbx15* cDNA sequence was observed to hybridize with two native transcripts of approximately 1.9 kb and 2.5 kb in all three cell lines, and with an additional transcript of approximately 6.6 kb in the COB54 cell line, representative of the gene trap allele (Figure 6.1A). Re-hybridization of the blot with a probe "*Gapdh*", corresponding to a 0.8 kb fragment of *Gapdh* coding sequence (section 2.1.2), provided a loading control for all RNA samples (Figure 6.1B). Approximate quantitation of signals by phosphoimage analysis (data not shown), suggested that for each of the three cell lines the ratio between the two native transcripts was approximately the same, and for the COB54 cell line the combined signal for the two native transcripts was about half of that for the two cell lines expressing only the endogenous alleles.

6.3 *IN VIVO* EXPRESSION PROFILE FOR THE COB54 INSERTION

To investigate the *in vivo* expression profile for the *Fbx15* gene, pre- and post-implantation stage embryos, neonates and adults were analysed by X-gal staining for expression of the COB54 *lacZ* insertion and by RT-PCR amplification for the *Fbx15* mRNA transcript. The ZIN40 transgenic mouse strain (section 2.1.6) provided a control for marked and ubiquitous nuclear localized *lacZ* expression in all embryonic and adult tissues, while wild type embryos and tissues provided a control for non-specific endogenous cytoplasmic βgal activity.

6.3.1 *lacZ* expression in preimplantation stage COB54 embryos

COB54 heterozygote adult males (section 5.7.2) were crossed with wild type females from C57Bl/6J x CBA matings (section 2.1.6). Prepubescent females were

164

A

6.6kb-
5.0
3.6
2.6
1.9
1.4
1.0

E14 COB36 COB54

B
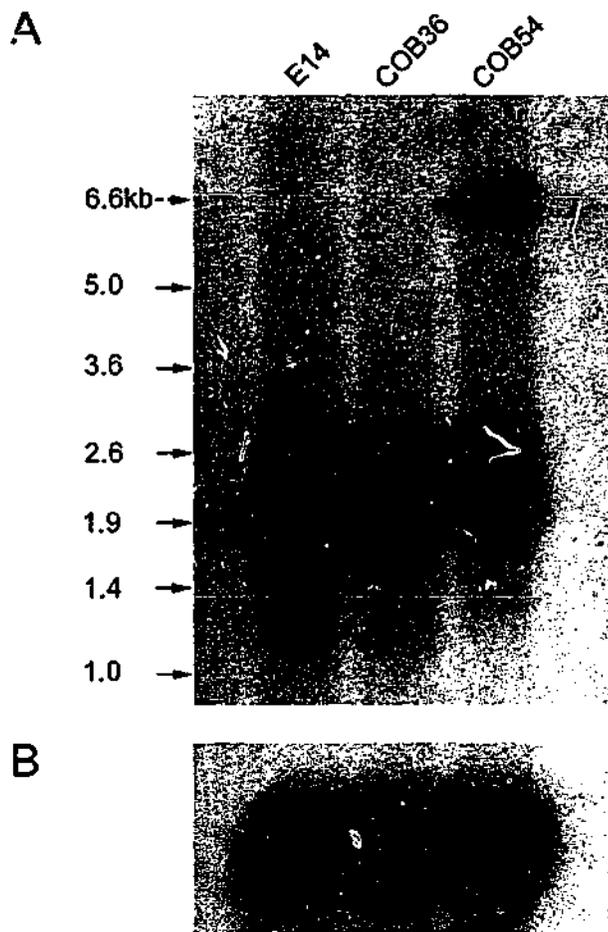
Figure 6.1: The detection of COB54 native and gene trap fusion transcripts in ES cells by Northern blot analysis. (A) For each of the E14, COB36 and COB54 ES cell lines, 3 µg of poly A⁺ mRNA was separated on a denaturing gel, and analysed by Northern blot hybridization with the *Cob54Race* probe. The probe's *Fbx15* cDNA sequence hybridized with two endogenous transcripts of approximately 1.9 kb and 2.5 kb in all cell lines, as well as with a single gene trap fusion transcript of approximately 6.6 kb in the COB54 cell line (see Figure 6.10 also, showing greater separation of the native transcripts). (B) Re-hybridization of the blot with the *Gapdh* probe served as an RNA loading control for each sample. Autoradiographic exposure was for 6 days (A), and 3 h (B), at -80°C with intensifying screens.

superovulated prior to matings (section 2.4.1) and generally yielded 20-40 embryos per litter. Fertilized oocytes were collected from the ampullary region of oviducts at 0.5 dpc (section 2.4.1), cultured *in vitro* and fixed at representative preimplantation stages of development for X-gal staining (sections 2.4.3a & 2.6.2). Alternatively, 3.5-4.5 dpc blastocysts were retrieved by flushing embryos from the uterine horns and fixed immediately for X-gal staining (sections 2.4.1 & 2.6.2). COB54 heterozygote females (section 5.7.2) were superovulated for the collection of unfertilized oocytes, to assess for any maternal expression of the gene during oocyte development. Heterozygous COB54 females were also mated with wild type adult males to evaluate preimplantation stage embryos as above.

Staining for *lacZ* expression in ZIN40- and COB54-derived oocytes confirmed there is no detectable βgal activity in the ovulated transgenic COB54 oocyte, nor in surrounding cumulus cells (Figure 6.2A & B). Similarly, no *lacZ* expression was observed in the 2PN and early 2-cell stage embryos (Figure 6.2C) derived from either male or female COB54 heterozygote parents. However, by the late 2-cell stage nuclear *lacZ* expression is seen at marked levels (Figure 6.2D), and is consistent with activation of gene expression at the major onset of zygotic transcription. Embryos cultured for assessment of expression at the first cleavage stage were fixed for X-gal staining at either 30-32 h, or 38-40 h postfertilization (section 2.4.1). Staining of the earlier *in vitro* 2-cell group detected βgal activity in only 3/54 embryos, from a female heterozygote parent in each case, with 2 showing mild expression and the other a moderate level of expression in both cells. For the late *in vitro* 2-cell group, 44/100 embryos stained positive for βgal activity, with all but 5 of the 44 showing a marked level of expression in both cells (Figure 6.2D). For these 5 embryos, derived from both female and male heterozygotes, 3 displayed a mild level of *lacZ* expression and 1 displayed moderate expression in both cells, while the remaining embryo showed mild expression in one cell and marked expression in the other. These results suggest that the expressed protein is a product of the embryonic genome, and that the developmental asynchrony associated with *in vitro* culture is likely to account for the observations seen between the early and late 2-cell stages.

The marked *lacZ* expression pattern appearing at the late 2-cell stage continues throughout preimplantation development (Figure 6.2E-H), with 4-cell, 6-8 cell, compacted morula and blastocyst stage embryos all staining strongly positive for βgal

165
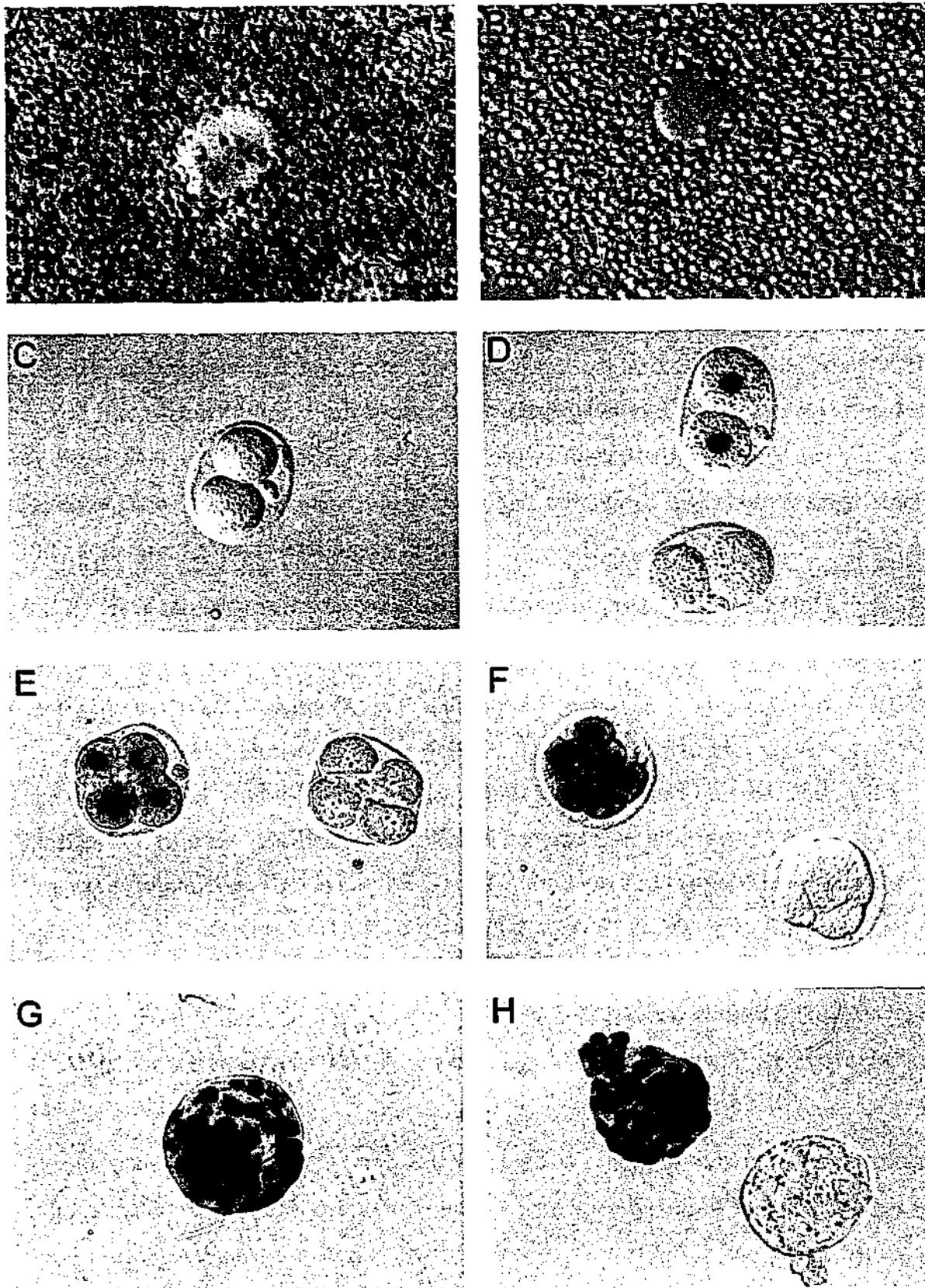
Figure 6.2: *The lacZ expression profile in COB54 oocytes and preimplantation embryos.* Heterozygote COB54 ovulated oocytes and embryos were fixed and stained with X-gal to detect βgal activity. Heterozygote ZIN40 and wild type oocytes and embryos were stained as positive and negative controls, respectively. Compared with a ZIN40 oocyte (A) no *lacZ* expression is seen in the mature COB54 oocyte (B), nor in the COB54 early 2-cell embryo following fertilization (C). By the late 2-cell stage, COB54 embryos display a marked level of nuclear *lacZ* expression (D), continuing through the preimplantation cleavage stages (E, F) to the blastocyst stage, in all cells of both the ICM and TE lineages (G). As the expanded blastocyst hatches from the zona, this strong expression is maintained in all cells (H). Non-staining wild type embryos are shown (D-F, H). Photographed under phase-contrast optics, 400x.

activity. Interestingly, in both the expanded and hatching blastocyst stage embryos, marked expression is seen in all cells of both the ICM and TE lineages (Figure 6.2H).

Preimplantation expression analyses identified healthy adult males that transmitted the transgenic allele to 100% of embryos in litters arising from superovulation, while others transmitted to the expected 50% (data not shown). These results confirmed evidence from the earlier molecular studies supporting the presence of viable and fertile homozygote animals generated from heterozygote intercrosses (section 5.8). To identify and maintain homozygote females for establishing a homozygote COB54 mouse line, weaned litters from backcross matings with wild type males were analysed by *βgeo* PCR analysis (section 5.7.1) for 100% transmission of the gene trap allele. Homozygosity was further confirmed by X-gal staining of a blastocyst stage embryo litter at the time of sacrificing these females.

6.3.2 *lacZ* expression in postimplantation stage COB54 embryos

The developmental expression pattern for the COB54 allele following implantation was analysed by whole-mount X-gal staining of embryos recovered from COB54 homozygote male and wild type (C57Bl/6J x CBA) female crosses. ZIN40 heterozygote and wild type (C57Bl/6J x CBA) males were crossed with wild type females to provide positive and negative controls, respectively. Embryos were dissected from uterine decidua at the time of gastrulation and from the uterus for stages up to 15.5 dpc (section 2.4.1), then fixed and stained according to their age (section 2.6.2). For embryos 12.5 dpc and older, some of each litter were cut sagitally to improve fixative and stain penetration. Whole-mount stained embryos were subsequently sectioned after paraffin embedding, and counterstained with eosin (section 2.6.3).

Results shown in Figure 6.3 confirm that there is a marked downregulation of *lacZ* expression in the COB54 transgenic embryo immediately following implantation. In the 6.5 dpc gastrulating embryo, weak βgal activity is detected only in the proximal extraembryonic region of the embryo (Figure 6.3A,B), with some staining seen in the VE cells and a few adjacent trophoblastic cells, but not throughout the ectoplacental cone (Figure 6.3C). By 7.5 dpc there is a complete absence of βgal activity in the embryo (Figure 6.3D,E), which is not detected again during developmental stages to 15.5 dpc (Figure 6.3F-H). Isolation of 5.5 dpc embryos from uterine decidua was

Figure 6.3: The lacZ expression profile in COB54 postimplantation embryos. Heterozygote
COB54 embryos were recovered at 6.5-15.5 dpc for whole-mount X-gal staining. Heterozygote
ZIN40 and wild type embryos were stained as positive and negative controls, respectively.
Embryos are shown at 6.5 dpc (A-C), 7.5 dpc (D,E), 9.5 dpc (F), 10.5 dpc (G), and 12.5 dpc (H-J).
Sections after paraffin embedding and eosin counterstaining are shown for 6.5 dpc (C) and 7.5
dpc (E) embryos, and for 12.5 dpc visceral yolk sac (I). ZIN40 control embryos show ubiquitous
staining at all stages (A, x3 embryos; J). In COB54 embryos, βgal activity is dramatically
downregulated by 6.5 dpc (A, x6 embryos). Mild staining is detected only in visceral endodermal
cells of the proximal extraembryonic region and a few adjacent trophoblastic cells, but not
throughout the ectoplacental cone (B,C). COB54-specific *lacZ* expression is not detected in the
embryo proper from 7.5 dpc onwards (D-H). COB54 expression is detected in extraembryonic
tissues from 10.5 dpc in columnar endodermal cells of the vascularised visceral yolk sac (I).
Photographed under light-field optics, 12.3x (A), 3.8x (F,G), 1.9x (H,J), 100x (D), 200x (B,C,E,I).
al=allantois; am=amnion; ch=chorion; ecp=ectoplacental cone; em=embryonic ectoderm;
ex=extraembryonic ectoderm; mes=embryonic mesoderm; tb=trophoblast; ve=visceral
endoderm; vys=visceral yolk sac.

A
COB54   ZIN40

B

C
tb
ecp
ex
ve
em

D

E
mes   al   ch
tb
em   vys   ve
am

F

G

H

I

J
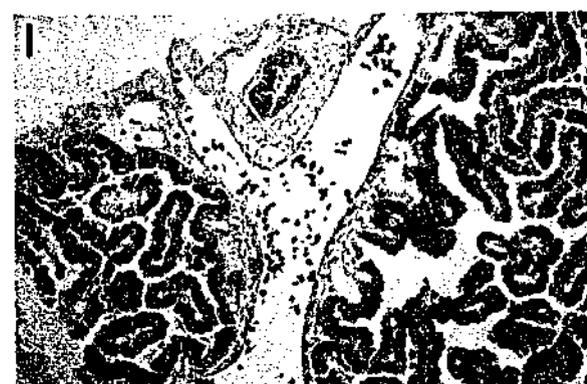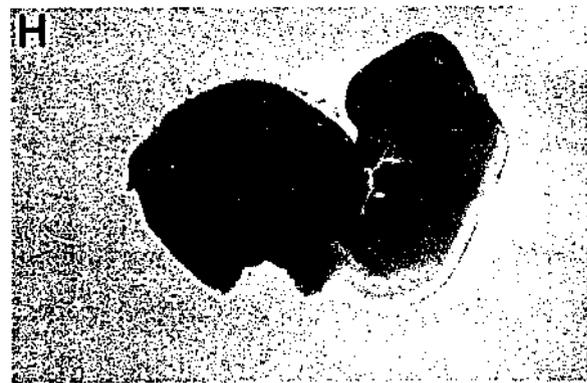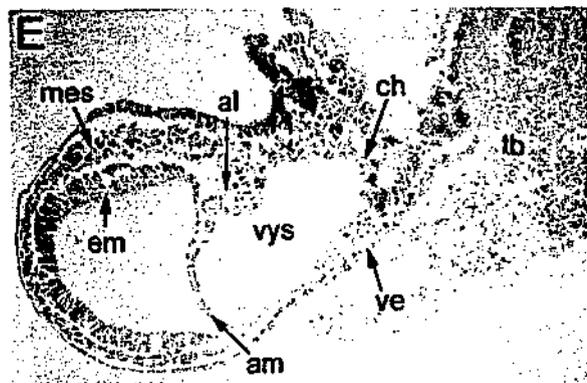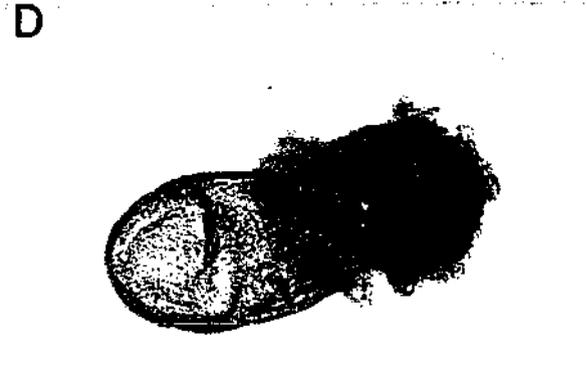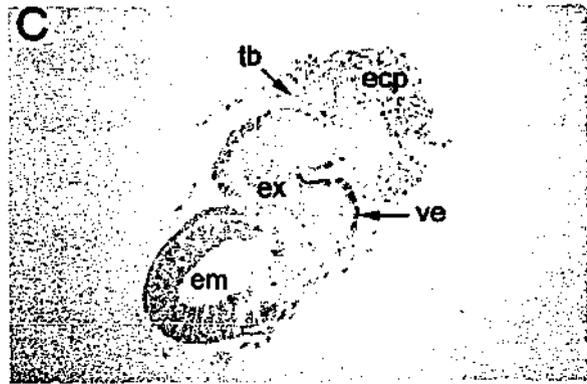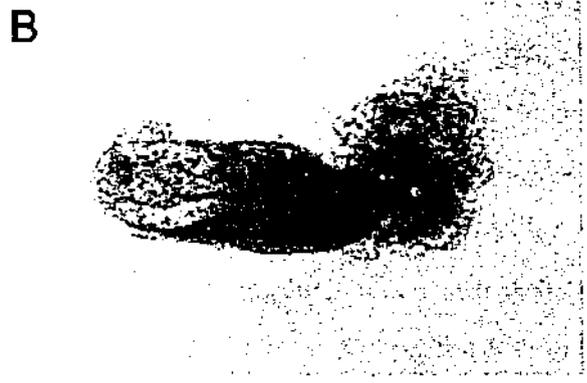
somewhat difficult, but X-gal staining of sagitally opened decidua did not detect any βgal activity in the embryonic region (data not shown). In the extraembryonic tissues, while a small degree of non-specific *lacZ* expression is detected in the maternal placenta, and more so in the Reichart's membrane covering the embryonic surface of the placenta, nuclear βgal activity is only evident in the columnar endodermal cells of the vascularised embryonic visceral yolk. This expression is first detected in a few cells at 10.5 dpc, and then throughout the yolk sac from 12.5 dpc onwards (Figure 6.3H,I). Weak βgal activity detected in the antimesometrial region of 6.5 ad 7.5 dpc dissected decidua and stronger activity in maternal uterine tissue, was also evident in negative control tissues and determined not to be associated with a transgenic contribution (data not shown).

6.3.3 RT-PCR confirmation of the embryonic expression profile for *Fbx15*

To confirm that the *lacZ* expression profile seen for the COB54 allele was representative of endogenous gene expression, an RT-PCR analysis for endogenous *Fbx15* mRNA was performed on single wild type oocytes and embryos through to 7.5 dpc. Assistance with this analysis from Dr. Rob Daniels (MIRD) is gratefully acknowledged. According to the protocol described in section 2.2.7c, RNA lysates for preimplantation stages and mRNA preparations for postimplantation stages were reverse transcribed from random primers, and cDNA amplified in 65 cycles ($T_a = 52°C$) with the previously used gene-specific primers COB54F1 and COB54R1 (section 2.1.3; Figure 5.3). RT-PCR analysis of *Fbx15* expression as demonstrated by the appearance of the expected 817 bp *Fbx15* cDNA product is shown in Figure 6.4A. Amplification of cDNA with *poly A polymerase* PCR primers (section 2.2.7b) controlled for template amount, and allowed a comparison of PCR positive signals between embryonic stages (Figure 6.4B). The downregulation of *poly A polymerase* transcripts following fertilization and its upregulation from the 4-cell stage, is as normally observed for the onset of embryonic transcription for this gene (R. Daniels, MIRD; *pers. comm.*). While the *Fbx15* gene is not maternally expressed, it does appear to be activated early at the maternal zygotic transition, with weak signals seen in 2-cell embryos at both 32 h and 38 h postfertilization, and a subsequent upregulation from the 4-cell to blastocyst stage. From the early to hatching blastocyst stages, although a reduction is seen in template *poly A polymerase*, the corresponding decrease in *Fbx15* expression appears slightly more so. No *Fbx15* expression was detected for a whole 5.5 dpc deciduum, although it

167
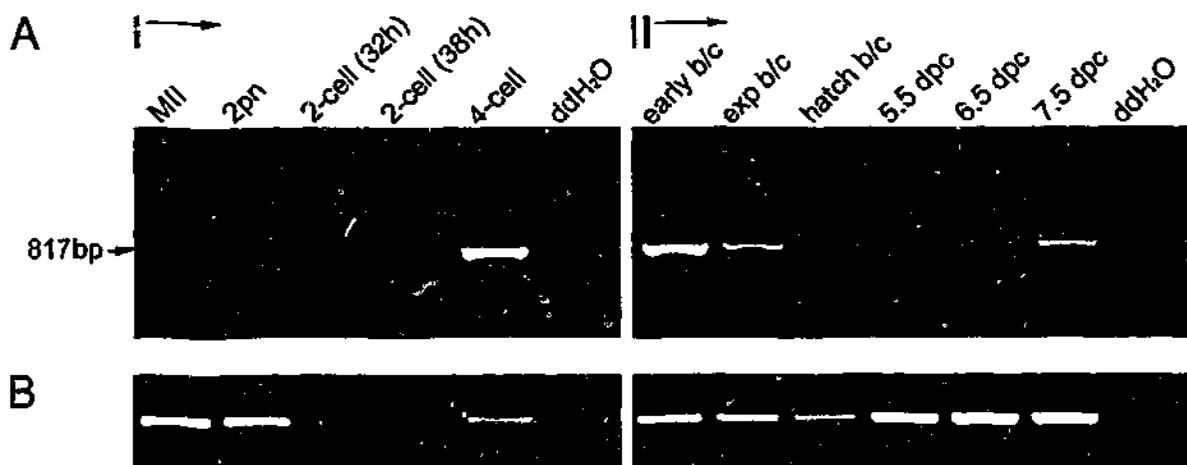
Figure 6.4: Analysis of Fbx15 expression in the early mouse embryo by RT-PCR. cDNA templates were generated from RNA lysates for single wild type oocytes and 0.5-7.5 dpc embryos with random primer RT. Gels I-II: (A) Following 65 PCR cycles, an 817 bp product corresponding to the *Fbx15* gene was detected for - the 4-cell embryo, the early, expanding (exp) and hatching (hatch) blastocyst (b/c) stages, only weakly in 32 h and 38 h 2-cell embryos, but not in the mature oocyte (MII) or zygote (2pn). For the postimplantation stages, no product was detectable for a whole 5.5 dpc deciduum (5.5 dpc), but a weak signal was detected for 6.5 dpc and 7.5 dpc stage embryos. (B) A control *poly A polymerase* product was amplified from the same amount of each cDNA template, and both assays included a water template control (ddH$_2$0). As for the normally observed embryonic onset of *poly A polymerase* transcription (BI), the weak detection and upregulation of *Fbx15* transcription from the 2-cell stage (AI-II) is consistent with early embryonic activation for this gene. The reduction in *Fbx15* transcription from the early b/c stage onwards, appears marked compared with corresponding *poly A polymerase* product for each template.

remains possible that this did not contain an embryo. *Fbx15* expression is detected in both the 6.5 dpc and 7.5 dpc stage embryos by PCR, but this expression is clearly downregulated from that observed for the blastocyst stages by comparison with the *poly A polymerase* template controls.

6.3.4 *lacZ* expression in implantation-delayed COB54 blastocysts

To investigate if the downregulation of COB54 expression was time- or implantation-dependent, *lacZ* expression was analysed in blastocyst stage embryos for which implantation had been artificially delayed. Plugged wild type females mated with homozygote COB54 males were ovariectomized at 2.5 dpc, and administered progesterone to render the uterus non-receptive to embryo implantation (section 2.4.2). Zona-free hatched blastocysts were gently flushed from the uteri (section 2.4.1) at 6.5 dpc and 8.5 dpc, and fixed for X-gal staining (section 2.6.2). Control ZIN40 heterozygote implantation-delayed embryos were assessed in the same way. As shown in Figure 6.5, strong *lacZ* expression, comparable to that previously demonstrated for 3.5-4.5 dpc blastocyst stage COB54 embryos (Figure 6.2G,H), is maintained in all cells of both 6.5 dpc and 8.5 dpc implantation-delayed embryos. Although in the latest stage embryos the blastocoel cavity has collapsed, the expression profile is very clear.

6.3.5 *lacZ* expression in adult COB54 tissues

Tissues were dissected from wild type and COB54 heterozygote males and females at 7-10 weeks of age, and fixed for whole-mount X-gal staining (sections 2.4.4 & 2.6.2). Where required, tissue pieces were cut to a size suitable for fixative and stain penetration, which was enhanced with the addition of detergents (section 2.6.2). Whole-mount stained tissues were paraffin embedded for sectioning and counterstained with eosin (section 2.6.3). Tissues assessed for *lacZ* expression included testis, ovary, brain, heart, lung, liver, kidney, spleen and fat.

The only adult tissue to display definitive COB54-specific *lacZ* expression was the testis. Both macroscopically and histologically, strong staining for βgal activity is seen in COB54 seminiferous tubules (Figure 6.6A-C), as opposed to the endogenous cytoplasmic βgal activity normally seen only in the interstitial tissue of wild type testis (Figure 6.6D). With reference to *Histological and histopathological evaluation of the testis* (Russell *et al.*, 1990) and with the assistance of Dr. Kate Loveland (MIRD), the COB54-specific staining was located in the spermatocyte populations, commencing
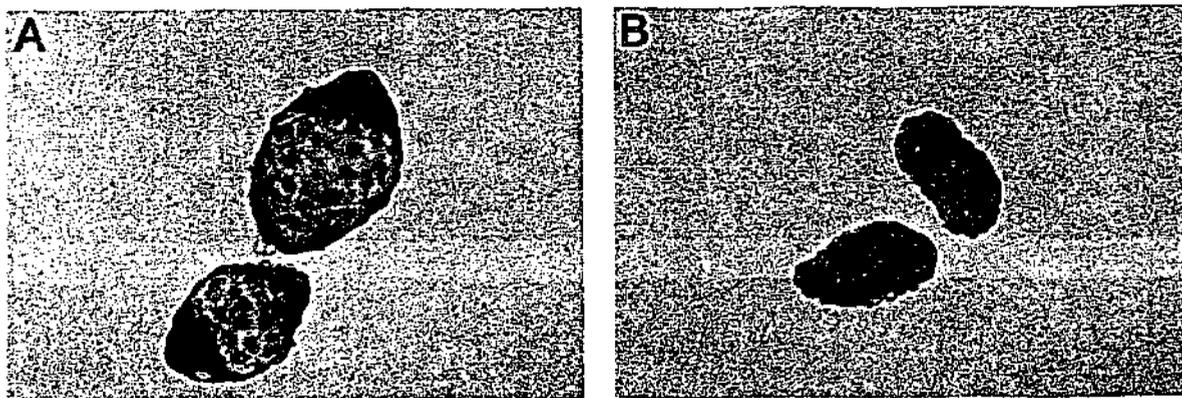
Figure 6.5: The lacZ expression profile in COB54 implantation-delayed embryos. Heterozygote COB54 embryos were flushed from ovariectomized females at 6.5 dpc (A) and 8.5 dpc (B) and stained with X-gal. Embryos at both stages have hatched from the zona and show marked *lacZ* expression in all cells, of both the ICM and TE lineages. In the later implantation-delayed stage embryos, the blastocoel cavity has collapsed. Photographed under light-field optics, 400x.
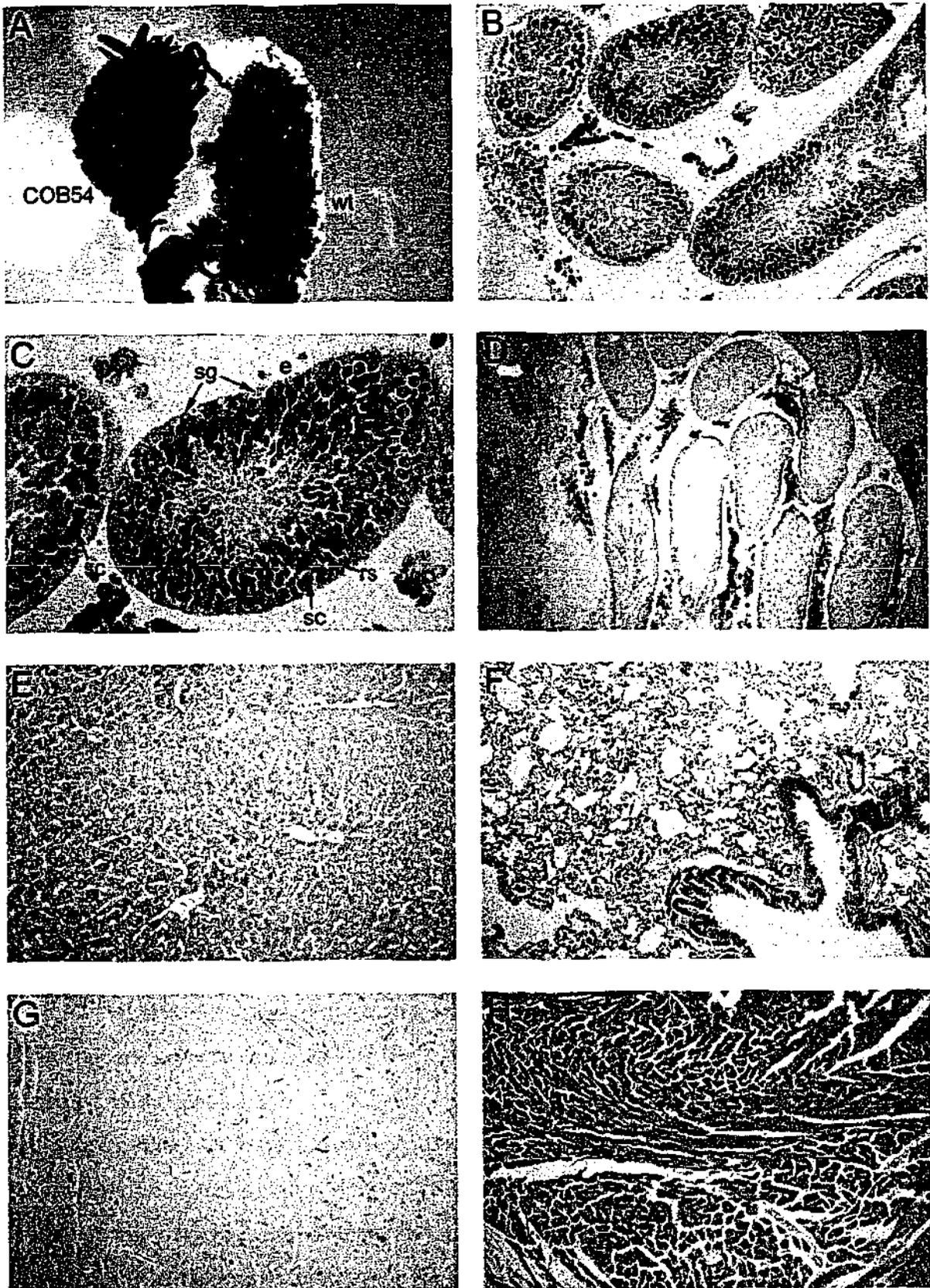
Figure 6.6: The lacZ expression profile in COB54 adult tissues. COB54 heterozygote adult mice were dissected at 7-10 weeks for whole-mount X-gal staining of tissues. Compared with wild type tissues (wt), COB54-specific *lacZ* expression is seen definitively in whole-mount testis (A). Eosin counterstained sections are shown for COB54 testis (B,C), wt testis (D), and COB54 liver (E), lung (F), brain (G) and heart (H). Strong staining in the seminiferous epithelia of the COB54 testis (B,C) is associated with the spermatocyte (sc) and round spermatid (rs) populations, but not with the spermatogonial germ cells (sg) and the basal lamina on which they lie, nor with the differentiated elongate spermatids (e). Non-specific staining is seen only in the interstitial tissue of wt testis (D). Except for a few cells in lung alveolar tissue (F), no staining is evident in the other tissues (E,G,H). Photographed under light-field optics, 2.5x (A), 100x (D), 200x (B,E-H), 400x (C).

some time in the mid-pachytene stage, prior to the first meiotic division. Staining is also seen in the round spermatids resulting from the second meiotic division, but not in the differentiated elongate spermatids. In staged transverse tubule sections (Russell *et al.*, 1990), staining in the spermatid cells appears less intense than for the spermatocytes in some tubules, while in others it is equally intense for both populations. No staining is seen in the basal lamina cells lining the tubules, nor in the spermatogonial germ cells along this basal layer (Figure 6.6C). Of the other tissues examined, strong staining for non-specific βgal activity was seen throughout ovarian tissue, and also in the epithelial lining of the nephric tubules in kidney tissue, for both COB54 and wild type tissues. A few occasional cells associated with the alveolar walls in adult lung tissue appeared to stain nuclear positive for βgal activity (Figure 6.6F).

6.3.6 *lacZ* expression in day 2 neonate COB54 tissues

A series of neonatal tissues were dissected from wild type and COB54 homozygote animals 2 days after birth (section 2.4.4) and fixed for whole-mount X-gal staining as described in section 6.3.5. Tissues were sectioned following paraffin embedding and counterstained with eosin (section 2.6.3). Tissues assessed for *lacZ* expression included testis, ovary, brain, heart, lung, liver, kidney, spleen, smooth muscle, skeletal muscle, colon, small intestine, stomach, tongue and skin. The lung and brain whole-mount tissues show a degree of βgal activity greater than that for wild type tissues (Figure 6.7A,B). Histologically, the COB54 lung tissue shows strong blue staining associated with cells within the parenchyma around the terminal respiratory sacs, and developing alveoli and alveolar ducts. Although a similar staining pattern is seen in wild type lung, this appears less marked than for the COB54 lung (Figure 6.7C,D). Occasional nuclear staining cells are seen by histological examination of COB54 neonate brain tissue (Figure 6.7E), but not in wild type neonate brain. Other tissues, including the prepubescent testis, did not show COB54-specific *lacZ* expression (Figure 6.7F-H). Non-specific staining was seen for the intestinal sections, on account of their bacterial content, and in the kidney tubules, as for the adult.

6.3.7 *lacZ* staining in 11.5-13.5dpc COB54 genital ridges

Genital ridges and their mesonephroi were dissected from heterozygote COB54 litters at 11.5 dpc, 12.5 dpc and 13.5 dpc (section 2.4.4), and fixed for whole-mount X-gal staining as described for the tissues (section 2.6.2). Similarly, these were sectioned and
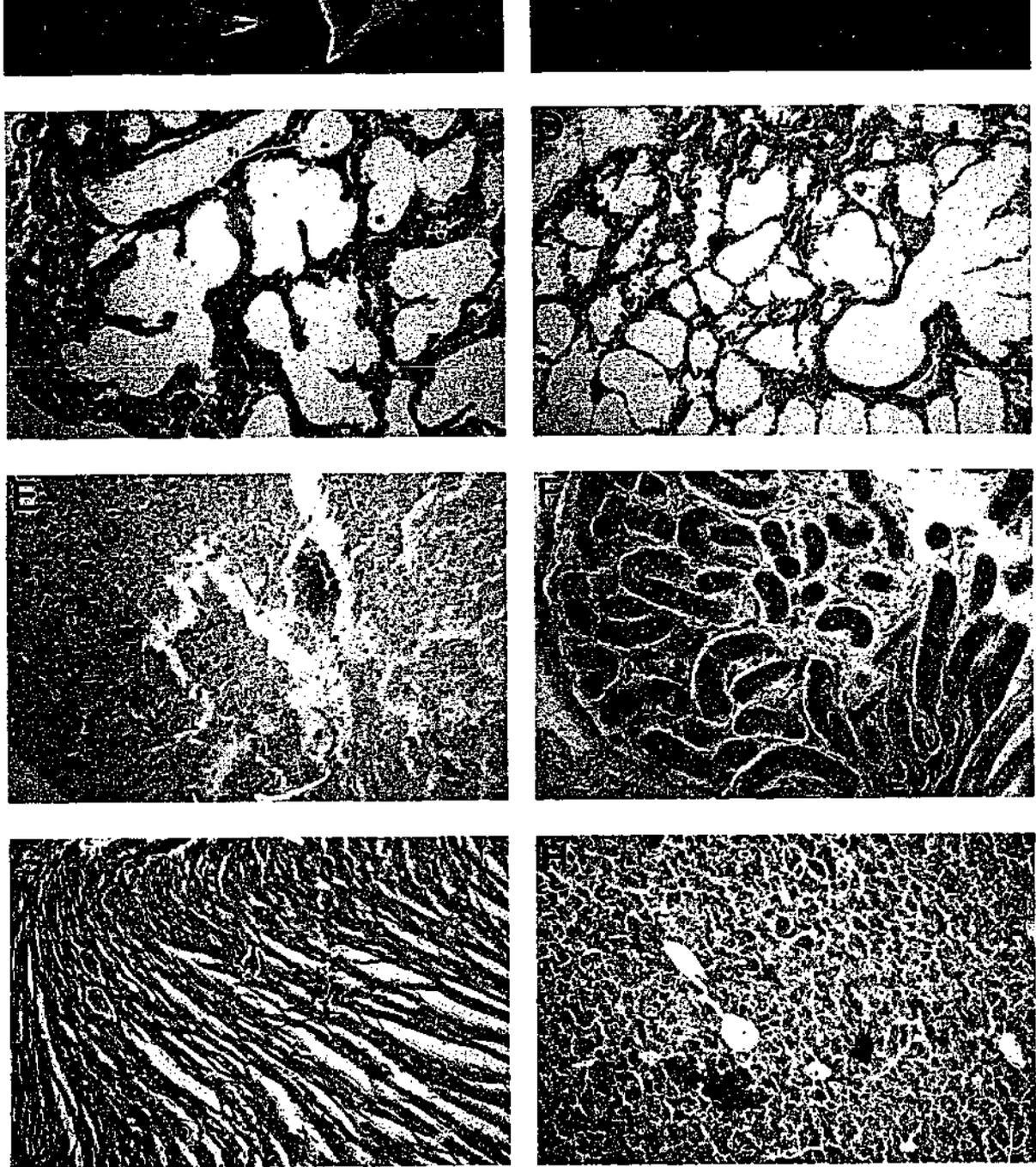
Figure 6.7: The lacZ expression profile in COB54 day 2 neonate tissues. COB54 homozygote offspring were dissected at day 2 for whole-mount X-gal staining of tissues. Compared with wild type tissues (wt), COB54-specific βgal activity is detectable in neonate lung (A) and brain (B) whole-mount tissues. Eosin counterstained sections are shown for COB54 (C) lung, (E) brain, (F) testis, (G) heart and (H) liver tissues, and for (D) wt lung. Non-specific staining in the wt lung appears less marked than the gene-specific expression seen in the COB54 lung parenchyma around the alveoli and alveolar ducts (C,D). Occasional nuclear staining cells are seen in COB54 neonate brain (E), but not in the other tissues (F-H). Staining seen in the interstitial tissue of the COB54 neonate testis is non-specific (F). Photographed under dark-field optics, 3.8x (A,B); light-field optics, 200x (C-H).
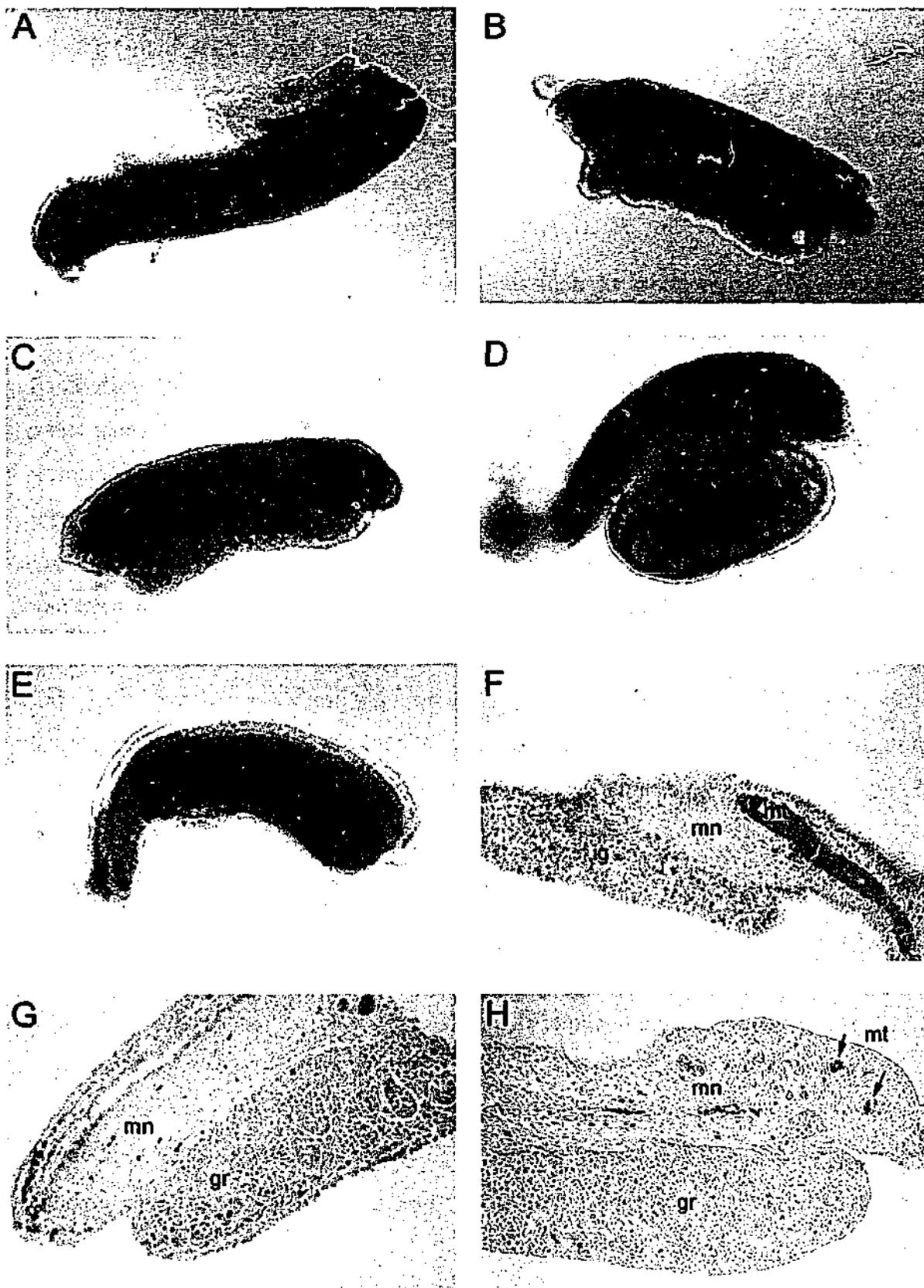
Figure 6.8: The lacZ expression profile in COB54 genital ridges. Whole-mount X-gal staining did not detect any COB54-specific βgal activity in: (A,F) the 11.5 dpc indifferent gonad, or the (B,G) 12.5 dpc male, (C) 12.5 dpc female, (D) 13.5 dpc male, or (E,H) 13.5 dpc female - genital ridges and overlying mesonephros from COB54 heterozygote embryos. Eosin counterstained sections are shown for (F) 11.5 dpc indifferent, (G) 12.5 dpc male, and (H) 13.5 dpc female gonads. Non-specific βgal activity is seen in the proximal mesonephric tubules of the mesonephros for 12.5 and 13.5 dpc male and female embryos (H, arrows). Photographed under light-field optics, 100x (A-E), 200x (F-H). ig=indifferent gonadal primordium; gr=genital ridge; mn=mesonephros; mt=mesonephric tubule.

counterstained with eosin following paraffin embedding (section 2.6.3). No *lacZ* expression was detected in whole-mounts or sections for genital ridges at any of these stages, for either sex (Figure 6.8A-H). Some non-specific staining seen in a few cells in the proximal mesonephros of both male and female 12.5 and 13.5 dpc genital ridges was associated with the developing mesonephric tubules of the embryonic kidney tissue.

6.3.8 RT-PCR screen for *Fbx15* expression in embryonic, neonatal and adult tissues.

To confirm the results of histochemical analyses for reporter gene expression and to detect any low-level *Fbx15* expression, RT-PCR analysis for the endogenous gene was performed on a library of poly A$^+$ mRNA samples from wild type 8-18 dpc postimplantation embryos, day 2 neonatal and adult tissues (section 2.4.4). Each RNA template was reverse transcribed from random primers (section 2.2.7b) and cDNA amplified in 30 cycles (section 2.2.7a, denaturing and annealing times = 60 s, T_a = 56°C) with the gene-specific primers COB54F1 and COB54R1 (section 2.1.3, Figure 5.3). E14 ES cell poly A$^+$ mRNA was used as a positive RT-PCR control template for the expected 817 bp *Fbx15* product (Figure 5.3), while PCR amplification with *poly A polymerase* PCR primers provided an internal control for each cDNA template generated (section 2.2.7b).

Results in Figure 6.9 (I-III) show there is some detectable *Fbx15* transcript in the 8, 10 and 12 dpc whole embryo, while there is clearly expression in the developing placenta, seen most strongly in late gestation. Expression is also detected in both the late gestation and neonatal lung and brain, although in each case this is much less than for amplification of *poly A polymerase* transcripts in the same samples. Brain expression is seen most strongly for the neonate, while lung expression is comparitive in both the 18 dpc and neonate samples. The adult testis shows definitive and strong expression of the *Fbx15* gene compared with *poly A polymerase* amplification, confirming the significant expression detected by *lacZ* analysis (Figure 6.6). While there is a suggestion of very weak *Fbx15* expression in several other samples, the abundant *poly A polymerase* product generated for all samples indicates that these signals are likely to have resulted from an excess amount of cDNA template. Two RNA samples, adult lung and adult tongue, could not be amplified most probably due to poor quality or insufficient template (data not shown).
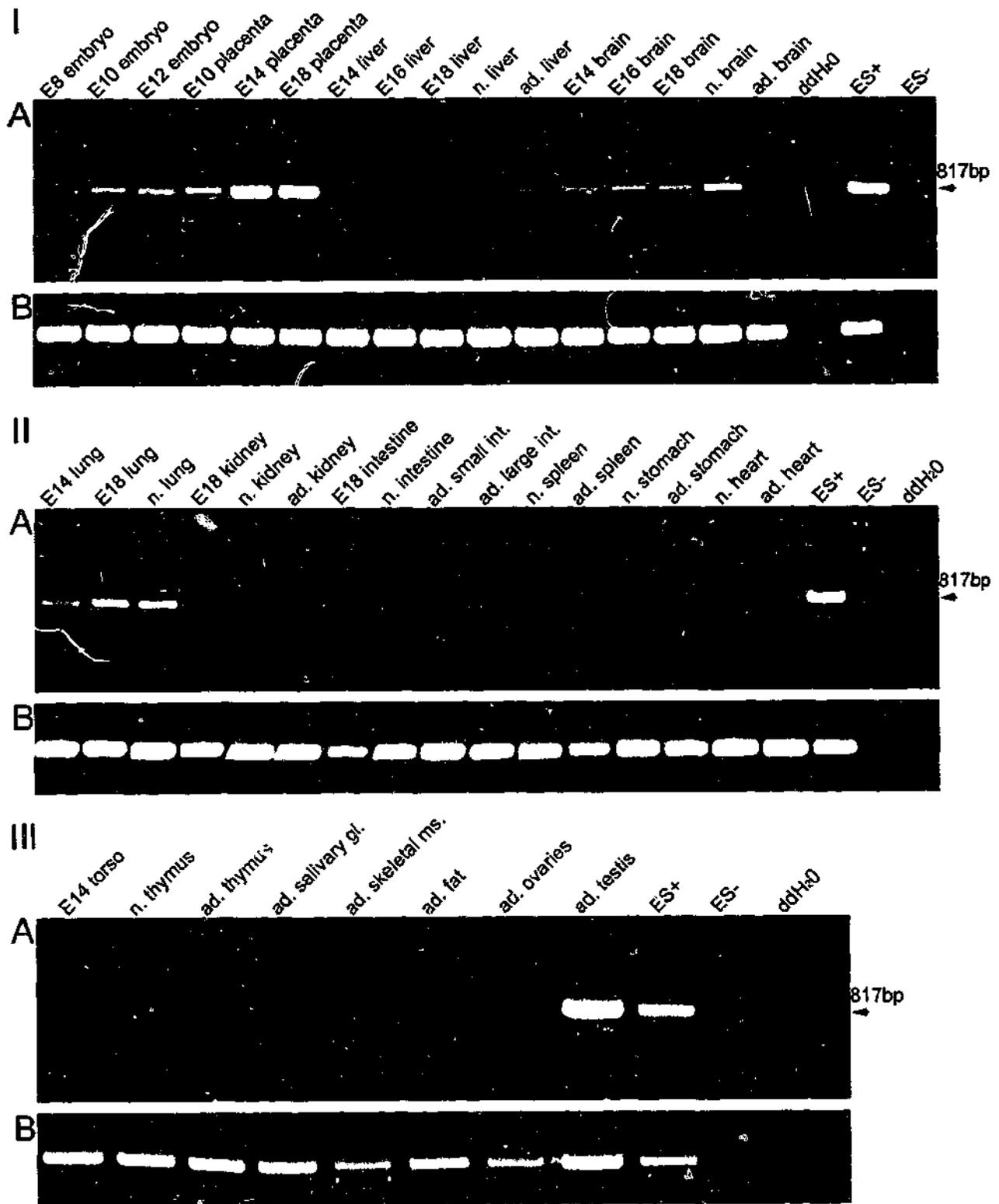
Figure 6.9: Analysis of Fbx15 expression in embryonic, neonatal and adult tissues by RT-PCR. cDNA templates were generated by random primer RT from ~1 μg each of wild type poly A⁺ mRNA samples for 8-18 dpc whole embryos (E8-18), and day 2 neonate (n.) and adult (ad.) tissues. Gels I-III show: (A) 30 PCR amplification cycles for detection of an 817 bp product corresponding to the *Fbx15* gene, against control E14 ES cell mRNA with (ES+) and without (ES-) RT enzyme, as well as a control water template with RT enzyme (ddH₂0). (B) For each cDNA sample, the same template amount amplified with primers for *poly A polymerase*. Compared with the abundant *poly A polymerase* transcripts amplified for each template, *Fbx15* expression is detected weakly in the E8-12 whole embryos and strongly in the late gestation placenta. Moderate *Fbx15* expression is detected for the late gestation and neonatal brain and lung, while the adult testis displays marked expression.

6.3.9 Northern analysis for the *in vivo* COB54 fusion transcript

The testis was selected as the most suitable tissue source for Northern blot analysis of mRNA sequences which hybridize with a probe for the COB54 5' entrapped sequence. Testicular poly A+ mRNA was prepared from 8-11 week old heterozygote and homozygote COB54 animals (sections 2.2.2f & 2.4.4). Control wild type testis total RNA was prepared from adult mice and ES cell poly A^+ mRNA from E14 wild type, and the COB36 and COB54 heterozygote gene trap cell lines (sections 2.2.2e & f, 2.4.4). A Northern blot was hybridized (sections 2.2.4 & 2.2.6b) with a probe "*Cob54Pcr*", corresponding to 817 bp of 5' entrapped *Fbx15* sequence lying 164 bp upstream of the gene trap insertion site (section 2.1.2). This probe was generated by RT-PCR amplification from E14 ES cell mRNA with the COB54F1 and COB54R1 primers (sections 2.2.7b & 2.1.3, Figure 5.3).

Northern blot hybridization results shown in Figure 6.10A confirm the expression of two *Fbx15* wild type transcripts of approximately 1.9 kb and 2.5 kb in the adult testis, as seen in ES cells (Figure 6.1). Importantly, this analysis also shows the absence of these transcripts in testis homozygous for the gene trap insertion, although with an increased gel RNA loading there is faint detection of an approximately 2.5 kb band which may represent the larger native transcript. The approximately 6.6 kb gene trap fusion transcript detected in COB54 ES cells (Figure 6.1) is also detected in COB54 transgenic testis tissue, but appears less abundantly expressed than the endogenous transcripts in heterozygote animals. The *Cob54Pcr* probe also detects a weak hybridization band just above the native transcripts in all transgenic testis mRNA samples, although it is not known whether this represents a tissue-specific transcript or a possible non-specific hybridization product. Re-hybridization of the blot with the *Gapdh* probe (section 2.1.2) provided a loading control for all RNA samples (Figure 6.10B).

6.4   *IN VITRO* EXPRESSION PROFILE FOR THE COB54 INSERTION

6.4.1 *lacZ* staining of blastocyst explant cultures

Blastocyst explant cultures were undertaken to investigate if expression for the COB54 gene trap allele would parallel postimplantation downregulation upon adherence and outgrowth *in vitro*. Blastocysts were collected at 3.5 dpc from superovulated wild type females mated with COB54 homozygote males, or control ZIN40 heterozygote males (section 2.4.1). Explant cultures were established and cultured as described in section

171
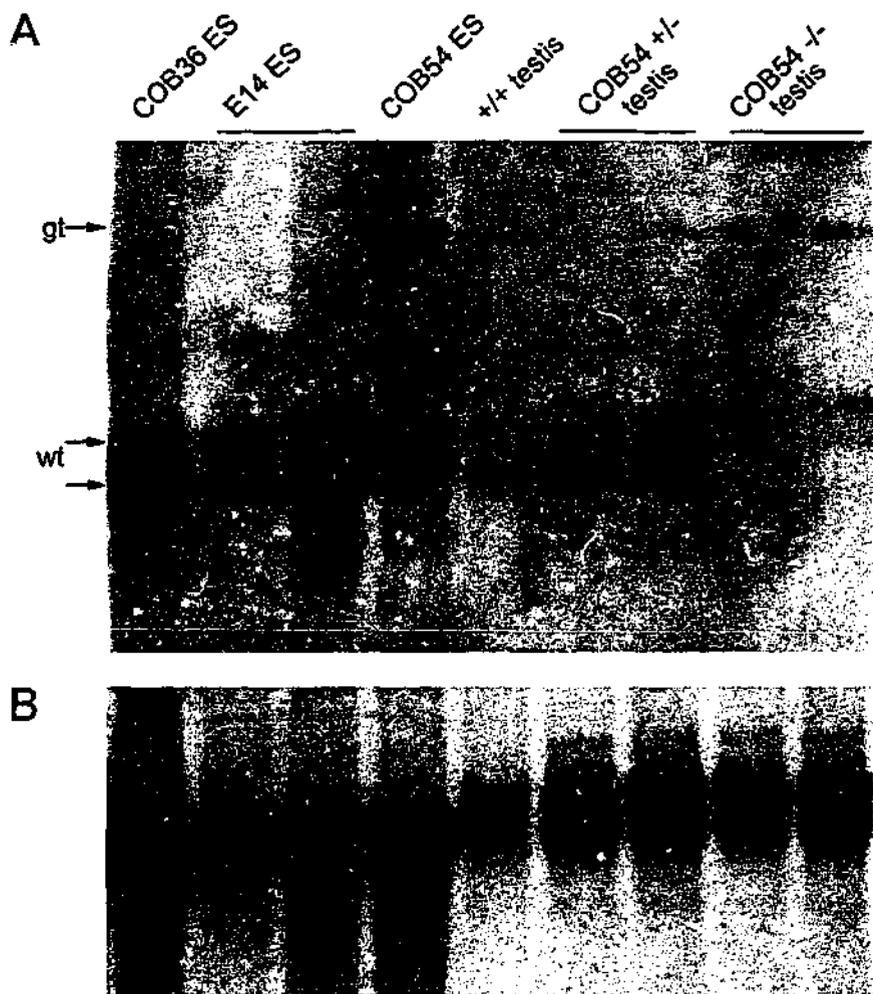
Figure 6.10: The detection of Fbx15 transcripts in the adult testes of COB54 mice by Northern blot analysis. Testicular poly A⁺ mRNA from adult mice heterozygote (+/-) and homozygote (-/-) for the COB54 gene trap allele, control total RNA from wild type (+/+) adult testis, and ES cell poly A⁺ mRNA from the COB36, E14 (~3 µg and a double loading) and COB54 ES cell lines (ES) were separated on a denaturing gel for Northern blot analysis. (A) Hybridization of the blot with the *Cob54Pcr* probe confirms *in vivo* testicular expression of the two wild type *Fbx15* transcripts (wt) previously detected in ES cells, and their absence in the COB54 -/- animal. However, the increased COB54 -/- RNA loading in one lane detects very low level expression of a band which may represent the larger wt transcript. The gene trap mRNA fusion transcript (gt) detected in COB54 ES cells is also evident in the transgenic animals, but appears less abundantly expressed. An extra weak hybridization band just larger than the wt transcripts is also detected in each transgenic testis RNA sample. (B) Re-hybridization of the blot with the *Gapdh* probe shows that the estimated gel loadings for tissue RNA are less than those for ES cell RNA. Autoradiographic exposure was for 3.5 days (A), and 24 h (B), at -80°C with intensifying screens.
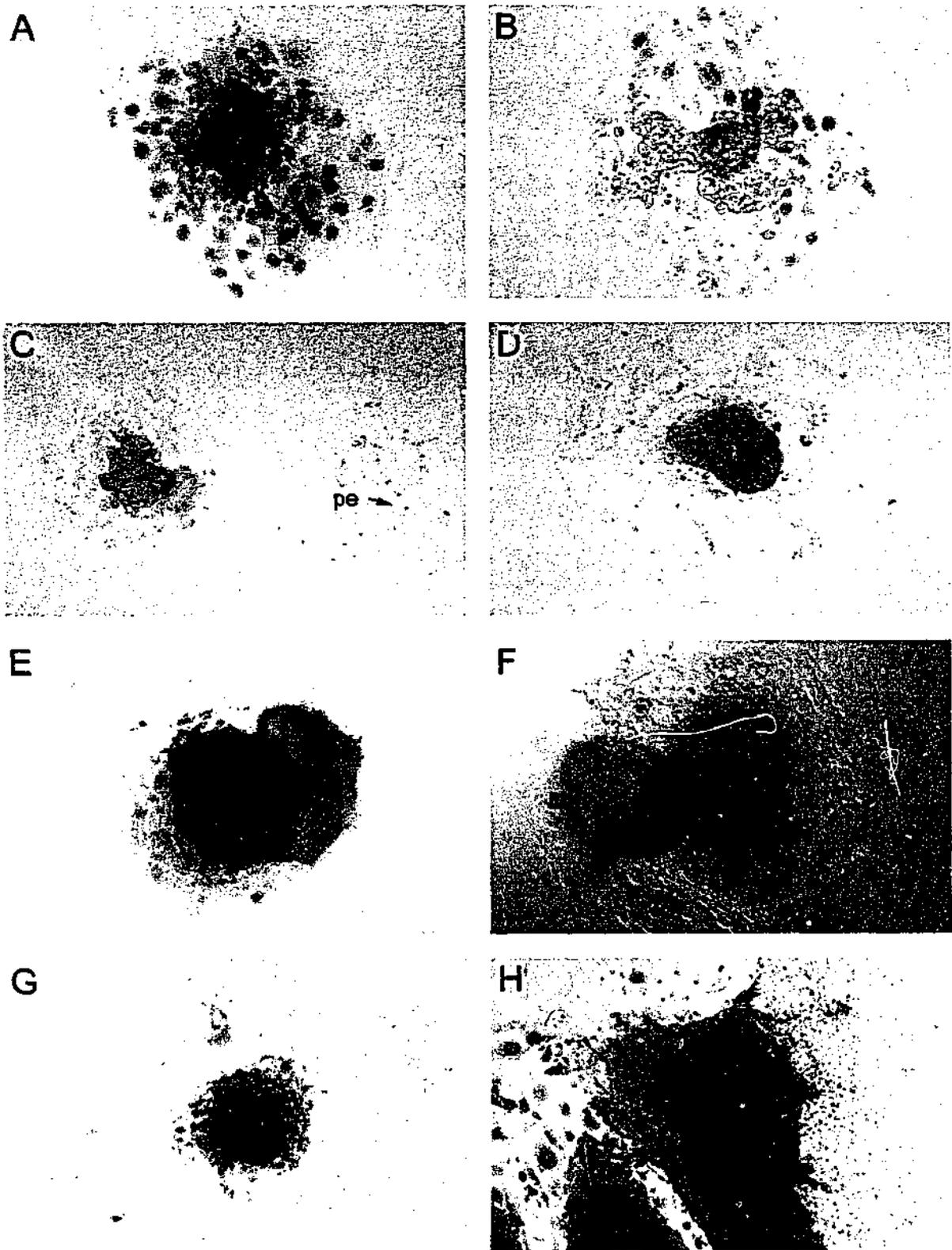
Figure 6.11: The lacZ expression profile in COB54 blastocyst explant cultures. COB54 heterozygote 3.5 dpc blastocysts were cultured for up to 6 days and stained with X-gal for βgal activity. Explants cultured without LIF (A-E) show: (A) Strong staining in all cells of the hatched embryo and TE outgrowths at 2 days post plating (dpp). (B) Marked downregulation of *lacZ* expression, with moderate staining in only a few outgrowth cells at 3 dpp. (C) Weak staining restricted to migratory parietal endoderm (pe) cells by 4 dpp. (D) Moderate staining is again evident in a few cells within the differentiating ICM-like clump at 5 dpp. (E) Upregulated and strong *lacZ* expression throughout the highly differentiating explant at 6 dpp. (F) Explants cultured with LIF maintained a larger ICM-like core, were less expansive and displayed a similar *lacZ* profile, with upregulated mild staining in a restricted population of ICM-like cells at 6 dpp. (G) Explants cultured with both LIF and G418 selected for a small subset of strongly staining cells from the ICM-like clump by 6 dpp. (H) Control ZIN40 heterozygote blastocysts, shown at 6 dpp, displayed marked βgal activity in all cell types at all stages. Photographed under light-field optics, 200x (A,B,G), 100x (C-E,F,H).

2.4.3b with the addition of LIF, with LIF and G418, or without LIF. Blastocyst explants were cultured for up to 6 days and fixed for X-gal staining (section 2.6.2) after each day of culture for COB54 explants, or after each second day for the control ZIN40 explants. Essentially all blastocysts hatched from their zonae within 24 h of plating and attached to the tissue culture surface during the second day of culture. Explants cultured without LIF expanded more quickly and showed a smaller ICM-like clump of cells within the first few days of culture, whereas those cultured with LIF maintained a larger ICM-like mass and were generally not as expansive. The ZIN40 explants maintained strong *lacZ* expression in all cells throughout all stages of culture (Figure 6.11H).

COB54 explants cultured with and without LIF showed a similar *lacZ* expression profile. During the first 2 days of culture, COB54 embryos maintained the strong *lacZ* expression pattern seen for 3.5 dpc blastocysts (Figure 6.2G; Figure 6.11A). However, after the 3rd day of culture, explants showed a marked downregulation in expression with moderate blue staining seen only in a few outgrowth cells, and in a small number of central ICM-like cells for those explants cultured in LIF (Figure 6.11B). After 4 days, mild staining was evident in parietal endoderm-like cells migrating away from the explants (Figure 6.11C). After 5 days, mild or moderate staining was again evident in a few cells within the differentiating ICM-like clump of cells (Figure 6.11D). After the 6th day, an upregulation of *lacZ* expression was quite apparent with strong blue staining seen throughout the highly differentiated explants cultured without LIF (Figure 6.11E), and mild blue staining seen in a subset of central cells in explants cultured with LIF (Figure 6.11F). For explants cultured with both LIF and G418, a similar growth and expression pattern was observed as for the LIF cultures, but after the 5th day there were few surviving outgrowth cells surrounding a small ICM-like core of cells. After the 6th day, only this small core of cells remained, for which a strong upregulation in *lacZ* expression was observed (Figure 6.11G).

6.4.2 *Oct-4* and *Fbx15* expression in differentiating ES cell aggregates

A time course Northern blot analysis was performed to compare expression of the *Fbx15* and *Oct-4* genes in differentiating ES cell cultures. E14 wild type ES cells were induced to differentiate in high density aggregation cultures in the absence of LIF as described in section 2.3.5b, and a time course Northern blot prepared for poly A^{+} mRNA from EBs collected after 2, 4, 6 and 8 days of culture (sections 2.2.2f & 2.2.4).

172
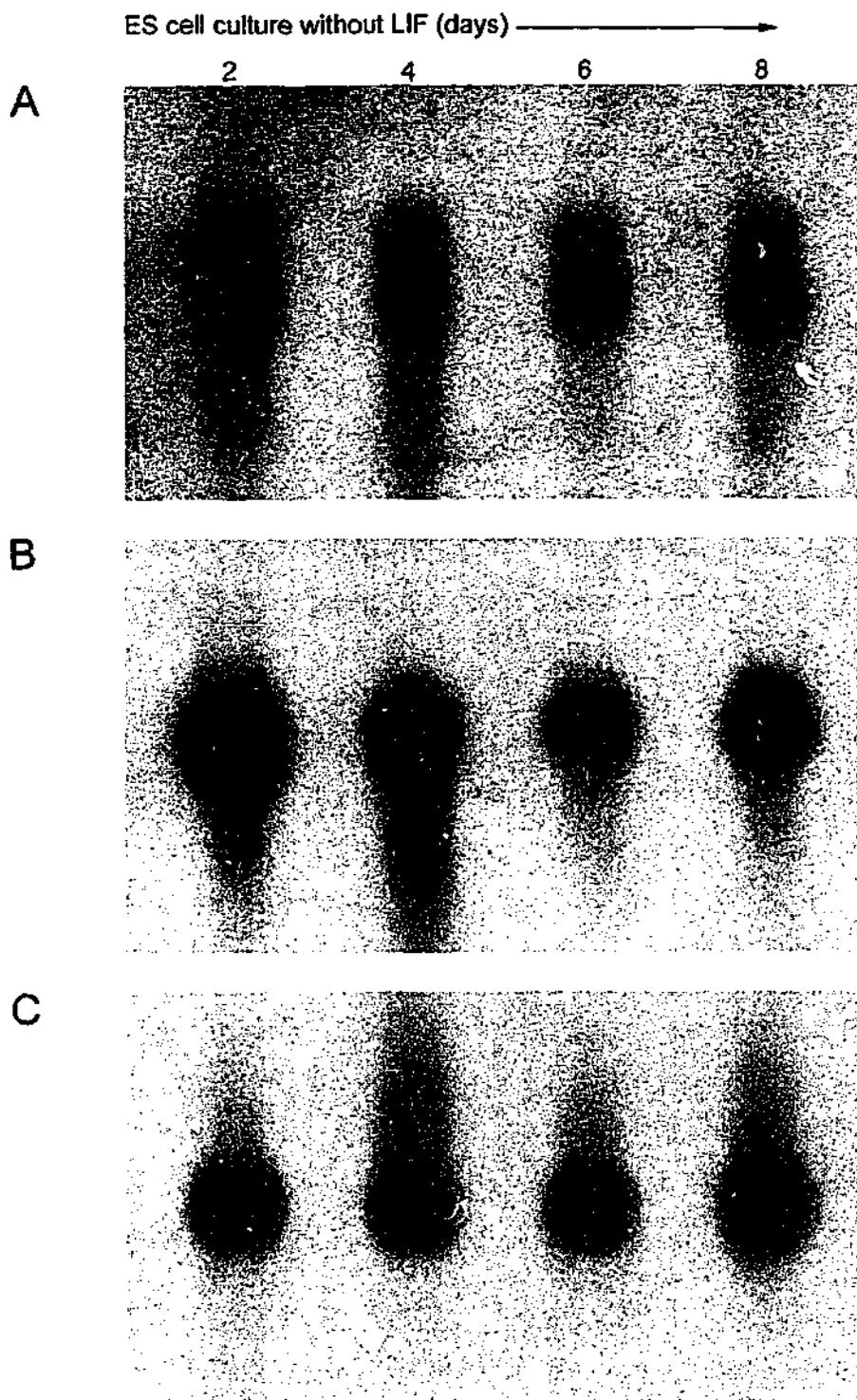
Figure 6.12: Northern blot analysis of Fbx15 and Oct-4 mRNA expression during ES cell differentiation. E14 wild type ES cell poly A⁺ mRNA (3 μg) from embryoid bodies cultured for 2, 4, 6 and 8 days without LIF, was separated on a denaturing gel for Northern blot analysis. The same blot was hybridized with (A) the *Cob54Pcr* probe specific to 817 bp of *Fbx15* sequence, (B) the *Oct4* probe specific to 1 kb of *Oct-4* coding sequence, and (C) the *Gapdh* probe for an RNA loading control. As for *Oct-4*, *Fbx15* expression is strong in ES cells, but appears to be downregulated at an earlier time point than for *Oct-4* in the absence of LIF. The slight upregulation for the 8 day sample with both ES cell-specific probes (A,B) is likely to be due to the expansion of residual ES cell clusters in aggregate cultures. Autoradiographic exposure was for 7 days (A), 24 h (B) and 18 h (C), at -80°C with intensifying screens.

The blot was initially hybridized (section 2.2.6b) with the *Cob54Pcr* probe corresponding to 817 bp of *Fbx15* cDNA sequence (section 2.1.2; Figure 5.3), then re-hybridized with a probe *"Oct4"* corresponding to a 1 kb fragment of *Oct-4* coding sequence (section 2.1.2). Finally, the blot was re-hybridized with the *Gapdh* probe (section 2.1.2) as a control for RNA loading at each time point.

Results in Figure 6.12 (A-C) show that as for *Oct-4*, expression of the *Fbx15* gene in ES cells is strong and is downregulated with *in vitro* differentiation. In the absence of LIF, the downregulation for *Fbx15* expression in ES cells appears to be occurring at an earlier time point than for *Oct-4*. A slight upregulation is seen for both genes from the 6 to the 8 day cultures. Although there is a slight variation in RNA sample load, as indicated by *Gapdh* hybridization, this upregulation is likely to be due to the expansion of residual ES cell clusters over time, a common anomaly seen in ES cell differentiation cultures.

6.5 DISCUSSION

Although the COB54 gene trap interruption has not resulted in an overt phenotype in animals homozygous for the transgenic allele, the results in this chapter present an interesting developmental expression profile which is not inconsistent with that anticipated by *in vitro* studies for the endogenous *Fbx15* gene.

Histochemical analysis of the COB54 reporter gene in oocytes and early cleavage stage embryos suggests the endogenous gene is activated at the onset of embryonic transcription (late 2-cell to early 4-cell stage), (Figure 6.2). This result is confirmed by RT-PCR analysis which shows *Fbx15* transcripts are weakly detected in cultured 2-cell stage embryos, with a strong upregulation at the 4-cell stage (Figure 6.4).

The marked nuclear localized *lacZ* expression first seen at the late 2-cell stage continues through all stages of preimplantation development and is evident in both the TE and ICM lineages of the 3.5 dpc blastocyst. It is an interesting observation, that a gene for which expression is strongly restricted to undifferentiated ICM-derived ES cells *in vitro*, is expressed strongly in the TE cells after the first major embryonic differentiation event has taken place. While this result is not inconsistent with the screening strategy used to select the gene trap clones (section 3.4), expression in TE cells provides a clear example of how under-representation of specific cell types *in vitro*, in this case extraembryonic cells, must be considered as a limitation in predicting expression profiles *in vivo*.

Strong COB54 reporter gene expression continues in all cells of both lineages as the embryo hatches from the zona on the 5th day of development (Figure 6.2) and is dramatically downregulated upon implantation.

The marked downregulation for *Fbx15* expression upon implantation results in a complete loss of expression in the embryonic epiblast by 6.5 dpc. Weak βgal activity is evident only in the primitive endoderm-derived cuboidal VE cells surrounding the proximal extraembryonic ectoderm and some trophoblastic cells immediately adjacent to this VE layer, but not throughout the ectoplacental cone. By 7.5 dpc, there is a complete absence of expression as determined by *lacZ* analysis in any cells of the gastrulating embryo (Figure 6.3). The dramatic downregulation of COB54 reporter gene expression is consistent with RT-PCR analysis which shows *Fbx15* transcripts are most strongly detected in the early blastocyst stage. The most substantial change in *Fbx15* and COB54 reporter gene expression identified in these studies occurs at the 5.5 dpc stage of development where no evidence of expression could be detected by RT-PCR or *lacZ* staining. Expression reappears in 6.5 dpc and 7.5 dpc embryos, although at markedly lower levels than for the preimplantation stages (Figures 6.3 & 6.4). It is of interest that the sharp embryonic downregulation of the F-box protein entrapped in this study, coincides with the time point in development when cell cycle length decreases and the epiblast undergoes rapid growth (Snow, 1977; Gardner and Beddington, 1988).

The *in vivo* loss of expression for *Fbx15* following implantation and gastrulation is paralleled in blastocyst explant cultures where heterozygote COB54 blastocysts, collected at 3.5 dpc and cultured in the absence of LIF, show a strong downregulation of βgal activity after the first two days of culture (Figure 6.11). After 4 days of culture, only a few migratory parietal endodermal cells show some mild staining. Interestingly, after the 5th day of culture, moderate staining is again seen in a small subset of cells within the ICM-like group of cells, and after 6 days, this expression is quite strongly upregulated throughout the highly differentiated 3-dimensional explant cultures. This upregulation after prolonged culture may, however, be associated with the disorganized or preferential differentiation to particular cell types *in vitro*.

To investigate whether the *in vivo* downregulation of *Fbx15* expression might be a function of developmental age or the event of implantation, *lacZ* expression was evaluated in heterozygote COB54 embryos for which implantation was artificially delayed. Implantation into the uterus on the 5th day of development is critically

174

dependent on the progesterone and estrogen levels of the mother's uterine environment. Removal of the estrogen producing corpus lutea by ovariectomy renders the uterus non-receptive and puts the blastocyst into diapause, known as "delay" (Mantalenakis and Ketchel, 1966; Yoshinaga and Adams, 1966). The metabolism of blastocysts slows down during this phase of "sustained development", and ICM mitotic activity is reduced (McLaren, 1968). While trophoblastic giant cell transformation is inhibited in an ICM-independent manner (Snow *et al.*, 1976; Surani and Barton, 1977), differentiation of the primitive endoderm does, however, continue to occur at the same time that it would in normal blastocysts (Gardner *et al.*, 1988). Implantation-delayed COB54 embryos collected at 6.5 dpc and 8.5 dpc stain strongly for βgal activity in all cells at both stages (Figure 6.5), suggesting a downregulation in *Fbx15* gene expression is dependent upon processes associated with implantation.

The dramatic downregulation of *Fbx15* embryonic expression at implantation, as evident by COB54 reporter gene expression and RT-PCR analysis of the *Fbx15* wild type allele, precedes the time in development where pluripotential stem cells are lost and *Oct-4* becomes restricted to the germ cell lineage (Rosner *et al.*, 1990; Schöler *et al.*, 1990; Yeom *et al.*, 1996; Pesce *et al.*, 1998). Time course Northern analysis of *Fbx15* and *Oct-4* transcripts in differentiating ES cell cultures is consistent with these observations (Figure 6.12). While both transcripts are expressed strongly in ES cells, upon induction of differentiation *Fbx15* expression appears to be downregulated at an earlier stage than for *Oct-4*.

Whole-mount X-gal staining and histological examination for postimplantation stage embryos did not detect any embryonic expression of the gene trap COB54 allele between 6.5 dpc stage and 15.5 dpc. Some nuclear localized βgal activity is evident during this period in the extraembryonic layer of columnar endodermal cells facing into the embryonic visceral yolk sac. This strong expression is initially detected at 10.5 dpc in a few of these cells, and then throughout the yolk sac from 12.5 dpc onwards (Figure 6.3).

No COB54 reporter gene expression was detected in genital ridges of 11.5 dpc, 12.5 dpc and 13.5 dpc COB54 heterozygote embryos (Figure 6.8). However, non-specific *lacZ* expression was seen in the developing tubules of both male and female 12.5dpc and 13.5 dpc proximal mesonephros, which is also consistent with the non-specific

175

endogenous expression seen in both neonatal and adult kidney tissue. *Fbxl5* expression does not, therefore, appear to be associated with the germ cell lineage.

RT-PCR results show *Fbxl5* transcripts are detectable in 10 dpc and 12 dpc whole embryos, and in embryonic placenta at 10 dpc. Very weak expression is detected for 16-18 dpc late gestation brain and more clearly in 18 dpc lung (Figure 6.9). These results suggest a likely upregulation for *Fbxl5* expression in at least a proportion of cells in these two tissues, just prior to birth.

COB54-specific *lacZ* expression was detected by whole-mount X-gal staining in day 2 neonatal lung and brain tissues, but only in a few residual cells in adult lung, and not in adult brain (Figures 6.6 & 6.7). By histological examination, strong blue staining cells were seen throughout the lung parenchyma around the developing neonatal alveolar structures. Although similar staining is seen in wild type neonatal lung, *lacZ* expression appeared to be nuclear localized and more predominant in the transgenic lung tissue. In the brain, only occasional clusters of blue staining cells were seen, but nuclear staining was quite clear. RT-PCR analysis confirmed the presence of *Fbxl5* transcripts in the neonate lung and brain tissues (Figure 6.9).

Of the adult tissues analysed in this study, the testis is the only tissue that shows definitive expression of the *Fbxl5* gene by RT-PCR analysis (Figure 6.9). This result correlates with a PCR analysis described for the initial reporting of this gene (Winston *et al.*, 1999a). Marked βgal activity is evident by whole-mount X-gal staining in the tubular tissue of COB54 heterozygote testis and is quite distinct from the endogenous *lacZ* expression normally seen in the Leydig cell-containing interstitial tissue of wild type testis. At a histological level, the marked COB54 *lacZ* expression in the seminiferous epithelia is seen specifically in the spermatocyte and round spermatid populations, with variations in their staining intensities according to the stage of spermatogenesis in individual tubules (Russell *et al.*, 1990). Expression is not seen in the basal lamina surrounding the seminiferous tubules, nor in the spermatogonial germ cells that lie along this basal layer (Figure 6.6).

Testicular expression of the COB54 allele appears to commence some time during the mid-pachytene stage of the long first meiotic prophase in the primary spermatocytes that have arisen from Type-B spermatogonia. At this stage, homologous chromosomes have paired and genetic crossing-over takes place. Dispersed X-gal staining seen in the

pachytene spermatocytes is due to an increased size of the nucleus, and more widespread chromosomal material at this time. *lacZ* expression continues to be detected in the secondary spermatocytes and the smaller haploid round spermatids. No expression is detected in the elongate spermatids, however, nor in any of the subsequent differentiation phases that give rise to mature spermatozoa. It therefore seems that *Fbx15* expression is associated with the meiotic phase, and not the proliferative or differentiation phases, of spermatogenesis. It correlates that no *lacZ* expression is seen in the seminiferous epithelia of the day 2 neonate COB54 testis (Figure 6.7), as the type-A spermatogonial stem cells first appear 3-7 days after birth and do not enter meiosis until adulthood (reviewed by Russell *et al.*, 1990; Hogan *et al.*, 1994).

The association of COB54 reporter gene expression with male meiosis further supports a possible role for the *Fbx15* gene in cell cycle regulation. Developmentally regulate gene expression during spermatogenesis has been reported for certain members of the cyclin family. Cyclin B1 is elevated in pachytene spermatocytes and early round spermatids (Chapman and Wolgemuth, 1992, 1993), while the Cyclin B1-dependent kinase Cdc2 has been found in pachytene spermatocytes, but not in spermatids (Chapman and Wolgemuth, 1994). Expression of another B-type cyclin, Cyclin B2, is highest in late pachytene and diplotene spermatocytes (Chapman and Wolgemuth, 1993). Cyclin A1, which is expressed in mice exclusively in the germ cell lineage (Sweeney *et al.*, 1996), and in humans at highest levels in the testis and in certain myeloid leukemia cells (Yang *et al.*, 1997; Kramer *et al.*, 1998), has been demonstrated as an essential requirement for spermatocyte passage into the first meiotic division (Liu *et al.*, 1998). A null mutation of the gene encoding for Cyclin A1, *Ccna1*, results in sterility in male, but not female, mice and is associated with a reduction of Cdc2 kinase activation at the end of meiotic prophase in spermatocytes (Liu *et al.*, 1998). Other proteins implicated in regulating meiosis in the testis include those encoded by some tumor-suppressor genes such as *p53*, *Rb1* and *Rbcal*, and the heat-shock protein HSP70-2 which has been shown to have an essential role in determining Cdc2 kinase activity in pachytene spermatocytes (Dix *et al.*, 1996), (reviewed by Eddy and O'Brien, 1998).

Northern analysis of heterozygote COB54 testis RNA confirms *in vivo* expression of the approximately 6.6 kb COB54 gene trap allele and two endogenous *Fbx15* transcripts of approximately 1.9 kb and 2.5 kb, as seen for RNA from the gene trap COB54 ES cell

line (Figures 6.1 & 6.10). The gene trap allele appears to be less abundantly expressed than the endogenous transcripts, though this difference in apparent expression levels may be attributable to differences in mRNA processing or stability. In the homozygote COB54 testis the wild type transcripts are not detected, although with an increased gel loading of RNA there is faint detection of a hybridization band which could represent the approximately 2.5 kb endogenous transcript. It is therefore possible that weak expression of this transcript, perhaps as a result of alternate splicing around the gene trap insertion, may contribute to an apparent lack of an overt phenotype in homozygote animals. An extra hybridization band of weak intensity is also detected in each transgenic testis sample just above the two endogenous transcripts. This band may represent an alternate testis-specific transcript or possibly a non-specific product, though further investigation is required to determine its nature.

In summary, the endogenous *Fbxl5* gene pertaining to the COB54 gene trap insertion displays an expression profile which is restricted to ES cells *in vitro*, and *in vivo* is activated with the onset of the embryonic genome, and expressed strongly in both the ICM and TE lineages of the preimplantation embryo. *Fbxl5* gene expression is markedly downregulated upon implantation and is most strikingly upregulated in the adult testis, where it appears to be associated with the meiotic phase of spermatogenesis. While no morphological or behavioural phenotype has been observed in COB54 homozygote animals, the possible presence of low levels of wild type mRNA and the insertion of the gene trap at the 3' end of the encoded F-box protein, are factors which may have curtailed any null mutant effect for the COB54 gene trap interruption.

Chapter Seven

CONCLUSION & FUTURE DIRECTIONS

Pluripotential stem cells derived from the preimplantation mammalian embryo have the capacity to give rise to differentiated progeny representative of the three embryonic germ layers, the germ cells and some extraembryonic tissues. Mouse ES cells are currently providing a powerful tool for the identification, genetic modification and functional analysis of genes involved in normal mammalian development, and an enormous potential is now evident for the application of human ES cells in cell-based transplantation therapies. To date, however, the large-scale propagation, efficient cloning, or genetic manipulation of human ES cells and their directed differentiation into specific cell lineages, has not yet been achieved. Realization of the envisaged applications for both ES and multipotent stem cells will require elucidation of the genetic mechanisms underpinning the regulation of self-renewal and differentiation in these cells, as well as the identification of extrinsic factors that influence their growth and differentiation *in vitro*.

A number of regulatory pathways and molecules have been identified which regulate the growth and differentiation of ES cells. The cytokine LIF promotes self-renewal of mouse ES cells *in vitro* through the activation of the gp130 receptor (Smith and Hooper, 1987; Smith *et al.*, 1988; Williams *et al.*, 1988; Gearing *et al.*, 1991), while human ES cells appear not to be responsive to exogenous LIF and instead require a feeder layer of

mouse embryonic fibroblast cells to support their survival and self-renewal (Thomson *et al.*, 1998; Reubinoff *et al.*, 2000). Of interest, however, is a recent report describing the *in vitro* growth and maintenance of undifferentiated human ES cells by culturing on a laminin-based matrix and in media conditioned by mouse embryonic fibroblasts (Xu *et al.*, 2001). *Oct-4*, expressed in both mouse and human ES cells, is crucial for maintaining the pluripotential cells in the mouse preimplantation epiblast (Nichols *et al.*, 1998) and is now known to play a key regulatory role in maintaining the pluripotent ES cell phenotype via precise regulation of its expression levels (Niwa *et al.*, 2000). This quantitative regulation may be mediated by a fine balance between the STAT3 and ERK effectors of the gp130 signaling pathway or by alternative self-renewal signals (Burdon *et al.*, 1999a; Niwa *et al.*, 2000). There are now several reports for candidate positive regulators of ES cell growth, unrelated to LIF, which may play a role in human ES cell maintenance, although none have been purified to homogeneity or cloned (Roach S. *et al.*, 1993; Dani *et al.*, 1998; Rathjen *et al.*, 1999). In an effort to identify genes that might play a regulatory role in controlling the growth and differentiation of ES cells, this study has utilized a gene trap cloning and *in vitro* pre-screening approach to identify genes which are expressed specifically in mouse ES cells, and not in their differentiated progeny.

There are now many reports demonstrating the successful use of gene trap vector constructs for simultaneously identifying and mutating novel developmentally important genes. This is well exemplified by the recent identification of the novel *Tbn* gene, crucial for survival of the ICM in preimplantation embryos (Voss *et al.*, 2000). The results in the current study demonstrate the use of two new optimized gene trap vectors, pGT1.8 (OPT) IRES*βgeo* and pGT1.8 (OPT/Nuclear) IRES*βgeo*, for the efficient generation of ES cell gene trap clones yielding detectable reporter expression, and in the case of the latter construct, nuclear localized *lacZ* expression. Taking advantage of the ability for ES cells to spontaneously differentiate *in vitro*, the three-tiered pre-screen established in this study has efficiently identified restricted *in vitro* expression profiles for gene trap events of interest, overcoming the need to generate large numbers of chimaeric offspring. This stringent *in vitro* screening system identified less than 1% (28/4345) of the generated gene trap events as insertions in transcription units that are active specifically in undifferentiated ES cells, and demonstrates for the first time the

powerful approach of screening gene trap clones for ES cell-specific neo resistance in embryoid body cultures.

Rapid analysis of 28 ES cell-restricted gene trap events by direct sequencing of 5'-RACE products generated from fusion transcript mRNAs, demonstrates correct use of the vector splice site and the identification of a single 5' sequence read for 36% of the clones (10/28). This result compares well with a previous report for a large-scale analysis of ES cell gene trap clones generated with the pGT1.8TM secretory gene trap vector (Townley *et al.*, 1997), and indicates that the inclusion of an IRES unit and a nuclear localization signal in the pGT1.8 (OPT/Nuclear) IRES*βgeo* construct has not been detrimental to the recovery of useful information regarding the trapped endogenous sequences.

Direct sequence analyses identified three independent gene trap integrations in the same gene, although in different locations within a likely common intron for this gene. In each case the entrapped sequence is in the 5'-flanking region of a gene that is related to the RBP-J$_k$ transcription factor, which itself interacts with the intracellular domain of the Notch receptor to form a signalling pathway between the cell surface and nucleus (reviewed by Honjo, 1996; Lendahl, 1998). The three insertions within the *RBP-J$_k$*-related gene in this study and in one further independent report (J. Brennan, Centre for Genome Research, Edinburgh, U.K; *pers. comm.*), are suggestive of a genomic hotspot for gene trap integration in ES cells. It is of interest that the trapped sequence lies immediately downstream of sequence with high homology to mouse L1 repetitive elements. L1 elements are replicating retrotransposons which are inserted throughout the mammalian genome and are likely to have a profound effect on the evolution, structure and function of the genome (reviewed by Furano, 2000). L1 elements within close proximity of the insertion site may be further evidence of recombination in this region and point to a possible explanation for the high incidence of gene trap integration at this site in ES cells. It is possible that with the use of vectors with different integration preferences, such as exon traps and retrovirally inserted traps, insertion into this sequence would not be favoured, but at present the significance and mechanism of integration in this apparent trapping hotspot in the ES cell genome is not clear.

For five novel ES cell-specific sequences identified, four of the corresponding ES cell lines were unable to contribute to germline transmitting chimaeras, including one line that did not yield any offspring from 69 blastocyst injections. Chimaeras for the

181

remaining three lines showed a high ES cell contribution by coat colour, however, and breeding success was poor. Having trapped genes which are preferentially expressed in the pluripotential cells *in vitro*, these results may point to a possible gene dosage effect in cells heterozygous for the entrapped allele which could result in an associated effect on the pluripotential cells of the preimplantation embryo. To date these four gene trap ES cell lines, COB28, COB36, COB37 and COB52, stand as insertions in uncharacterized ES cell-restricted sequences.

The COB54 gene trap cell line displayed a greater permissibility to contribute to chimaeras, and to the germ cells in offspring. The endogenous sequence, initially of interest for its homology with an EST sequence from a mouse 2-cell cDNA library (Rothstein *et al.*, 1992) and a tightly restricted *in vitro* expression profile, has been recently reported as the gene *Fbx15*, belonging to a novel family of mammalian proteins containing a conserved F-box motif (Cenciarelli *et al.*, 1999; Winston *et al.*, 1999a). F-box proteins are an expanding family of adaptor subunit proteins that are responsible for substrate-recognition by E3 ubiquitin ligase complexes, which target proteins for ubiquitin-mediated degradation (reviewed by Patton *et al.*, 1998a; Craig and Tyers, 1999; Tyers and Jorgenson, 2000). While no function has yet been assigned for the *Fbx15* gene, the COB54 gene trap insertion has allowed detailed investigation of gene expression at the cellular level and provided clear evidence of a spatially and temporally restricted expression profile, both *in vitro* and *in vivo*.

The strong expression for the COB54 gene trap allele in the preimplantation embryo, commencing with activation of the early embryonic genome, is markedly downregulated following implantation and is only later detected in the developing visceral yolk sac of the extraembryonic tissues, late gestation and neonatal lung and brain tissues, and most strikingly, in the adult testis. In collaboration with Dr. Kate Loveland (MIRD), *in situ* tissue hybridization analyses are now being undertaken to confirm testicular *Fbx15* gene expression, and ICM and TE cell expression in the blastocyst stage embryo.

While results of the current study have not identified a functional role for the *Fbx15* gene product, it is interesting to draw from the information available to speculate on possible activities within the cell. One such role, which correlates well with the tightly regulated *Fbx15* gene expression evident at the meiotic phase of spermatogenesis and the complete downregulation of expression in the rapidly dividing cells of the

182

postimplantation embryonic epiblast, might be for this gene product to target cell cycle regulatory protein(s) for ubiquitin-mediated proteolysis. The future identification of such substrate protein(s) remains of much interest, and may eventually shed light on a role for the novel F-box protein trapped in this study and a unique pathway for ES cell regulation.

While the COB54 gene trap insertion did not result in an overt homozygote phenotype, an important developmental role for the *Fbx15* gene is not excluded. The possible *in vivo* presence of very low levels of wild type transcript in animals homozygous for the gene trap insertion, a previously observed phenomenon attributable to alternate splicing around the gene trap integration or failure to make use of the construct's poly A signal (reviewed by Voss *et al.*, 1998a), may have curtailed a phenotypic effect. Furthermore, the gene trap insertion occurs at the carboxy terminal end of the predicted amino acid sequence and therefore, would not necessarily result in the generation of a dysfunctional endogenous gene product.

Definitive demonstration of a redundant homozygous phenotype requires the use of an *Fbx15* gene replacement strategy to ensure a complete deletion of gene function. Forced over-expression both *in vitro* and *in vivo*, and/or *in vitro* strategies to investigate the effect of upregulation and downregulation of *Fbx15* expression in ES cells, may provide further insight into a possible regulatory role for this gene.

Future improvements in gene trap vectors and screening strategies will no doubt enhance the efficiency of large-scale mutagenic screens for genes of interest in ES cell systems, particularly in specific cell lineages of interest. Replacement of the gene trap βgal reporter with a vital reporter gene, such as GFP, would facilitate real time monitoring of reporter gene expression and direct visualization of the reporter in pre- and post-implantation embryos, and tissues.

The COB54 transgenic mouse line developed in this study provides a valuable resource for future investigation of *Fbx15* gene expression *in vivo* and a particularly useful marker for the analysis of early embryonic activation in nuclear reprogramming studies. In collaboration with Ms. Megan Munsie (MIRD), *lacZ* negative adult cumulus cells from heterozygote COB54 female mice have been used as a somatic cell source for monitoring reactivation of this early embryonic gene, following nuclear transfer into enucleated mouse oocytes. The COB54 gene trap allele shows marked nuclear

localized *lacZ* expression in the 2-cell nuclear transfer embryo where reprogramming has been successful (*manuscript in preparation*). Similarly, in collaboration with Dr. Andrew French (MIRD), fibroblast cell lines derived from adult COB54 heterozygote males, are presently being used as an adult somatic cell source for nuclear transfer into enucleated bovine oocytes.

In summary, this study has identified and investigated the developmental and *in vitro* expression profile of a novel member of the recently identified family of mammalian F-box proteins (Cenciarelli *et al.*, 1999; Winston *et al.*, 1999a). There is presently much interest in the role of these proteins in targeting cell regulatory proteins for ubiquitin-mediated proteolysis and there is now significant evidence to support a role for F-box proteins in regulating the mammalian cell cycle. The transition to alternate cell cycles seen upon ES cell differentiation *in vitro*, in embryonic epiblast cells following implantation and, in cells undergoing spermatogenic meiosis, is consistent with a possible cell cycle regulatory role for the *Fbx15* gene product. The substrate(s) targeted by the Fbx15 protein for degradation both in ES cell cultures and *in vivo*, once identified, will shed further light on the function of the Fbx15 protein. A better understanding of the processes associated with this pathway of regulation may eventually provide new opportunites to enhance ES cell culture systems for human and other mammalian species, and in turn help realize some of the exciting possibilities now evident in the burgeoning field of stem cell biology.

Appendix

Recipes for commonly use reagents

50x TAE Buffer

Add   484 g   Tris-base
           114.2 ml   glacial acetic acid
           200 ml   0.5 M EDTA, pH 8.0

Bring to a final volume of 2 litres with ddH_2O.

For a <u>1x TAE working solution</u>, dilute 200 ml of 50x TAE to 10 litres with ddH_2O and store at RT°C.

10x TBE Buffer

Add   540 g   Tris-base
           275 g   Boric acid
           200 ml   0.5 M EDTA, pH 8.0 *or* 46.5 g $Na_2EDTA.2H_2O$, pH 8.0

Bring to a final volume of 5 litres, adjusting the pH to 8.3.

For a <u>1x TBE working solution</u>, dilute 100 ml of 10x TBE to 1 litre with ddH_2O and store at RT°C.

10x MOPS Buffer

Add   41.8 g   MOPS [3-(N-morpholino) prop-panesulfonic acid]   (0.2 M)
           4.1 g   Sodium acetate   (50 mM)
           3.72 g   EDTA salt   (10 mM)

Bring to a final volume of 1 litre in ddH_2O, adjusting the pH to 7.0 with 10 M NaOH (~10 ml). Autoclave (the solution turns a yellow colour), and store at RT°C.

For a <u>1x MOPS working solution</u>, dilute 100 ml of 10x MOPS to 1 litre with DEPC-treated ddH_2O, making fresh prior to use.

20x SSPE buffer

Add   696 g   NaCl
           110.4 g $NaH_2PO_4.H_2O$
           30 g   EDTA salt

Bring to a final volume of 4 litres in ddH_2O, adjusting the pH to 7.4 with 10 M NaOH. Leave on a magnetic stirrer overnight. Autoclave and store at RT°C.

20x SSC buffer

Add   175.5 g   NaCl
           88.2 g $Na_3Citrate$

Bring to a final volume of 1 litre, adjusting the pH to 7.0 with 10 M NaOH. Autoclave and store at RT°C.

1x TE Buffer

Add   10 ml 1 M Tris-HCl, pH 8.0
         2 ml 0.5 M EDTA, pH 8.0
Bring to a final volume of 1 litre with ddH$_2$0, autoclave and store at RT°C.

PBS Buffer

Add   40 g   NaCl   (137 mM)
       7.25 g Na$_2$HPO$_4$.7H$_2$0   (4.3 mM)
       1 g   KCl   (2.7 mM)
       1 g   KH$_2$PO$_4$ (1.4 mM)
Bring to a final volume of 5 litres with ddH$_2$0, adjusting the pH to 7.4. Autoclave and store at RT°C.

Alternatively, dissolve five PBS (*Dulbecco A*) tablets (Oxoid, Basingstoke, UK) in 500 ml ddH$_2$0 to give a PBS solution at pH 7.3, and autoclave.

DEPC-treated ddH$_2$0/solutions

Prepare in a fume hood by adding 1 ml of diethyl-pyrocarbonate (stored at 4°C, Sigma) to 1 litre of ddH$_2$0, shake thoroughly at several intervals and leave overnight at RT°C. Autoclave and keep at RT°C.
Similarly,other solutions required to be kept RNase-free are treated with DEPC (1:1000 v/v) and autoclaved as above.

10% SDS

Dissolve 100 g of SDS in 900 ml ddH$_2$0, heating in a 65°C water bath to assist dissolution. Adjust the pH to 7.2 by adding a few drops of concentrated HCl. Make up to 1 litre with ddH$_2$0 and store at RT°C.

5 M NaCl

Dissolve 292.2 g of NaCl in 800 ml ddH$_2$0. Make up to 1 litre with ddH$_2$0, autoclave and store at RT°C.

10 M NaOH

Dissolve 400 g of NaOH pellets in 800 ml of ddH$_2$0, taking care with generated heat. Make up to 1 litre in ddH$_2$0 and store at RT°C.

1 M Tris-HCl, pH 7.4 – 8.0

Dissolve 121.1 g of Tris base in 800 ml of ddH$_2$0. Adjust the pH at RT°C to the desired value by adding concentrated HCl:
                pH 7.4   70 ml
                pH 7.6   60 ml
                pH 8.0   42 ml
Make up to 1 litre with ddH$_2$0, autoclave and store at RT°C.

0.5 M EDTA, pH 8.0

Add 186.1 g of disodium ethylenediaminetetra-acetate.$2H_2O$ (EDTA salt) to 800 ml ddH_2O and stir vigorously on a magnetic stirrer. Adjust the pH to 8.0 with NaOH (~20 g of NaOH pellets), bringing the EDTA salt into solution. Autoclave and store at RT°C.

3 M Sodium Acetate, pH 5.2

Dissolve 408.1 g of NaOAc.$3H_2O$ in 800 ml ddH_2O. Adjust the pH to 5.2 with glacial acetic acid and make up to 1 litre in ddH_2O. Autoclave and store at RT°C.

RNase A 10mg/ml

To 100 mg DNase-free pancreatic RNase A (Sigma) add:
        100 µl 1 M sterile Tris-HCl, pH 7.5
        30 µl 5 M sterile NaCl
Add sterile ddH_2O until RNA is dissolved (~1 ml), and make up to 10 ml with sterile ddH_2O in a 50ml tube. Loosen the cap and boil in a water bath for ~20 min. Allow to cool to RT°C and store in 50 µl aliquots at -20°C.
For a 10 µg/ml working stock, dilute a 50 µl 10 mg/ml aliquot to 50 ml in TE buffer and store 1 ml aliquots at -20°C.

DNA electrophoresis loading dye

For a 10x DNA gel-loading buffer, the following mix is made up in ddH_2O and stored at RT°C:
0.25% (w/v)   Orange G (Sigma)
15% (v/v)   Ficoll (Type 400, Pharmacia)

RNA electrophoresis loading dye

For a 6x RNA gel-loading buffer freshly prepare the following mix prior to loading samples, or store in small aliquots at -20°C.
1.5 ml	deionized formamide
300 µl	10x MOPS buffer
480 µl	37% formaldehyde solution
320 µl	DEPC-treated sterile ddH_2O
200 µl	glycerol
160 µl	2.5% (w/v) bromphenol blue (Sigma)

Genomic DNA lysis buffer

Bring the following mix to a final volume of 100 ml in ddH$_2$0 and store at 4°C.

10 ml	1M Tris-HCl, pH 8.5	(100 mM)
1 ml	0.5 M EDTA, pH 8.0	(5 mM)
2 ml	10% SDS	(0.2%)
4 ml	5 M NaCl	(200 mM)

Prior to use Proteinase K is added at a final concentration of 100 µg/ml.

Low salts genomic DNA lysis buffer

Bring the following mix to a final volume of 100 ml in ddH$_2$0 and store at 4°C.

1 ml	1 M Tris-HCl, pH 8.3	(10 mM)
200 µl	Tween 20	(0.2% v/v)
1 ml	5 M NaCl	(50 mM)

Prior to use Proteinase K is added at a final concentration of 100 µg/ml.

Poly A$^+$ mRNA STE lysis buffer

2 ml	5 M NaCl	(0.1 M)
2 ml	1 M Tris-HCl (pH 7.5)	(20 mM)
2 ml	0.5 M EDTA	(10 mM)

Make up to final volume of 100 ml in ddH$_2$0. DEPC-treat (1:1000 v/v) overnight, autoclave and store at 4°C.

Oligo(dT) loading buffer

16 ml	5 M NaCl	(0.4 M)
4 ml	1 M Tris-HCl (pH 7.5)	(20 mM)
4 ml	0.5 M EDTA	(10 mM)
4 ml	10% SDS	(0.2%)

Make up to final volume of 200 ml in ddH$_2$0. DEPC-treat (1:1000 v/v) overnight, autoclave and store at RT°C. Add the SDS after autoclaving.

Oligo(dT) washing buffer

1.0 ml	5 M NaCl	(0.1 M)
0.5 ml	1 M Tris-HCl (pH 7.5)	(10 mM)
0.1 ml	0.5 M EDTA	(1 mM)
1.0 ml	10% SDS	(0.2%)

Make up to final volume of 50 ml in ddH$_2$0. DEPC-treat (1:1000 v/v) overnight, autoclave and store at RT°C. Add the SDS after autoclaving.

Oligo(dT) elution buffer

50 µl	1 M Tris-HCl (pH 7.5)	(1 mM)
100 µl	0.5 M EDTA	(1 mM)
1.0 ml	10% SDS	(0.2%)

Make up to final volume of 50 ml in ddH$_2$0. DEPC-treat (1:1000 v/v) overnight, autoclave and store at RT°C. Add the SDS after autoclaving.

DNA hybridization buffer

50 ml   1 M Na_2HPO_4 (pH 7.2)   0.5 M
35 ml   20% SDS   7%
1 g   BSA   1%
0.2 ml   0.5 M EDTA (pH 8.0)   1 mM
Make up to a final volume of 100 ml in ddH_2O and dissolve on a heated stirrer. Store at RT°C and heat to 65°C prior to use.

RNA hybridization buffer

5 ml   deionized formamide   50% (v/v)
1 ml   50x Denhardt's   5x (stored at -20°C)
1 ml   10% SDS   1%
1 ml   30% dextran sulphate   3% (stored at 4°C)
2.5 ml   20x SSPE   5x
Make up a fresh 10.5 ml mix just prior to use, heat to 42°C and add denatured salmon sperm DNA (10 mg/ml, Boehringer) at 100 µg/ml.

DNA wash buffer I

100 ml   20x SSC   2x
10 ml   10% SDS   0.1% (w/v)
Make up to a final volume of 1 litre in ddH_2O and store at RT°C. Heat to 65°C prior to use.

DNA wash buffer II

5 ml   20x SSC   0.1x
10 ml   10% SDS   0.1% (w/v)
Make up to a final volume of 1 litre in ddH_2O and store at RT°C. Heat to 65°C prior to use.

RNA wash buffer I

25 ml   20x SSPE   1x
25 ml   10% SDS   0.5% (w/v)
Make up to a final volume of 1 litre in ddH_2O and store at RT°C. Heat to 65°C prior to use.

RNA wash buffer II

2.5 ml   20x SSPE   0.1x
25 ml   10% SDS   0.5% (w/v)
Make up to a final volume of 1 litre in ddH_2O and store at RT°C. Heat to 60°C prior to use.

Appendix

Bibliography

Acampora, D., Mazan, S., Lallemand, Y., Avantaggiato, V., Maury, M., Simeone, A., and Brulet, P. (1995). Forebrain and midbrain regions are deleted in Otx2-/- mutants due to a defective anterior neuroectoderm specification during gastrulation. *Development* **121**, 3279-3290.

Akam, M. (1987). The molecular basis for metameric pattern in the Drosophila embryo. [Review] [165 refs]. *Development* **101**, 1-22.

Albano, R.M., Arkell, R., Beddington, R.S., and Smith, J.C. (1994). Expression of inhibin subunits and follistatin during postimplantation mouse development: decidual expression of activin and expression of follistatin in primitive streak, somites and hindbrain. *Development* **120**, 803-813.

Allen, N.D., Cran, D.G., Barton, S.C., Hettle, S., Reik, W., and Surani, M.A. (1988). Transgenes as probes for active chromosomal domains in mouse development. *Nature* **333**, 852-855.

Allikmets, R., Gerrard, B., Hutchinson, A., and Dean, M. (1996). Characterization of the human ABC superfamily: isolation and mapping of 21 new genes using the expressed sequence tags database. *Hum. Mol. Genet.* **5**, 1649-1655.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. [Review] [90 refs]. *Nucl. Acids. Res.* **25**, 3389-3402.

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.

Ambrosetti, D.C., Basilico, C., and Dailey, L. (1997). Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. *Mol. Cell. Biol.* **17**, 6321-6329.

Anderson, R., Fassler, R., Georges-Labouesse, E., Hynes, R.O., Bader, B.L., Kreidberg, J.A., Schaible, K., Heasman, J., and Wylie, C. (1999). Mouse primordial germ cells lacking beta1 integrins enter the germline but fail to migrate normally to the gonads. *Development* **126**, 1655-1664.

Andrews, P.W. (1988). Human teratocarcinomas. [Review] [163 refs]. *Biochem Biophys Acta* **948**, 17-36.

Andrews, P.W., Damjanov, I., Simon, D., Banting, G.S., Carlin, C., Dracopoli, N.C., and Fogh, J. (1984). Pluripotent embryonal carcinoma clones derived from the human teratocarcinoma cell line Tera-2. Differentiation *in vivo* and *in vitro*. *Lab. Invest.* **50**, 147-162.

Ang, S.L., Jin, O., Rhinn, M., Daigle, N., Stevenson, L., Rossant, and J. (1996). A targeted mouse Otx2 mutation leads to severe defects in gastrulation and formation of axial mesoderm and to deletion of rostral brain. *Development* **122**, 243-252.

Ang, S.L. and Rossant, J. (1994). HNF-3 beta is essential for node and notochord formation in mouse development. *Cell* **78**, 561-574.

Ang, S.L., Conlon, R.A., Jin, O., and Rossant, J. (1994). Positive and negative signals from mesoderm regulate the expression of mouse Otx2 in ectoderm explants. *Development* **120**, 2979-2989.

Araki, K., Imaizumi, T., Sekimoto, T., Yoshinobu, K., Yoshimuta, J., Akizuki, M., Miura, K., Araki, M., and Yamamura, K. (1999). Exchangeable gene trap using the Cre/mutated lox system. Cell. Mol. Biol. 45, 737-750.

Aso, T., Haque, D., Barstead, R.J., Conaway, R.C., and Conaway, J.W. (1996). The inducible elongin A elongation activation domain: structure, function and interaction with the elongin BC complex. EMBO. J. 15, 5557-5566.

Aviv, H. and Leder, P. (1972). Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA. 69, 1408-1412.

Bagutti, C., Wobus, A.M., Fassler, R., and Watt, F.M. (1996). Differentiation of embryonal stem cells into keratinocytes: comparison of wild-type and beta 1 integrin-deficient cells. Dev. Biol. 179, 184-196.

Bai, C., Sen, P., Hofmann, K., Ma, L., Goebl, M., Harper, J.W., and Elledge, S.J. (1996). SKP1 connects cell cycle regulators to the ubiquitin proteolysis machinery through a novel motif, the F-box. Cell 86, 263-274.

Bailey, J.A., Carrel, L., Chakravarti, A., and Eichler, E.E. (2000) Molecular evidence for a relationship between LINE-1 elements and X chromosome inactivation: the Lyon repeat hypothesis. Proc. Natl. Acad. Sci. USA. 97, 6634-6639.

Bain, G., Ray, W.J., Yao, M., and Gottlieb, D.I. (1996). Retinoic acid promotes neural and represses mesodermal gene expression in mouse embryonic stem cells in culture. Biochem & Biophys Res Comm. 223, 691-694.

Bain, G., Kitchens, D., Yao, M., Huettner, J.E., and Gottlieb, D.I. (1995). Embryonic stem cells express neuronal properties in vitro. Dev. Biol. 168, 342-357.

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

Balling, R., Deutsch, U., and Gruss, P. (1988). undulated, a mutation affecting the development of the mouse skeleton, has a point mutation in the paired box of Pax 1. Cell 55, 531-535.

Barnes, F.L., Crombie, A., Gardner, D.K., Kausche, A., Lacham-Kaplan, O., Suikkari, A.M., Tiglias, J., Wood, C., and Trounson, A.O. (1995). Blastocyst development and birth after in-vitro maturation of human primary oocytes, intracytoplasmic sperm injection and assisted hatching. Hum. Reprod. 10, 3243-3247.

Barnes, J.D., Crosby, J.L., Jones, C.M., Wright, C.V., and Hogan, B.L. (1994). Embryonic expression of Lim-1, the mouse homolog of Xenopus Xlim-1, suggests a role in lateral mesoderm differentiation and neurogenesis. Dev. Biol. 161, 168-178.

Baumeister, W., Walz, J., Zuhl, F., and Seemuller, E. (1998). The proteasome: paradigm of a self-compartmentalizing protease. [Review] [113 refs]. Cell 92, 367-380.

Beck, F., Erler, T., Russell, A., and James, R. (1995). Expression of Cdx-2 in the mouse embryo and placenta: possible role in patterning of the extra-embryonic membranes. Dev. Dynamics 204, 219-227.

Beddington, R.S. and Robertson, E.J. (1999). Axis development and early asymmetry in mammals. [Review] [112 refs]. Cell 96, 195-209.

Beddington, R. (1998). Cripto-analysis of embryonic codes [news; comment]. Nature 395, 641.

Beddington, R.S. and Robertson, E.J. (1998). Anterior patterning in mouse. [Review] [69 refs]. *Trends. Genetics* 14, 277-284.

Beddington, R.S., Rashbass, P., and Wilson, V. (1992). Brachyury-a gene affecting mouse gastrulation and early organogenesis. *Development -Supp.* 157-165.

Beddington, R.S. and Robertson, E.J. (1989). An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development* 105, 733-737.

Beddington, R.S., Morgernstern, J., Land, H., and Hogan, A. (1989). An in situ transgenic enzyme marker for the midgestation mouse embryo and the visualization of inner cell mass clones during early organogenesis. *Development* 106, 37-46.

Beddington, R.S. (1987). Isolation, culture and manipulation of post-implantation mouse embryos. In Mammalian development. M. Monk, editor. IRL press, Oxford, UK. 43-69.

Beddington, R.S. (1983). Histogenetic and neoplastic potential of different regions of the mouse embryonic egg cylinder. *J. Embryol. Exp. Morph.* 75, 189-204.

Bedell, M.A., Jenkins, N.A., and Copeland, N.G. (1997). Mouse models of human disease. Part I: techniques and resources for genetic analysis in mice. [Review] [65 refs]. *Genes. Dev.* 11, 1-10.

Belaoussoff, M., Farrington, S.M., and Baron, M.H. (1998). Hematopoietic induction and respecification of A-P identity by visceral endoderm signaling in the mouse embryo. *Development* 125, 5009-5018.

Bellen, H.J., Wilson, C., and Gehring, W.J. (1990). Dissecting the complexity of the nervous system by enhancer detection. [Review] [34 refs]. *Bioessays* 12, 199-204.

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.

Beio, J.A., Bouwmeester, T., Leyns, L., Kertesz, N., Gallo, M., Follettie, M., and De Robertis, E.M. (1997). Cerberus-like is a secreted factor with neutralizing activity expressed in the anterior primitive endoderm of the mouse gastrula. *Mech. Dev.* 68, 45-57.

Ben-Shushan, E., Thompson, J.R., Gudas, L.J., and Bergman, Y. (1998). Rex-1, a gene encoding a transcription factor expressed in the early embryo, is regulated via Oct-3/4 and Oct-6 binding to an octamer site and a novel protein, Rox-1, binding to an adjacent site. *Mol. Cell. Biol.* 18, 1866-1878.

Berstine, E.G., Hooper, M.L., Grandchamp, S., and Ephrussi, B. (1973). Alkaline phosphatase activity in mouse teratoma. *Proc. Natl. Acad. Sci. USA.* 70, 3899-3903.

Bhatt, H., Brunet, L.J., and Stewart, C.L. (1991). Uterine expression of leukemia inhibitory factor coincides with the onset of blastocyst implantation. *Proc. Natl. Acad. Sci. USA.* 88, 11408-11412.

Biben, C., Stanley, E., Fabri, L., Kotecha, S., Rhinn, M., Drinkwater, C., Lah, M., Wang, C.C., Nash, A., Hilton, D., Ang, S.L., Mohun, T., and Harvey, R.P. (1998). Murine cerberus homologue mCer-1: a candidate anterior patterning molecule. *Dev. Biol.* 194, 135-151.

Bier, E., Vaessin, H., Shepherd, S., Lee, K., McCall, K., Barbel, S., Ackerman, L., Carretto, R., Uemura, T., Grell, E., and et al. (1989). Searching for pattern and mutation in the Drosophila genome with a P-lacZ vector. *Genes. Dev.* 3, 1273-1287.

Bjornson, C.R., Rietze, R.L., Reynolds, B.A., Magli, M.C., and Vescovi, A.L. (1999). Turning brain into blood: a hematopoietic fate adopted by adult neural stem cells *in vivo*. *Science* **283**, 534-537.

Blum, M., Gaunt, S.J., Cho, K.W., Steinbeisser, H., Blumberg, B., Bittner, D., De, and Robertis, E.M. (1992). Gastrulation in the mouse: the role of the homeobox gene goosecoid. *Cell* **69**, 1097-1106.

Bonaldo, M.F., Lennon, G., and Soares, M.B. (1996). Normalization and subtraction: two approaches to facilitate gene discovery. *Genome. Res.* **6**, 791-806.

Bonaldo, P., Chowdhury, K., Stoykova, A., Torres, M., and Gruss, P. (1998). Efficient gene trap screening for novel developmental genes using IRES beta geo vector and *in vitro* preselection. *Exp. Cell Res.* **244**, 125-136.

Bowman, P. and McLaren, A. (1970). Cleavage rate of mouse embryos *in vivo* and *in vitro*. *J. Embryol. Exp. Morph.* **24**, 203-207.

Bradley, A. (1987). Production and analysis of chimaeric mice. In: Teratocarcinoma and embryonic stem cells. E.J. Robertson, editor. IRL Press, Oxford, UK. 113-152.

Bradley, A. and Robertson, E. (1986). Embryo-derived stem cells: a tool for elucidating the developmental genetics of the mouse. [Review] [50 refs]. *Curr. Topics. Dev. Biol.* **20**, 357-371.

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

Brandon, E.P., Idzerda, R.L., and McKnight, G.S. (1995). Targeting the mouse genome: a compendium of knockouts (Part II) [published erratum appears in *Curr Biol* 1995 1;5(9):1073]. [Review] [189 refs]. *Curr. Biol.* **5**, 758-765.

Brehm, A., Ovitt, C.E., and Scholer, H.R. (1998). Oct-4: more than just a POUerful marker of the mammalian germline?. [Review] [94 refs]. *APMIS* **106**, 114-24; discussion 124-6.

Brenner, D.G., Lin-Chao, S., and Cohen, S.N. (1989). Analysis of mammalian cell genetic regulation in situ by using retrovirus-derived "portable exons" carrying the Escherichia coli lacZ gene. *Proc. Natl. Acad. Sci. USA.* **86**, 5517-5521.

Brilliant, M.H., Gondo, Y., and Eicher, E.M. (1991). Direct molecular identification of the mouse pink-eyed unstable mutation by genome scanning. *Science* **252**, 566-569.

Brinster, R.L., Chen, H.Y., Trumbauer, M., Senear, A.W., Warren, R., and Palmiter, R.D. (1981). Somatic expression of herpes thymidine kinase in mice following injection of a fusion gene into eggs. *Cell* **27**, 223-231.

Brinster, R.L., Chen, H.Y., Trumbauer, M.E., and Avarbock, M.R. (1980). Translation of globin messenger RNA by the mouse ovum. *Nature* **283**, 499-501.

Brinster, R.L. (1974). The effect of cells transferred into the mouse blastocyst on subsequent development. *J Exptl Med* **140**, 1049-1056.

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.

Brown, A., Bernier, G., Mathieu, M., Rossant, J., and Kothary, R. (1995). The mouse dystonia musculorum gene is a neural isoform of bullous pemphigoid antigen 1. *Nature Genet.* **10**, 301-306.

Brown, S.D. and Peters, J. (1996). Combining mutagenesis and genomics in the mouse - closing the phenotype gap. [Review] [22 refs]. *Trends. Genetics* 12, 433-435.

Brustle, O., Jones, K.N., Learish, R.D., Karram, K., Choudhary, K., Wiestler, O.D., Duncan, I.D., and McKay, R.D. (1999). Embryonic stem cell-derived glial precursors: a source of myelinating transplants [see comments]. *Science* 285, 754-756.

Brustle, O., Spiro, A.C., Karram, K., Choudhary, K., Okabe, S., and McKay, R.D. (1997). *In vitro*-generated neural precursors participate in mammalian brain development. *Proc. Natl. Acad. Sci. USA.* 94, 14809-14814.

Buehr, M. (1997). The primordial germ cells of mammals: some current perspectives. [Review] [129 refs]. *Exp. Cell Res.* 232, 194-207.

Burdon, T., Chambers, I., Stracey, C., Niwa, H., and Smith, A. (1999a). Signaling mechanisms regulating self-renewal and differentiation of pluripotent embryonic stem cells. [Review] [97 refs]. *Cells Tissues Organs* 165, 131-143.

Burdon, T., Stracey, C., Chambers, I., Nichols, J., and Smith, A. (1999b). Suppression of SHP-2 and ERK signalling promotes self-renewal of mouse embryonic stem cells. *Dev. Biol.* 210, 30-43.

Camus, A., Kress, C., Babinet, C., and Barra, J. (1996). Unexpected behavior of a gene trap vector comprising a fusion between the Sh ble and the lacZ genes. *Mol. Reprod. Dev.* 45, 255-263.

Candia, A.F., Hu, J., Crosby, J., Lalley, P.A., Noden, D., Nadeau, J.H., and Wright, C.V. (1992). Mox-1 and Mox-2 define a novel homeobox gene subfamily and are differentially expressed during early mesodermal patterning in mouse embryos. *Development* 116, 1123-1136.

Capecchi, M.R. (1989). Altering the genome by homologous recombination. [Review] [24 refs]. *Science* 244, 1288-1292.

Carrano, A.C., Eytan, E., Hershko, A., and Pagano, M. (1999). SKP2 is required for ubiquitin-mediated degradation of the CDK inhibitor p27. *Nature Cell Biol.* 1, 193-199.

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

Casadaban, M.J. and Cohen, S.N. (1979). Lactose genes fused to exogenous promoters in one step using a Mu-lac bacteriophage: *in vivo* probe for transcriptional control sequences. *Proc. Natl. Acad. Sci. USA.* 76, 4530-4533.

Cenciarelli, C., Chiaur, D.S., Guardavaccaro, D., Parks, W., Vidal, M., and Pagano, M. (1999). Identification of a family of human F-box proteins. *Curr. Biol.* 9, 1177-1179.

Chabot, B., Stephenson, D.A., Chapman, V.M., Besmer, P., Bernstein, and A. (1988). The proto-oncogene c-kit encoding a transmembrane tyrosine kinase receptor maps to the mouse W locus. *Nature* 335, 88-89.

Chambers, C.A. (1994). TKO'ed: lox, stock and barrel. *Bioessays* 16, 865-868.

Chang, W., Hubbard, S.C., Friedel, C., and Ruley, H.E. (1993). Enrichment of insertional mutants following retrovirus gene trap selection. *Virology* 193, 737-747.

Chapman, D.L. and Wolgemuth, D.J. (1994). Regulation of M-phase promoting factor activity during development of mouse male germ cells. *Dev. Biol.* **165**, 500-506.

Chapman, D.L. and Wolgemuth, D.J. (1993). Isolation of the murine cyclin B2 cDNA and characterization of the lineage and temporal specificity of expression of the B1 and B2 cyclins during oogenesis, spermatogenesis and early embryogenesis. *Development* **118**, 229-240.

Chapman, D.L. and Wolgemuth, D.J. (1992). Identification of a mouse B-type cyclin which exhibits developmentally regulated expression in the germ line. *Mol. Reprod. Dev.* **33**, 259-269.

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 tyrosine kinase by gene trap retroviral insertion. *Oncogene* **12**, 979-988.

Chen, W.S., Manova, K., Weinstein, D.C., Duncan, S.A., Plump, A.S., Prezioso, V.R., Bachvarova, R.F., and Darnell, J.E., Jr. (1994). Disruption of the HNF-4 gene, expressed in visceral endoderm, leads to cell death in embryonic ectoderm and impaired gastrulation of mouse embryos. *Genes. Dev.* **8**, 2466-2477.

Chen, Z. (1996). Simple modifications to increase specificity of the 5' RACE procedure. *Trends. Genetics* **12**, 87-88.

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. *Genes. Dev.* **8**, 2293-2301.

Chiquoine, A. D. (1954). The identification, origin and migration of the primordial germ cells in the mouse embryo. *Anat. Rec.* **118**, 135-146.

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. *Nucl. Acids. Res.* **25**, 1531-1536.

Chu, G. and Sharp, P.A. (1981). A gene chimaera of SV40 and mouse beta-globin is transcribed and properly spliced. *Nature* **289**, 378-382.

Chung, C.T., Niemela, S.L., and Miller, R.H. (1989). One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. USA.* **86**, 2172-2175.

Cibelli, J.B., Stice, S.L., Golueke, P.J., Kane, J.J., Jerry, J., Blackwell, C., Ponce de Leon, F.A., and Robl, J.M. (1998). Cloned transgenic calves produced from nonquiescent fetal fibroblasts. *Science* **280**, 1256-1258.

Clarke, D.L., Johansson, C.B., Wilbertz, J., Veress, B., Nilsson, E., Karlstrom, H., Lendahl, U., and Frisen, J. (2000). Generalized potential of adult neural stem cells. *Science* **288**, 1660-1663.

Conlon, F.L., Barth, K.S., and Robertson, E.J. (1991). A novel retrovirally induced embryonic lethal mutation in the mouse: assessment of the developmental fate of embryonic stem cells homozygous for the 413.d proviral integration. *Development* **111**, 969-981.

Conlon, R.A., Reaume, A.G., and Rossant, J. (1995). Notch1 is required for the coordinate segmentation of somites. *Development* **121**, 1533-1545.

Conover, J.C., Ip, N.Y., Poueymirou, W.T., Bates, B., Goldfarb, M.P., DeChiara, T.M., and Yancopoulos, G.D. (1993). Ciliary neurotrophic factor maintains the pluripotentiality of embryonic stem cells. *Development* 119, 559-565.

Cooke, J., Nowak, M.A., Boerlijst, M., and Maynard-Smith, J. (1997). Evolutionary origins and maintenance of redundant gene expression during metazoan development. [Review] [44 refs]. *Trends. Genetics* 13, 360-364.

Cooley, L., Kelley, R., and Spradling, A. (1988). Insertional mutagenesis of the Drosophila genome with single P elements. [Review] [45 refs]. *Science* 239, 1121-1128.

Copeland, N.G., Jenkins, N.A., Gilbert, D.J., Eppig, J.T., Maltais, L.J, Miller, J.C., Dietrich, W.F., Weaver, A., Lincoln, S.E., Steen, R.G., and et al. (1993). A genetic linkage map of the mouse: current applications and future prospects. [Review] [81 refs]. *Science* 262, 57-66.

Copp, A.J. (1979). Interaction between inner cell mass and trophectoderm of the mouse blastocyst. II. The fate of the polar trophectoderm. *J. Embryol. Exp. Morph.* 51, 109-120.

Copp, A.J. (1978). Interaction between inner cell mass and trophectoderm of the mouse blastocyst. I. A study of cellular proliferation. *J. Embryol. Exp. Morph.* 48, 109-125.

Costa, G.L.&. Weiner M.P (1995). Cloning and analysis of PCR-generated DNA fragments. In PCR Primer. Dieffenbach C.W & Dvesksler G.S., editors: CSH Laboratory Press, Cold Spring Harbor, NY, USA. 555-580.

Costantini, F. and Lacy, E. (1981). Introduction of a rabbit beta-globin gene into the mouse germ line. *Nature* 294, 92-94.

Coucouvanis, E. and Martin, G.R. (1999). BMP signaling plays a role in visceral endoderm differentiation and cavitation in the early mouse embryo. *Development* 126, 535-546.

Coucouvanis, E. and Martin, G.R. (1995). Signals for death and survival: a two-step mechanism for cavitation in the vertebrate embryo. *Cell* 83, 279-287.

Couldrey, C., Carlton, M.B., Ferrier, J., Colledge, W.H., and Evans, M.J. (1998). Disruption of murine alpha-enolase by a retroviral gene trap results in early embryonic lethality. *Dev. Dynamics* 212, 284-292.

Craig, K.L. and Tyers, M. (1999). The F-box: a new motif for ubiquitin dependent proteolysis in cell cycle regulation and signal transduction. [Review] [160 refs]. *Progr. Biophys. Mol. Biol.* 72, 299-328.

Craigie, R. (1992). Hotspots and warm spots: integration specificity of retroelements. [Review] [27 refs]. *Trends. Genetics* 8, 187-190.

Cross, J.C. (1998). Formation of the placenta and extraembryonic membranes. [Review] [59 refs]. *Ann. NY. Acad. Sci.* 857, 23-32.

Crossley, P.H. and Martin, G.R. (1995). The mouse Fgf8 gene encodes a family of polypeptides and is expressed in regions that direct outgrowth and patterning in the developing embryo. *Development* 121, 439-451.

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. *Dev. Biol.* 170, 664-678.

Damjanov, I. and Solter, D. (1974). Experimental teratoma. [Review] [3 refs]. *Curr. Topics. Pathol.* 59, 69-130.

Damjanov, I., Solter, D., and Skreb, N. (1971). Teratocarcinogenesis as related to the age of embryos grafted under the kidney capsule. *Roux's Arch. Dev. Biol.* **173**, 228-234.

Dani, C., Chambers, I., Johnstone, S., Robertson, M., Ebrahimi, B, Saito, M., Taga, T., Li, M., Burdon, T., Nichols, J., and Smith, A. (1998). Paracrine induction of stem cell renewal by LIF-deficient cells: a new ES cell regulatory pathway. *Dev. Biol.* **203**, 149-162.

Dani, C., Smith, A.G., Dessolin, S., Leroy, P., Staccini, L., Villageois, P., Darimont, C., and Ailhaud, G. (1997). Differentiation of embryonic stem cells into adipocytes *in vitro*. *J. Cell. Sci.* **110**, 1279-1285.

Daniels, R., Zuccotti, M., Kinis, T., Serhal, P., and Monk, M. (1997). XIST expression in human oocytes and preimplantation embryos. *Am. J. Hum. Genet.* **61**, 33-39.

Darnell, J.E., Jr. (1997). STATs and gene regulation. [Review] [99 refs]. *Science* **277**, 1630-1635.

Darnell, J.E., Jr., Kerr, I.M., and Stark, G.R. (1994). Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. [Review] [97 refs]. *Science* **264**, 1415 1421.

Davidson, D.R. and Hill, R.E. (1991). Msh-like genes: A family of homeobox genes with wide ranging expression during vertebrate development. *Semin. Dev. Biol.* **2**, 405-414.

Davis, S., Aldrich, T.H., Stahl, N., Pan, L., Taga, T., Kishimoto, T., Ip, N.Y., and Yancopoulos, G.D. (1993). LIFR beta and gp130 as heterodimerizing signal transducers of the tripartite CNTF receptor. *Science* **260**, 1805-1808.

Dean, W., Bowden, L., Aitchison, A., Klose, J., Moore, T., Meneses, J.J., Reik, W., and Feil, R. (1998). Altered imprinted gene methylation and expression in completely ES cell-derived mouse fetuses: association with aberrant phenotypes. *Development* **125**, 2273-2282.

DeGregori, J., Russ, A., von Melchner, H., Rayburn, H., Priyaranjan, P., Jenkins, N.A., Copeland, N.G., and Ruley, H.E. (1994). A murine homolog of the yeast RNA1 gene is required for postimplantation development. *Genes. Dev.* **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.

Deshaies, R.J., Chau, V., and Kirschner, M. (1995). Ubiquitination of the G1 cyclin Cln2p by a Cdc34p-dependent pathway. *EMBO. J.* **14**, 303-312.

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W., Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J., and Simpson, A.J. (2000). Shotgun sequencing of the human transcriptome with ORF expressed sequence tags. *Proc. Natl. Acad. Sci. USA.* **97**, 3491-3496.

Dietrich, W.F., Miller, J., Steen, R., Merchant, M.A., Damron-Boles, D., Husain, Z., Dredge, R., Daly, M.J., Ingalls, K.A., O'Connor, T.J., and et al. (1996). A comprehensive genetic map of the mouse genome [published erratum appears in Nature 1996 May 9,381(6578):172]. *Nature* **380**, 149-152.

Ding, J., Yang, L., Yan, Y.T., Chen, A., Desai, N., Wynshaw-Boris, A, and Shen, M.M. (1998). Cripto is required for correct orientation of the anterior-posterior axis in the mouse embryo. *Nature* **395**, 702-707.

Diwan, S.B. and Stevens, L.C. (1976). Development of teratomas from the ectoderm of mouse egg cylinders. *J. Nat. Cancer. Inst.* **57**, 937-942.

Dix, D.J., Allen, J.W., Collins, B.W., Mori, C., Nakamura, N., Poorman-Allen, P., Goulding, E.H., and Eddy, E.M. (1996). Targeted gene disruption of Hsp70-2 results in failed meiosis, germ cell apoptosis, and male infertility. *Proc. Natl. Acad. Sci. USA.* **93**, 3264-3268.

Dixon, K.E. (1994). Evolutionary aspects of primordial germ cell formation. [Review] [80 refs]. *Ciba Foundation Symposium* **182**, 92-110; discussion 110-120.

Doetschman, T., Gregg, R.G., Maeda, N., Hooper, M.L., Melton, D.W., Thompson, S., and Smithies, O. (1987). Targeted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature* **330**, 576-578.

Doetschman, T., Gossler, A., Serfling, E., Schaffner, W., Marcu, K, Stanton, L., and Kemler, R. (1986). Introduction of genes into mouse embryonic stem cells. *Progr. Clin. Biol. Res.* **217A**, 47-50.

Doetschman, T.C., Eistetter, H., Katz, M., Schmidt, W., and Kemler, 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.

Dressler, G.R. and Gruss, P. (1988). Do multigene families regulate vertebrate development?. [Review] [42 refs]. *Trends. Genetics* **4**, 214-219.

Driever, W., Solnica-Krezel, L., Schier, A.F., Neuhauss, S.C., Malicki, J., Stemple, D.L., Stainier, D.Y., 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.

Dunwoodie, S.L., Rodriguez, T.A., and Beddington, R.S. (1998). Msg1 and Mrg1, founding members of a gene family, show distinct patterns of gene expression during mouse embryogenesis. *Mech. Dev.* **72**, 27-40.

Dush, M.K. and Martin, G.R. (1992). Analysis of mouse Evx genes: Evx-1 displays graded expression in the primitive streak. *Dev. Biol.* **151**, 273-287.

Eddy, E.M. and O'Brien, D.A. (1998). Gene expression during mammalian meiosis. [Review] [336 refs]. *Curr. Topics. Dev. Biol.* **37**, 141-200.

Evans, M.J. (1998). Gene trapping--a preface. *Dev. Dynamics* **212**, 167-169.

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

Evans, M.J. (1972). The isolation and properties of a clonal tissue culture strain of pluripotent mouse teratoma cells. *J. Embryol. Exp. Morph.* **28**, 163-176.

Faisst, A.M. and Gruss, P. (1998). Bootnin: a novel murine gene expressed in restricted areas of the brain. *Dev. Dynamics* **212**, 293-303.

Falconer, D.S. and Avery, P.J. (1978). Variability of chimaeras and mosaics. *J. Embryol. Exp. Morph.* **43**, 195-219.

Fan, Y., Melhem, M.F., and Chaillet, J.R. (1999). Forced expression of the homeobox-containing gene Pem blocks differentiation of embryonic stem cells. *Dev. Biol.* **210**, 481-496.

Faria, T.N., LaRosa, G.J., Wilen, E., Liao, J., and Gudas, L.J. (1998). Characterization of genes which exhibit reduced expression during the retinoic acid-induced differentiation of F9 teratocarcinoma cells: involvement of cyclin D3 in RA-mediated growth arrest. *Mol. Cell. Endocrinol.* **143**, 155-166.

Faust, C. and Magnuson, T. (1993). Genetic control of gastrulation in the mouse. [Review] [74 refs]. *Curr. Opin. Genet. Devel.* **3**, 491-498.

Favor, J., Neuhauser-Klaus, A., and Ehling, U.H. (1990). The frequency of dominant cataract and recessive specific-locus mutations and mutation mosaics in F1 mice derived from post-spermatogonial treatment with ethylnitrosourea. *Mutation Research* **229**, 105-114.

Feinberg, A.P. and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**, 6-13.

Feldman, B., Poueymirou, W., Papaioannou, V.E., DeChiara, T.M., and Goldfarb, M. (1995). Requirement of FGF-4 for postimplantation mouse development. *Science* **267**, 246-249.

Feldman, R.M., Correll, C.C., Kaplan, K.B., and Deshaies, R.J. (1997). A complex of Cdc4p, Skp1p, and Cdc53p/cullin catalyzes ubiquitination of the phosphorylated CDK inhibitor Sic1p. *Cell* **91**, 221-230.

Feng, D.F. and Doolittle, R.F. (1987). Progressive sequence alignment as a prerequisite to correct phylogenetic trees. *J. Mol. Evolution* **25**, 351-360.

Ferrari, G., Cusella-De Angelis, G., Coletta, M., Paolucci, E., Stornaiuolo, A., Cossu, G., and Mavilio, F. (1998). Muscle regeneration by bone marrow-derived myogenic progenitors [published erratum appears in Science 1998 Aug 14; 281(5379):923]. *Science* **279**, 1528-1530.

Finley, M.F., Kulkarni, N., and Huettner, J.E. (1996). Synapse formation and establishment of neuronal polarity by P19 embryonic carcinoma cells and embryonic stem cells. *J. Neuroscience* **16**, 1056-1065.

Flach, G., Johnson, M.H., Braude, P.R., Taylor, R.A., and Bolton, V.N. (1982). The transition from maternal to embryonic control in the 2-cell mouse embryo. *EMBO. J.* **1**, 681-686.

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.

Fourney, R. M., J. Miyakoshi, R. S. Day III, and M. C. Paterson. (1988). Northern blotting: Efficient RNA staining and transfer. *Focus* 10:5-7.(Abstr.)

Fraichard, A., Chassande, O., Bilbaut, G., Dehay, C., Savatier, P., and Samarut, J. (1995). *In vitro* differentiation of embryonic stem cells into glial cells and functional neurons. *J. Cell. Sci.* **108**, 3181-3188.

Freemont, P.S. (2000). RING for destruction?. [Review] [20 refs]. *Curr. Biol.* **10**, R84-7.

Fried, M., Griffiths, M., Davies, B., Bjursell, G., La Mantia, G., and Lania, L. (1983). Isolation of cellular DNA sequences that allow expression of adjacent genes. *Proc. Natl. Acad. Sci. USA.* **80**, 2117-2121.

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.

Frohman, M.A. (1995). Rapid amplification of cDNA ends. In PCR Primer. Dieffenbach C.W. & Dvesksler G.S, editors. CSH laboratory press, Cold Spring Harbor, NY, USA. 381-410.

199

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.

Fuchs, E. and Segre, J.A. (2000). Stem cells: a new lease on life. [Review] [96 refs]. *Cell* **100**, 143-155.

Fujii, J.T. and Martin, G.R. (1983). Developmental potential of teratocarcinoma stem cells in utero following aggregation with cleavage-stage mouse embryos. *J. Embryol. Exp. Morph.* **74**, 79-96.

Fukada, T., Hibi, M., Yamanaka, Y., Takahashi-Tezuka, M., Fujitani, Y., Yamaguchi, T., Nakajima, K., and Hirano, T. (1996). Two signals are necessary for cell proliferation induced by a cytokine receptor gp130: involvement of STAT3 in anti-apoptosis. *Immunity* **5**, 449-460.

Furano, A.V. (2000). The biological properties and evolutionary dynamics of mammalian LINE-1 retrotransposons. [Review] [163 refs]. *Prg. Nucl. Acids. Res. Mol. Biol.* **64**, 255-294.

Furukawa, T., Kimura, K., Kobayakawa, Y., Tamura, K., Kawaichi, M., Tanimura, T., and Honjo, T. (1994). Genetic characterization of Drosophila RBP-J kappa (suppressor of hairless) as a neurogenic gene in adult PNS development. *Jpn. J. Genet.* **69**, 701-711.

Galan, J.M. and Peter, M. (1999). Ubiquitin-dependent degradation of multiple F-box proteins by an autocatalytic mechanism. *Proc. Natl. Acad. Sci. USA.* **96**, 9124-9129.

Gardner, R.L. and Beddington, R.S. (1988). Multi-lineage 'stem' cells in the mammalian embryo. [Review] [133 refs]. *J. Cell. Sci. Supp.* **10**, 11-27.

Gardner, R.L., Davies, T.J., and Carey, M.S. (1988). Effect of delayed implantation on differentiation of the extra-embryonic endoderm in the mouse blastocyst. *Placenta* **9**, 343-359.

Gardner, R.L., Lyon, M.F., Evans, E.P., and Burtenshaw, M.D. (1985). Clonal analysis of X-chromosome inactivation and the origin of the germ line in the mouse embryo. *J. Embryol. Exp. Morph.* **88**, 349-363.

Gardner, R.L. (1983). Origin and differentiation of extraembryonic tissues in the mouse. *Int. Rev. Exp. Pathol.* **24**, 63-133.

Gardner, R.L. and Rossant, J. (1979). Investigation of the fate of 4-5 day post-coitum mouse inner cell mass cells by blastocyst injection. *J. Embryol. Exp. Morph.* **52**, 141-152.

Gardner, R.L & Papaioannou V.E. (1975). Differentiation in the trophectoderm and inner cell mass. In The early development of mammals. M. Ball and A.E. Wild, editors. Cambridge University Press, Cambridge, UK. 107-132.

Gardner, R.L. and Johnson, M.H. (1972). An investigation of inner cell mass and trophoblast tissues following their isolation from the mouse blastocyst. *J. Embryol. Exp. Morph.* **28**, 279-312.

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. Genet.* **17**, 141-154.

Gearing, D.P. and Bruce, A.G. (1992). Oncostatin M binds the high-affinity leukemia inhibitory factor receptor. *New Biologist* **4**, 61-65.

Gearing, D.P., Thut, C.J., VandeBos, T., Gimpel, S.D., Delaney, P.B., King, J., Price, V., Cosman, D., and Beckmann, M.P. (1991). Leukemia inhibitory factor receptor is structurally related to the IL-6 signal transducer, gp130. *EMBO. J.* **10**, 2839-2848.

Gearing, D.P., Gough, N.M., King, J.A., Hilton, D.J., Nicola, N.A., Simpson, R.J., Nice, E.C., Kelso, A., and Metcalf, D. (1987). Molecular cloning and expression of cDNA encoding a murine myeloid leukaemia inhibitory factor (LIF). *EMBO. J.* 6, 3995-4002.

Geissler, E.N., Ryan, M.A., and Housman, D.E. (1988). The dominant-white spotting (W) locus of the mouse encodes the c-kit proto-oncogene. *Cell* 55, 185-192.

Ghattas, I.R., Sanes, J.R., and Majors, J.E. (1991). The encephalomyocarditis virus internal ribosome entry site allows efficient coexpression of two genes from a recombinant provirus in cultured cells and in embryos. *Mol. Cell. Biol.* 11, 5848-5859.

Ginsburg, M., Snow, M.H., and McLaren, A. (1990). Primordial germ cells in the mouse embryo during gastrulation. *Development* 110, 521-528.

Gogos, J.A., Thompson, R., Lowry, W., Sloane, B.F., Weintraub, H., and Horwitz, M. (1996). Gene trapping in differentiating cell lines: regulation of the lysosomal protease cathepsin B in skeletal myoblast growth and fusion. *J. Cell. Biol.* 134, 837-847.

Gordon, J.W. and Ruddle, F.H. (1981). Integration and stable germ line transmission of genes injected into mouse pronuclei. *Science* 214, 1244-1246.

Gordon, J.W., Scangos, G.A., Plotkin, D.J., Barbosa, J.A., and Ruddle, F.H. (1980). Genetic transformation of mouse embryos by microinjection of purified DNA. *Proc. Natl. Acad. Sci. USA.* 77, 7380-7384.

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., Doetschman, T., Korn, R., Serfling, E., and Kemler, R. (1986). Transgenesis by means of blastocyst-derived embryonic stem cell lines. *Proc. Natl. Acad. Sci. USA.* 83, 9065-9069.

Graham, C.F. (1977). Teratocarcinoma cells and normal mouse embryogenesis. In Concepts in mammalian embryogenesis. M.I. Sherman, editor. MIT Press, Cambridge, MA, USA. 315-394.

Gray, W.M., del Pozo, J.C., Walker, L., Hobbie, L., Risseeuw, E., Banks, T., Crosby, W.L., Yang, M., Ma, H., and Estelle, M. (1999). Identification of an SCF ubiquitin-ligase complex required for auxin response in Arabidopsis thaliana. *Genes. Dev.* 13, 1678-1691.

Greenwald, I. (1998). LIN-12/Notch signaling: lessons from worms and flies. [Review] [117 refs]. *Genes. Dev.* 12, 1751-1762.

Gridley, T. (1991). Insertional versus targeted mutagenesis in mice. [Review] [81 refs]. *New Biologist* 3, 1025-1034.

Gridley, T., Gray, D.A., Orr-Weaver, T., Soriano, P., Barton, D.E., Francke, U., and Jaenisch, R. (1990). Molecular analysis of the Mov 34 mutation: transcript disrupted by proviral integration in mice is conserved in Drosophila. *Development* 109, 235-242.

Gridley, T., P. Soriano & Jaenisch.R (1987). Insertional mutagenesis in mice. *Trends. Genetics* 3, 162-166.

Grossman, S.R., Johannsen, E., Tong, X., Yalamanchili, R., and Kieff, E. (1994). The Epstein-Barr virus nuclear antigen 2 transactivator is directed to response elements by the J kappa recombination signal binding protein. *Proc. Natl. Acad. Sci. USA.* 91, 7568-7572.

Grunstein, M. and Hogness, D.S. (1975). Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. *Proc. Natl. Acad. Sci. USA.* 72, 3961-3965.

Gu, H., Marth, J.D., Orban, P.C., Mossmann, H., and Rajewsky, K. (1994). Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting. *Science* 265, 103-106.

Gu, H., Zou, Y.R., and Rajewsky, K. (1993). Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting. *Cell* 73, 1155-1164.

Gubbay, J., Collignon, J., Koopman, P., Capel, B., Economou, A., Munsterberg, A., Vivian, N., Goodfellow, P., and Lovell-Badge, R. (1990). A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature* 346, 245-250.

Guillemot, F., Nagy, A., Auerbach, A., Rossant, J., and Joyner, A.L. (1994). Essential role of Mash-2 in extraembryonic development. *Nature* 371, 333-336.

Gussoni, E., Soneoka, Y., Strickland, C.D., Buzney, E.A., Khan, M.K., Flint, A.F., Kunkel, L.M., and Mulligan, R.C. (1999). Dystrophin expression in the mdx mouse restored by stem cell transplantation. *Nature* 401, 390-394.

Haar, J.L. and Ackerman, G.A. (1971). Ultrastructural changes in mouse yolk sac associated with the initiation of vitelline circulation. *Anat. Rec.* 170, 437-455.

Habets, G.G., Scholtes, E.H., Zuydgeest, D., van der Kammen, R.A., Stam, J.C., Berns, A., and Collard, J.G. (1994). Identification of an invasion-inducing gene, Tiam-1, that encodes a protein with homology to GDP-GTP exchangers for Rho-like proteins. *Cell* 77, 537-549.

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

Hahnel, A.C., Rappolee, D.A., Millan, J.L., Manes, T., Ziomek, C.A., Theodosiou, N.G., Werb, Z., Pedersen, R.A., and Schultz, G.A. (1990). Two alkaline phosphatase genes are expressed during early development in the mouse embryo. *Development* 110, 555-564.

Hamada, H. (1986a). Random isolation of gene activator elements from the human genome. *Mol. Cell. Biol.* 6, 4185-4194.

Hamada, H. (1986b). Activation of an enhancerless gene by chromosomal integration. *Mol. Cell. Biol.* 6, 4179-4184.

Handyside, A. H., O'Neil., Jones M & Hooper M.L. (1989). Use of BRL-conditioned medium in combination with feeder layers to isolate a diploid embryonal stem cell line. *Roux's Arch. Dev. Biol.* 198, 48-56.

Handyside, A.H. (1978). Time of commitment of inside cells isolated from preimplantation mouse embryos. *J. Embryol. Exp. Morph.* 45, 37-53.

Harbers, K., Kuehn, M., Delius, H., and Jaenisch, R. (1984). Insertion of retrovirus into the first intron of alpha 1 (I) collagen gene to embryonic lethal mutation in mice. *Proc. Natl. Acad. Sci. USA.* 81, 1504-1508.

Harbers, K., Jahner, D., and Jaenisch, R. (1981). Microinjection of cloned retroviral genomes into mouse zygotes: integration and expression in the animal. *Nature* 293, 540-542.

Hardouin, N. and Nagy, A. (2000). Gene-trap-based target site for cre-mediated transgenic insertion [letter]. *Genesis.* **26**, 245-252.

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

Hartenstein, V., Younossi-Hartenstein, A., and Lekven, A. (1994). Delamination and division in the Drosophila neurectoderm: spatiotemporal pattern, cytoskeletal dynamics, and common control by neurogenic and segment polarity genes. *Dev. Biol.* **165**, 480-499.

Hasty, P., Ramirez-Solis, R., Krumlauf, R., and Bradley, A. (1991). Introduction of a subtle mutation into the Hox-2.6 locus in embryonic stem cells [published erratum appears in Nature 1991, 353(6339):94]. *Nature* **350**, 243-246.

Hellen, C.U., Witherell, G.W., Schmid, M., Shin, S.H., Pestova, T.V., Gil, A., and Wimmer, E. (1993). A cytoplasmic 57-kDa protein that is required for translation of picornavirus RNA by internal ribosomal entry is identical to the nuclear pyrimidine tract-binding protein. *Proc. Natl. Acad. Sci. USA.* **90**, 7642-7646.

Henkel, T., Ling, P.D., Hayward, S.D., and Peterson, M.G. (1994). Mediation of Epstein-Barr virus EBNA2 transactivation by recombination signal-binding protein J kappa. *Science* **265**, 92-95.

Hermesz, E., Mackem, S., and Mahon, K.A. (1996). Rpx: a novel anterior-restricted homeobox gene progressively activated in the prechordal plate, anterior neural plate and Rathke's pouch of the mouse embryo. *Development* **122**, 41-52.

Herrmann, B.G., Labeit, S., Poustka, A., King, T.R., and Lehrach, H. (1990). Cloning of the T gene required in mesoderm formation in the mouse. *Nature* **343**, 617-622.

Hershko, A. and Ciechanover, A. (1998). The ubiquitin system. [Review] [298 refs]. *Annu. Rev. Biochem.* **67**, 425-479.

Hershko, A. (1997). Roles of ubiquitin-mediated proteolysis in cell cycle control. [Review] [86 refs]. *Curr. Opin. Cell. Biol.* **9**, 788-799.

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 Genet.* **16**, 338-344.

Hill, D.P. and Wurst, W. (1993). Screening for novel pattern formation genes using gene trap approaches. *Methods in Enzymology* **225**, 664-681.

Hilton, D.J., Richardson, R.T., Alexander, W.S., Viney, E.M., Willson, T.A., Sprigg, N.S., Starr, R., Nicholson, S.E., Metcalf, D., and Nicola, N.A. (1998). Twenty proteins containing a C-terminal SOCS box form five structural classes. *Proc. Natl. Acad. Sci. USA.* **95**, 114-119.

Hilton, D.J. and Gough, N.M. (1991). Leukemia inhibitory factor: a biological perspective. [Review] [58 refs]. *J. Cell. Biochem.* **46**, 21-26.

Hitotsumachi, S., Carpenter, D.A., and Russell, W.L. (1985). Dose-repetition increases the mutagenic effectiveness of N-ethyl-N-nitrosourea in mouse spermatogonia. *Proc. Natl. Acad. Sci. USA.* **82**, 6619-6621.

Hochstrasser, M. (1996). Ubiquitin-dependent protein degradation. [Review] [115 refs]. *Annu. Rev. Genet.* **30**, 405-439.

Hodgkinson, C.A., Moore, K.J., Nakayama, A., Steingrimsson, E., Copeland, N.G., Jenkins, N.A., and Arnheiter, H. (1993). Mutations at the mouse microphthalmia locus are associated with defects in a gene encoding a novel basic-helix-loop-helix-zipper protein. *Cell* **74**, 395-404.

Hogan, B., R. Beddington, F. Costantini, & Lacy E. (1994). Manipulating the mouse embryo. CSH Laboratory Press, Cold Spring Harbor, NY, USA.

Hogan, B. and Tilly, R. (1978). *In vitro* development of inner ce•' masses isolated immunosurgically from mouse blastocysts. I. Inner cell masses from 3.5-day p.c. blastocysts incubated for 24 h before immunosurgery. *J. Embryol. Exp. Morph.* **45**, 93-105.

Holzschu, D., Lapierre, L., Neubaum, D., and Mark, W.H. (1997). A molecular strategy designed for the rapid screening of gene traps based on sequence identity and gene expression pattern in adult mice. *Trans. Res.* **6**, 97-106.

Honjo, T. (1996). The shortest path from the surface to the nucleus: RBP-J kappa/Su(H) transcription factor. [Review] [49 refs]. *Genes to Cells* **1**, 1-9.

Hooper, M., Hardy, K., Handyside, A., Hunter, S., and Monk, M. (1987). HPRT-deficient (Lesch-Nyhan) mouse embryos derived from germline colonization by cultured cells. *Nature* **326**, 292-295.

Hrabe de Angelis, M., McIntyre, J., 2nd, and Gossler, A. (1997). Maintenance of somite borders in mice requires the Delta homologue Dll1. *Nature* **386**, 717-721.

Huang, E., Nocka, K., Beier, D.R., Chu, T.Y., Buck, J., Lahm, H.W., Wellner, D., Leder, P., and Besmer, P. (1990). The hematopoietic growth factor KL is encoded by the Sl locus and is the ligand of the c-kit receptor, the gene product of the W locus. *Cell* **63**, 225-233.

Hubbard, E.J., Wu, G., Kitajewski, J., and Greenwald, I. (1997). sel-10, a negative regulator of lin-12 activity in Caenorhabditis elegans, encodes a member of the CDC4 family of proteins. *Genes. Dev.* **11**, 3182-3193.

Hynes, R.O. (1992). Integrins: versatility, modulation, and signaling in cell adhesion. [Review] [180 refs]. *Cell* **69**, 11-25.

Ishibashi, M., Ang, S.L., Shiota, K., Nakanishi, S., Kageyama, R., and Guillemot, F. (1995). Targeted disruption of mammalian hairy and Enhancer of split homolog-1 (HES-1) leads to up-regulation of neural helix-loop-helix factors, premature neurogenesis, and severe neural tube defects. *Genes. Dev.* **9**, 3136-3148.

Ishida, Y. and Leder, P. (1999). RET: a poly A-trap retrovirus vector for reversible disruption and expression monitoring of genes in living cells. *Nucl. Acids. Res.* **27**, e35

Iwai, K., Yamanaka, K., Kamura, T., Minato, N., Conaway, R.C., Conaway, J.W., Klausner, R.D., and Pause, A. (1999). Identification of the von Hippel-lindau tumor-suppressor protein as part of an active E3 ubiquitin ligase complex *Proc. Natl. Acad. Sci. USA.* **96**, 12436-12441.

Jackson, K.A., Mi, T., and Goodell, M.A. (1999). Hematopoietic potential of stem cells isolated from murine skeletal muscle. *Proc. Natl. Acad. Sci. USA.* **96**, 14482-14486.

Jaenisch, R. (1988). Transgenic animals. [Review] [108 refs]. *Science* **240**, 1468-1474.

Jaenisch, R., Jahner, D., Nobis, P., Simon, I., Lohler, J., Harbers, K., and Grotkopp, D. (1981). Chromosomal position and activation of retroviral genomes inserted into the germ line of mice. *Cell* **24**, 519-529.

Jaenisch, R. (1976). Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus. *Proc. Natl. Acad. Sci. USA.* **73**, 1260-1264.

Jähner, D. and Jaenisch, R. (1985). Retrovirus-induced de novo methylation of flanking host sequences correlates with gene inactivity. *Nature* **315**, 594-597.

Jang, S.K. and Wimmer, E. (1990). Cap-independent translation of encephalomyocarditis virus RNA: structural elements of the internal ribosomal entry site and involvement of a cellular 57-kD RNA-binding protein. *Genes. Dev.* **4**, 1560-1572.

Jang, S.K., Davies, M.V., Kaufman, R.J., and Wimmer, E. (1989). Initiation of protein synthesis by internal entry of ribosomes into the 5' nontranslated region of encephalomyocarditis virus RNA *in vivo. J. Virol.* **63**, 1651-1660.

Jang, S.K., Krausslich, H.G., Nicklin, M.J., Duke, G.M., Palmenberg, A.C., and Wimmer, E. (1988). A segment of the 5' nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during *in vitro* translation. *J. Virol.* **62**, 2636-2643.

Jaquenoud, M., Gulli, M.P., Peter, K., and Peter, M. (1998). The Cdc42p effector Gic2p is targeted for ubiquitin-dependent degradation by the SCFGrr1 complex. *EMBO. J.* **17**, 5360-5373.

Jiang, J. and Struhl, G. (1998). Regulation of the Hedgehog and Wingless signalling pathways by the F-box/WD40-repeat protein Slimb. *Nature* **391**, 493-496.

Johnson, M.H. and Ziomek, C.A. (1981). The foundation of two distinct cell lineages within the mouse morula. *Cell* **24**, 71-80.

Jonsson, J.I., Wu, Q., Nilsson, K., and Phillips, R.A. (1996). Use of a promoter-trap retrovirus to identify and isolate genes involved in differentiation of a myeloid progenitor cell line *in vitro. Blood* **87**, 1771-1779.

Joyner, A.L., Auerbach, A., and Skarnes, W.C. (1992). The gene trap approach in embryonic stem cells: the potential for genetic screens in mice. [Review] [23 refs]. *Ciba Foundation Symposium* **165**, 277-288; discussion 288-289.

Joyner, A.L., Skarnes, W.C., and Rossant, J. (1989). Production of a mutation in mouse En-2 gene by homologous recombination in embryonic stem cells. *Nature* **338**, 153-156.

Kaiser, P., Sia, R.A., Bardes, E.G., Lew, D.J., and Reed, S.I. (1998). Cdc34 and the F-box protein Met30 are required for degradation of the Cdk-inhibitory kinase Swe1. *Genes. Dev.* **12**, 2587-2597.

Kamura, T., Koepp, D.M., Conrad, M.N., Skowyra, D., Moreland, R.J., Iliopoulos, O., Lane, W.S., Kaelin, W.G., Jr., Elledge, S.J., Conaway, R.C., Harper, J.W., and Conaway, J.W. (1999). Rbx1, a component of the VHL tumor suppressor complex and SCF ubiquitin ligase. *Science* **284**, 657-661.

Kamura, T., Sato, S., Haque, D., Liu, L., Kaelin, W.G., Jr., Conaway, R.C., and Conaway, J.W. (1998). The Elongin BC complex interacts with the conserved SOCS-box motif present in members of the SOCS, ras, WD-40 repeat, and ankyrin repeat families. *Genes. Dev.* **12**, 3872-3881.

Kaplan, K.B., Hyman, A.A., and Sorger, P.K. (1997). Regulating the yeast kinetochore by ubiquitin-dependent degradation and Skp1p-mediated phosphorylation. *Cell* **91**, 491-500.

Kato, Y., Rideout, W.M., 3rd, Hilton, K., Barton, S.C., Tsunoda, Y., and Surani, M.A. (1999). Developmental potential of mouse primordial germ cells. *Development* **126**, 1823-1832.

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

Kawaichi, M., Oka, C., Shibayama, S., Koromilas, A.E., Matsunami, N., Hamaguchi, Y., and Honjo, T. (1992). Genomic organization of mouse J kappa recombination signal binding protein (RBP-J kappa) gene. J. Biol. Chem. 267, 4016-4022.

Kawasaki, H., Mizuseki, K., Nishikawa, S., Kaneko, S., Kuwana, Y., Nakanishi, S., Nishikawa, S.I., and Sasai, Y. (2000). Induction of midbrain dopaminergic neurons from ES cells by stromal cell-derived inducing activity. Neuron 28(1), 31-40.

Keller, G.M. (1995). In vitro differentiation of embryonic stem cells. [Review] [50 refs]. Curr. Opin. Cell. Biol. 7, 862-869.

Keller, G., Kennedy, M., Papayannopoulou, T., and Wiles, M.V. (1993). Hematopoietic commitment during embryonic stem cell differentiation in culture. Mol. Cell. Biol. 13, 473-486.

Kelly, S.J. (1977). Studies of the developmental potential of 4- and 8-cell stage mouse blastomeres. J. Exptl. Zool. 200, 365-376.

Kemler, R. (1993). From cadherins to catenins: cytoplasmic protein interactions and regulation of cell adhesion. [Review] [42 refs]. Trends. Genetics 9, 317-321.

Kerr, W.G., Nolan, G.P., Serafini, A.T., and Herzenberg, L.A. (1989). Transcriptionally defective retroviruses containing lacZ for the in situ detection of endogenous genes and developmentally regulated chromatin. Cold Spring Harb. Symp. Quant. Biol. 54 Pt 2, 767-776.

Kessel, M. and Gruss, P. (1990). Murine developmental control genes. Science 249, 374-379.

Kidder, G.M. (1992). The genetic program for preimplantation development. [Review] [54 refs]. Dev. Genet. 13, 319-325.

Kim, D.G., Kang, H.M., Jang, S.K., and Shin, H.S. (1992). Construction of a bifunctional mRNA in the mouse by using the internal ribosomal entry site of the encephalomyocarditis virus [published erratum appears in Mol Cell Biol 1992 10:4807]. Mol. Cell. Biol. 12, 3636-3643.

Kimura, S., Niwa, H., Moriyama, M., Araki, K., Abe, K., Miike, T., and Yamamura, K. (1994). Improvement of germ line transmission by targeting b-galactosidase to nuclei in transgenic mice. Dev. Growth & Differ. 36, 521-527.

King, R.W., Deshaies, R.J., Peters, J.M., and Kirschner, M.W. (1996). How proteolysis drives the cell cycle. [Review] [120 refs]. Science 274, 1652-1659.

King, T.R., Dove, W.F., Herrmann, B., Moser, A.R., and Shedlovsky, A. (1989). Mapping to molecular resolution in the T to H-2 region of the mouse genome with a nested set of meiotic recombinants. Proc. Natl. Acad. Sci. USA. 86, 222-226.

Kingsley, D.M., Bland, A.E., Grubber, J.M., Marker, P.C., Russell, L.B., Copeland, N.G., and Jenkins, N.A. (1992). The mouse short ear skeletal morphogenesis locus is associated with defects in a bone morphogenetic member of the TGF beta superfamily. Cell 71, 399-410.

Kipreos, E.T., Lander, L.E., Wing, J.P., He, W.W., and Hedgecock, E.M. (1996). cul-1 is required for cell cycle exit in C. elegans and identifies a novel gene family. Cell 85, 829-839.

Klug, M.G., Soonpaa, M.H., Koh, G.Y., and Field, L.J. (1996). Genetically selected cardiomyocytes from differentiating embryonic stem cells form stable intracardiac grafts. J. Clin. Invest. 98, 216-224.

Kontgen, F., Suss, G., Stewart, C., Steinmetz, M., and Bluethmann, H. (1993). Targeted disruption of the MHC class II Aa gene in C57BL/6 mice. *International Immunology* 5, 957-964.

Koopman, P., Munsterberg, A., Capel, B., Vivian, N., and Lovell-Badge, R. (1990). Expression of a candidate sex-determining gene during mouse testis differentiation. *Nature* 348, 450-452.

Kopen, G.C., Prockop, D.J., and Phinney, D.G. (1999). Marrow stromal cells migrate throughout forebrain and cerebellum, and they differentiate into astrocytes after injection into neonatal mouse brains. *Proc. Natl. Acad. Sci. USA.* 96, 10711-10716.

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 chimaeric embryos with a high frequency and detect endogenous genes. *Mech. Dev.* 39, 95-109.

Koshimizu, U., Taga, T., Watanabe, M., Saito, M., Shirayoshi, Y., Kishimoto, T., and Nakatsuji, N. (1996). Functional requirement of gp130-mediated signaling for growth and survival of mouse primordial germ cells *in vitro* and derivation of embryonic germ (EG) cells. *Development* 122, 1235-1242.

Kothary, R., Clapoff, S., Brown, A., Campbell, R., Peterson, A., and Rossant, J. (1988). A transgene containing lacZ inserted into the dystonia locus is expressed in neural tube. *Nature* 335, 435-437.

Kozak, M. (1989). The scanning model for translation: an update. [Review] [287 refs]. *J. Cell. Biol.* 108, 229-241.

Kramer, A., Hochhaus, A., Saussele, S., Reichert, A., Willer, A., and Hehlmann, R. (1998). Cyclin A1 is predominantly expressed in hematological malignancies with myeloid differentiation. *Leukemia* 12, 893-898.

Krek, W. (1998). Proteolysis and the G1-S transition: the SCF connection. [Review] [59 refs]. *Curr. Opin. Genet. Devel.* 8, 36-42.

Kuehn, M.R., Bradley, A., Robertson, E.J., and Evans, M.J. (1987). A potential animal model for Lesch-Nyhan syndrome through introduction of HPRT mutations into mice. *Nature* 326, 295-298.

Labosky, P.A., Barlow, D.P., and Hogan, B.L. (1994a). Embryonic germ cell lines and their derivation from mouse primordial germ cells. [Review] [52 refs]. *Ciba Foundation Symposium* 182, 157-168; discussion 168-178.

Labosky, P.A., Barlow, D.P., and Hogan, B.L. (1994b). Mouse embryonic germ (EG) cell lines: transmission through the germline and differences in the methylation imprint of insulin-like growth factor 2 receptor (Igf2r) gene compared with embryonic stem (ES) cell lines. *Development* 120, 3197-3204.

Lagasse, E., Connors, H., Al-Dhalimy, M., Reitsma, M., Dohse, M., Osborne, L., Wang, X., Finegold, M., Weissman, I.L., and Grompe, M. (2000). Purified hematopoietic stem cells can differentiate into hepatocytes *in vivo*. *Nature Medicine* 6, 1229-1234.

Laird, P.W., Zijderveld, A., Linders, K., Rudnicki, M.A., Jaenisch, R, and Berns, A. (1991). Simplified mammalian DNA isolation procedure. *Nucl. Acids. Res.* 19, 4293

Lakso, M., Sauer, B., Mosinger, B., Jr., Lee, E.J., Manning, R.W., Yu, S.H., Mulder, K.L., and Westphal, H. (1992). Targeted oncogene activation by site-specific recombination in transgenic mice. *Proc. Natl. Acad. Sci. USA.* 89, 6232-6236.

Laney, J.D. and Hochstrasser, M. (1999). Substrate targeting in the ubiquitin system. [Review] [25 refs]. *Cell* **97**, 427-430.

Lardelli, M. (2000). Regulation of development in the zebrafish. *Australian Biochemist* **31**, 7-9.

Lawson, K.A., Dunn, N.R., Roelen, B.A., Zeinstra, L.M., Davis, A.M., Wright, C.V., Korving, J.P., and Hogan, B.L. (1999). Bmp4 is required for the generation of primordial germ cells in the mouse embryo. *Genes. Dev.* **13**, 424-436.

Lawson, K.A. and Hage, W.J. (1994). Clonal analysis of the origin of primordial germ cells in the mouse. [Review] [33 refs]. *Ciba Foundation Symposium* **182**, 68-84; discussion 84-91.

Lawson, K.A., Meneses, J.J., and Pedersen, R.A. (1991). Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development* **113**, 891-911.

Ledermann, B. and Burki, K. (1991). Establishment of a germ-line competent C57BL/6 embryonic stem cell line. *Exp. Cell Res.* **197**, 254-258.

Lee, K., Belinsky, M.G., Bell, D.W., Testa, J.R., and Kruh, G.D. (1998). Isolation of MOAT-B, a widely expressed multidrug resistance-associated protein/canalicular multispecific organic anion transporter-related transporter. *Cancer. Res.* **58**, 2741-2747.

Lei, W., Hirose, T., Zhang, L.X., Adachi, H., Spinella, M.J., Dmitrovsky, E., and Jetten, A.M. (1997). Cloning of the human orphan receptor germ cell nuclear factor/retinoid receptor-related testis-associated receptor and its differential regulation during embryonal carcinoma cell differentiation. *J. Mol. Endocrinol.* **18**, 167-176.

Lendahl, U. (1998). A growing family of Notch ligands. *Bioessays* **20**, 103-107.

Li, F.N. and Johnston, M. (1997). Grr1 of Saccharomyces cerevisiae is connected to the ubiquitin proteolysis machinery through Skp1: coupling glucose sensing to gene expression and the cell cycle. *EMBO. J.* **16**, 5629-5638.

Li, M., Pevny, L., Lovell-Badge, R., and Smith, A. (1998). Generation of purified neural precursors from embryonic stem cells by lineage selection. *Curr. Biol.* **8**, 971-974.

Li, M., Sendtner, M., and Smith, A. (1995). Essential function of LIF receptor in motor neurons. *Nature* **378**, 724-727.

Lisztwan, J., Imbert, G., Wirbelauer, C., Gstaiger, M., and Krek, W. (1999). The von Hippel-Lindau tumor suppressor protein is a component of an E3 ubiquitin-protein ligase activity. *Genes. Dev.* **13**, 1822-1833.

Lisztwan, J., Marti, A., Sutterluty, H., Gstaiger, M., Wirbelauer, C., and Krek, W. (1998). Association of human CUL-1 and ubiquitin-conjugating enzyme CDC34 with the F-box protein p45(SKP2): evidence for evolutionary conservation in the subunit composition of the CDC34-SCF pathway. *EMBO. J.* **17**, 368-383.

Liu, D., Matzuk, M.M., Sung, W.K., Guo, Q., Wang, P., and Wolgemuth, D.J. (1998). Cyclin A1 is required for meiosis in the male mouse. *Nature Genet.* **20**, 377-380.

Love, J.M., Knight, A.M., McAleer, M.A., and Todd, J.A. (1990). Towards construction of a high resolution map of the mouse genome using PCR-analysed microsatellites. *Nucl. Acids. Res.* **18**, 4123-4130.

Lyden, D., Young, A.Z., Zagzag, D., Yan, W., Gerald, W., O'Reilly, R, Bader, B.L., Hynes, R.O., Zhuang, Y., Manova, K., and Benezra, R. (1999). Id1 and Id3 are required for neurogenesis, angiogenesis and vascularization of tumour xenografts. *Nature* **401**, 670-677.

Lyon, M.F. (2000). LINE-1 elements and X chromosome inactivation: a function for "junk" DNA? *Proc. Natl. Acad. Sci. USA.* **97,** 6248-6249.

Lyon, M. F., Rastan S and Brown. S.D.M. (1996). Genetic variations and strains of the laboratory mouse. 3rd edition. Oxford University Press, Oxford, UK.

McClive, P., Pall, G., Newton, K., Lee, M., Mullins, J., and Forrester, L. (1998). Gene trap integrations expressed in the developing heart: insertion site affects splicing of the PT1-ATG vector. *Dev. Dynamics* **212,** 267-276.

McCoshen, J.A. and McCallion, D.J. (1975). A study of the primordial germ cells during their migratory phase in Steel mutant mice. *Experientia* **31,** 589-590.

McGrath, J. and Solter, D. (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* **37** 179-183.

McLaren, A. (1968). A study of blastocysts during delay and subsequent implantation in lactating mice. *J. Endocrinology* **42,** 453-463.

McWhir, J., Schnieke, A.E., Ansell, R., Wallace, H., Colman, A., Scott, A.R., and Kind, A.J. (1996). Selective ablation of differentiated cells permits isolation of embryonic stem cell lines from murine embryos with a non-permissive genetic background. *Nature Genet.* **14,** 223-226.

Mac Auley, A., Werb, Z., and Mirkes, P.E. (1993). Characterization of the unusually rapid cell cycles during rat gastrulation. *Development* **117,** 873-883.

Macleod, D., Lovell-Badge, R., Jones, S., and Jackson, I. (1991). A promoter trap in embryonic stem (ES) cells selects for integration of DNA into CpG islands. *Nucl. Acids. Res.* **19,** 17-23.

Mantalenakis, S.J. and Ketchel, M.M. (1966). Frequency and extent of delayed implantation in lactating rats and mice. *J. Reprod. Fertil.* **12,** 391-394.

Margottin, F., Bour, S.P., Durand, H., Selig, L., Benichou, S., Richard, V., Thomas, D., Strebel, K., and Benarous, R. (1998). A novel human WD protein, h-beta TrCp, that interacts with HIV-1 Vpu connects CD4 to the ER degradation pathway through an F-box motif. *Mol. Cell.* **1,** 565-574.

Marikawa, Y. and Elinson, R.P. (1998). beta-TrCP is a negative regulator of Wnt/beta-catenin signaling pathway and dorsal axis formation in Xenopus embryos. *Mech. Dev.* **77,** 75-80.

Marti, A., Wirbelauer, C., Scheffner, M., and Krek, W. (1999). Interaction between ubiquitin-protein ligase SCFSKP2 and E2F-1 underlies the regulation of E2F-1 degradation. *Nature Cell Biol.* **1,** 14-19.

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl. Acad. Sci. USA.* **78,** 7634-7638.

Martin, G.R. (1980). Teratocarcinomas and mammalian embryogenesis. [Review] [58 refs]. *Science* **209,** 768-776.

Martin, G.R., Wiley, L.M., and Damjanov, I. (1977). The development of cystic embryoid bodies *in vitro* from clonal teratocarcinoma stem cells. *Dev. Biol.* **61,** 230-244.

Martin, G.R. (1975). Teratocarcinomas as a model system for the study of embryogenesis and neoplasia. [Review] [77 refs]. *Cell* **5,** 229-243.

Martin, G.R. and Evans. M.J. (1975). The formation of embryoid bodies *in vitro* by homogeneous embryonal carcinoma cell cultures derived from isolated single cells. In Teratomas and Differentiation. M.I. Sherman and D. Solter, editors. Academic Press, New York, USA. 169-187.

Matsui, Y., Zsebo, K., and Hogan, B.L. (1992). Derivation of pluripotential embryonic stem cells from murine primordial germ cells in culture. *Cell* **70**, 841-847.

Matsunami, N., Hamaguchi, Y., Yamamoto, Y., Kuze, K., Kangawa, K., Matsuo, H., Kawaichi, M., and Honjo, T. (1989). A protein binding to the J kappa recombination sequence of immunoglobulin genes contains a sequence related to the integrase motif. *Nature* **342**, 934-937.

Maxwell, P.H., Wiesener, M.S., Chang, G.W., Clifford, S.C., Vaux, E.C., Cockman, M.E., Wykoff, C.C., Pugh, C.W., Maher, E.R., and Ratcliffe, P.J. (1999). The tumour suppressor protein VHL targets hypoxia-inducible factors for oxygen-dependent proteolysis. *Nature* **399**, 271-275.

Mayer, U., R. A. T. Ruiz, T. Berleth, S. Misera, and Jurgens. G. 1991. Mutations affecting body organization in the *Arabidopsis* embryo. *Nature* **353**, 402-407

Meerovitch, K., Svitkin, Y.V., Lee, H.S., Lejbkowicz, F., Kenan, D.J., Chan, E.K., Agol, V.I., Keene, J.D., and Sonenberg, N. (1993). La autoantigen enhances and corrects aberrant translation of poliovirus RNA in reticulocyte lysate. *J. Virol.* **67**, 3798-3807.

Meijer, D., Graus, A., Kraay, R., Langeveld, A., Mulder, M.P., and Grosveld, G. (1990). The octamer binding factor Oct6: cDNA cloning and expression in early embryonic cells. *Nucl. Acids. Res.* **18**, 7357-7365.

Meisler, M.H. (1992). Insertional mutation of 'classical' and novel genes in transgenic mice. [Review] [46 refs]. *Trends. Genetics* **8**, 341-344.

Mercer, J.A., Seperack, P.K., Strobel, M.C., Copeland, N.G., and Jenkins, N.A. (1991). Novel myosin heavy chain encoded by murine dilute coat colour locus [published erratum appears in Nature 1991 Aug 8;352(6335): 547]. *Nature* **349**, 709-713.

Michael, S.K., Brennan, J., and Robertson, E.J. (1999). Efficient gene-specific expression of cre recombinase in the mouse embryo by targeted insertion of a novel IRES-Cre cassette into endogenous loci [published erratum appears in Mech Dev 1999 86(1-2):213]. *Mech. Dev.* **85**, 35-47.

Mintz, B. and Illmensee, K. (1975). Normal genetically mosaic mice produced from malignant teratocarcinoma cells. *Proc. Natl. Acad. Sci. USA.* **72**, 3585-3589.

Mintz, B. and Russell. E.S. (1957). Gene induced embryological modifications of primordial germ cells. *J. Exp. Zool.* **134**, 207-239.

Molla, A., Jang, S.K., Paul, A.V., Reuer, Q., and Wimmer, E. (1992). Cardioviral internal ribosomal entry site is functional in a genetically engineered dicistronic poliovirus. *Nature* **356**, 255-257.

Monk, M., Boubelik, M., and Lehnert, S. (1987). Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. *Development* **99**, 371-382.

Monk, M. and Harper, M.I. (1979). Sequential X chromosome inactivation coupled with cellular differentiation in early mouse embryos. *Nature* **281**, 311-313.

Morrison, S.J., Shah, N.M., and Anderson, D.J. (1997). Regulatory mechanisms in stem cell biology. [Review] [112 refs]. *Cell* 88, 287-298.

Mountford, P., Nichols, J., Zevnik, B., O'Brien, C., and Smith, A. (1998). Maintenance of pluripotential embryonic stem cells by stem cell selection. *Reprod. Fertil. Dev.* 10, 527-533.

Mountford, P.S. and Smith, A.G. (1995). Internal ribosome entry sites and dicistronic RNAs in mammalian transgenesis. [Review] [31 refs]. *Trends. Genetics* 11, 179-184.

Mountford, P., Zevnik, B., Duwel, A., Nichols, J., Li, M., Dani, C, Robertson, M., Chambers, I., and Smith, A. (1994). Dicistronic targeting constructs: reporters and modifiers of mammalian gene expression. *Proc. Natl. Acad. Sci. USA.* 91, 4303-4307.

Mummery, C.L., van Achterberg, T.A., van den Eijnden-van Raaij, A.J., van Haaster, L., Willemse, A., de Laat, S.W., and Piersma, A.H. (1991). Visceral-endoderm-like cell lines induce differentiation of murine P19 embryonal carcinoma cells. *Differentiation* 46, 51-60.

Muth, K., Bruyns, R., Thorey, I.S., and von Melchner, H. (1998). Disruption of genes regulated during hematopoietic differentiation of mouse embryonic stem cells. *Dev. Dynamics* 212, 277-283.

Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W., and Roder, J.C. (1993). Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl. Acad. Sci. USA.* 90, 8424-8428.

Natarajan, D. and Boulter, C.A. (1995). A lacZ-hygromycin fusion gene and its use in a gene trap vector for marking embryonic stem cells. *Nucl. Acids. Res.* 23, 4003-4004.

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.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. *Cell* 95, 379-391.

Nichols, J., Chambers, I., and Smith, A. (1994). Derivation of germline competent embryonic stem cells with a combination of interleukin-6 and soluble interleukin-6 receptor. *Exp. Cell Res.* 215, 237-239.

Nichols, J., Evans, E.P., and Smith, A.G. (1990). Establishment of germ-line-competent embryonic stem (ES) cells using differentiation inhibiting activity. *Development* 110, 1341-1348.

Nichols, J. and Gardner, R.L. (1984). Heterogeneous differentiation of external cells in individual isolated early mouse inner cell masses in culture. *J. Embryol. Exp. Morph.* 80, 225-240.

Nicolet, G. (1971). The young notochord can induce somite genesis by means of diffusible substances in the chick. *Experientia* 27, 938-939.

Nieto, M.A., Gilardi-Hebenstreit, P., Charnay, P., and Wilkinson, D.G. (1992). A receptor protein tyrosine kinase implicated in the segmental patterning of the hindbrain and mesoderm. *Development* 116, 1137-1150.

Nishikawa, S.I., Nishikawa, S., Hirashima, M., Matsuyoshi, N., and Kodama, H. (1998). Progressive lineage analysis by cell sorting and culture identifies FLK1+VE-cadherin+ cells at a diverging point of endothelial and hemopoietic lineages. *Development* 125, 1747-1757.

Nishimoto, M., Fukushima, A., Okuda, A., and Muramatsu, M. (1999). The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. *Mol. Cell. Biol.* **19**, 5453-5465.

Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. *Nature Genet.* **24**, 372-376.

Niwa, H., Burdon, T., Chambers, I., and Smith, A. (1998). Self-renewal of pluripotent embryonic stem cells is mediated via activation of STAT3. *Genes. Dev.* **12**, 2048-2060.

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.

Noben-Trauth, N., Kohler, G., Burki, K., and Ledermann, B. (1996). Efficient targeting of the IL-4 gene in a BALB/c embryonic stem cell line. *Trans. Res.* **5**, 487-491.

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

O'Gorman, S., Fox, D.T., and Wahl, G.M. (1991). Recombinase-mediated gene activation and site-specific integration in mammalian cells. *Science* **251**, 1351-1355.

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

Ohta, T., Michel, J.J., Schottelius, A.J., and Xiong, Y. (1999). ROC1, a homolog of APC11, represents a family of cullin partners with an associated ubiquitin ligase activity. *Mol. Cell.* **3**, 535-541.

Oka, C., Nakano, T., Wakeham, A., de la Pompa, J.L., Mori, C., Sakai, T., Okazaki, S., Kawaichi, M., Shiota, K., Mak, T.W., and Honjo, T. (1995). Disruption of the mouse RBP-J kappa gene results in early embryonic death. *Development* **121**, 3291-3301.

Okabe, S., Forsberg-Nilsson, K., Spiro, A.C., Segal, M., and McKay, R.D. (1996). Development of neuronal precursor cells and functional postmitotic neurons from embryonic stem cells in vitro. *Mech. Dev.* **59**, 89-102.

Okamoto, K., Okazawa, H., Okuda, A., Sakai, M., Muramatsu, M., and Hamada, H. (1990). A novel octamer binding transcription factor is differentially expressed in mouse embryonic cells. *Cell* **60**, 461-472.

Orban, P.C., Chui, D., and Marth, J.D. (1992). Tissue- and site-specific DNA recombination in transgenic mice. *Proc. Natl. Acad. Sci. USA.* **89**, 6861-6865.

Pages, G., Lenormand, P., L'Allemain, G., Chambard, J.C., Meloche, S., and Pouyssegur, J. (1993). Mitogen-activated protein kinases p42mapk and p44mapk are required for fibroblast proliferation. *Proc. Natl. Acad. Sci. USA.* **90**, 8319-8323.

Palacios, R., Golunski, E., and Samaridis, J. (1995). In vitro generation of hematopoietic stem cells from an embryonic stem cell line. *Proc. Natl. Acad. Sci. USA.* **92**, 7530-7534.

Papaioannou, V.E., McBurney, M.W., Gardner, R.L., and Evans, M.J. (1975). Fate of teratocarcinoma cells injected into early mouse embryos. *Nature* **258**, 70-73.

Patton, E.E., Willems, A.R., and Tyers, M. (1998a). Combinatorial control in ubiquitin-dependent proteolysis: don't Skp the F-box hypothesis. [Review] [74 refs]. *Trends. Genetics* **14**, 236-243.

Patton, E.E., Willems, A.R., Sa, D., Kuras, L., Thomas, D., Craig, K.L., and Tyers, M. (1998b). Cdc53 is a scaffold protein for multiple Cdc34/Skp1/F-box protein complexes that regulate cell division and methionine biosynthesis in yeast [published erratum appears in Genes Dev 1998 Oct 1;12(19):3144]. *Genes. Dev.* 12, 692-705.

Pease, S., Braghetta, P., Gearing, D., Grail, D., and Williams, R.L. (1990). Isolation of embryonic stem (ES) cells in media supplemented with recombinant leukemia inhibitory factor (LIF). *Dev. Biol.* 141, 344-352.

Pedersen, R.A. (1986). Potency, lineage and allocation in preimplantation mouse embryos. In Experimental approaches to mammalian embryonic development. J. Rossant and Pederson R.A., editors. Cambridge University Press, New York, USA. 3-33.

Peifer, M. (1993). Cancer, catenins, and cuticle pattern: a complex connection. *Science* 262, 1667-1668.

Pelletier, J. and Sonenberg, N. (1988). Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature* 334, 320-325.

Pennica, D., Shaw, K.J., Swanson, T.A., Moore, M.W., Shelton, D.L., Zioncheck, K.A., Rosenthal, A., Taga, T., Paoni, N.F., and Wood, W.I. (1995). Cardiotrophin-1. Biological activities and binding to the leukemia inhibitory factor receptor/gp130 signaling complex. *J. Biol. Chem.* 270, 10915-10922.

Pera, M. F. (1999). Biology of human testicular germ cell tumours. *Repro. Med. Rev.* 7, 141-154.

Pera, M.F. and Herszfeld, D. (1998). Differentiation of human pluripotent teratocarcinoma stem cells induced by bone morphogenetic protein-2. *Reprod. Fertil. Dev.* 10, 551-555.

Pera, M.F., Roach, S., and Elliss, C.J. (1990). Comparative biology of mouse and human embryonal carcinoma. [Review] [76 refs]. *Cancer Surv.* 9, 243-262.

Pera, M.F., Cooper, S., Mills, J., and Parrington, J.M. (1989). Isolation and characterization of a multipotent clone of human embryonal carcinoma cells. *Differentiation* 42, 10-23.

Pesce, M., Wang, X., Wolgemuth, D.J., and Scholer, H. (1998). Differential expression of the Oct-4 transcription factor during mouse germ cell differentiation. *Mech. Dev.* 71, 89-98.

Placzek, M., Tessier-Lavigne, M., Yamada, T., Jessell, T., Dodd, and J. (1990). Mesodermal control of neural cell identity: floor plate induction by the notochord. *Science* 250, 985-988.

Potten, C.S. and Loeffler, M. (1990). Stem cells: attributes, cycles, spirals, pitfalls and uncertainties. Lessons for and from the crypt. [Review] [83 refs]. *Development* 110, 1001-1020.

Potter, H., Weir, L., and Leder, P. (1984). Enhancer-dependent expression of human kappa immunoglobulin genes introduced into mouse pre-B lymphocytes by electroporation. *Proc. Natl. Acad. Sci. USA.* 81, 7161-7165.

Qu, C.K. and Feng, G.S. (1998). Shp-2 has a positive regulatory role in ES cell differentiation and proliferation. *Oncogene* 17, 433-439.

Qu, C.K., Shi, Z.Q., Shen, R., Tsai, F.Y., Orkin, S.H., and Feng, G.S. (1997). A deletion mutation in the SH2-N domain of Shp-2 severely suppresses hematopoietic cell development. *Mol. Cell. Biol.* 17, 5499-5507.

Quinn, P., Barros, C., and Whittingham, D.G. (1982). Preservation of hamster oocytes to assay the fertilizing capacity of human spermatozoa. *J. Reprod. Fertil.* 66, 161-168.

Ramirez-Solis, R., Liu, P., and Bradley, A. (1995). Chromosome engineering in mice. *Nature* 378, 720-724.

Rappolee, D.A., Sturm, K.S., Behrendtsen, O., Schultz, G.A., Pedersen, R.A., and Werb, Z. (1992). Insulin-like growth factor II acts through an endogenous growth pathway regulated by imprinting in early mouse embryos. *Genes. Dev.* 6, 939-952.

Rappolee, D.A., Brenner, C.A., Schultz, R., Mark, D., and Werb, Z. (1988). Developmental expression of PDGF, TGF-alpha, and TGF-beta genes in preimplantation mouse embryos. *Science* 241, 1823-1825.

Rathjen, J., Lake, J.A., Bettess, M.D., Washington, J.M., Chapman, G, and Rathjen, P.D. (1999). Formation of a primitive ectoderm like cell population, EPL cells, from ES cells in response to biologically derived factors. *J. Cell. Sci.* 112, 601-612.

Reddy, S., DeGregori, J.V., von Melchner, H., and Ruley, H.E. (1991). Retrovirus promoter-trap vector to induce lacZ gene fusions in mammalian cells. *J. Virol.* 65, 1507-1515.

Reeve, W.J. and Ziomek, C.A. (1981). Distribution of microvilli on dissociated blastomeres from mouse embryos: evidence for surface polarization at compaction. *J. Embryol. Exp. Morph.* 62, 339-350.

Regan-Reimann, J.D., Duong, Q.V., and Jackson, P.K. (1999). Identification of novel F-box proteins in Xenopus laevis. *Curr. Biol.* 9, R762-3.

Regenstreif, L.J. and Rossant, J. (1989). Expression of the c-fms proto-oncogene and of the cytokine, CSF-1, during mouse embryogenesis. *Dev. Biol.* 133, 284-294.

Resnick, J.L., Bixler, L.S., Cheng, L., and Donovan, P.J. (1992). Long-term proliferation of mouse primordial germ cells in culture. *Nature* 359, 550-551.

Reubinoff, B.E., Pera, M.F., Fong, C.Y., Trounson, A., and Bongso, A. (2000). Embryonic stem cell lines from human blastocysts: somatic differentiation *in vitro*. *Nature Biotech.* 18, 399-404.

Riley, P., Anson-Cartwright, L., and Cross, J.C. (1998). The Hand1 bHLH transcription factor is essential for placentation and cardiac morphogenesis. *Nature Genet.* 18, 271-275.

Rinchik, E.M. (1991). Chemical mutagenesis and fine-structure functional analysis of the mouse genome. [Review] [45 refs]. *Trends. Genetics* 7, 15-21.

Rinchik, E.M., Bangham, J.W., Hunsicker, P.R., Cacheiro, N.L., Kwon, B.S., Jackson, I.J., and Russell, L.B. (1990). Genetic and molecular analysis of chlorambucil-induced germ-line mutations in the mouse. *Proc. Natl. Acad. Sci. USA.* 87, 1416-1420.

Rinchik, E.M. and Russell. L.B. (1990). Germ-line deletion mutations in the mouse: Tools for intensive functional and physical mapping of regions of the mammalian genome. in:Genome analysis: Genetic and physical mapping. K. Davies and Tilghman S, editors. CSH Laboratory Press, Cold Spring Harbor, NY, USA. 121-158.

Roach, M.L., Stock, J.L., Byrum, R., Koller, B.H., and McNeish, J.D. (1995). A new embryonic stem cell line from DBA/1lacJ mice allows genetic modification in a murine model of human inflammation. *Exp. Cell Res.* 221, 520-525.

Roach, S., Cooper, S., Bennett, W., and Pera, M.F. (1993). Cultured cell lines from human teratomas: windows into tumour growth and differentiation and early human development. [Review] [23 refs]. *European Urology* 23, 82-87; discussion 87-88.

Robertson, E.J. (1987). Embryo-derived stem cell lines. In:Teratocarcinoma and embryonic stem cells. E.J. Robertson, editor. IRL Press, Oxford, UK. 71-112.

Robertson, E., Bradley, A., Kuehn, M., and Evans, M. (1986). Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. *Nature* 323, 445-448.

Rohdewohld, H., Weiher, H., Reik, W., Jaenisch, R., and Breindl, M. (1987). Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. *J. Virol.* 61, 336-343.

Rohrer, G.A., Alexander, L.J., Hu, Z., Smith, T.P., Keele, J.W., and Beattie, C.W. (1996). A comprehensive map of the porcine genome. [Review] [112 refs]. *Genome. Res.* 6, 371-391.

Rohwedel, J., Maltsev, V., Bober, E., Arnold, H.H., Hescheler, J., and Wobus, A.M. (1994). Muscle cell differentiation of embryonic stem cells reflects myogenesis *in vivo*: developmentally regulated expression of myogenic determination genes and functional expression of ionic currents. *Dev. Biol.* 164, 87-101.

Rose, T.M., Weiford, D.M., Gunderson, N.L., and Bruce, A.G. (1994). Oncostatin M (OSM) inhibits the differentiation of pluripotent embryonic stem cells *in vitro*. *Cytokine* 6, 48-54.

Rosenquist, T.A. and Martin, G.R. (1995). Visceral endoderm-1 (VE-1): an antigen marker that distinguishes anterior from posterior embryonic visceral endoderm in the early post-implantation mouse embryo. *Mech. Dev.* 49, 117-121.

Rosner, M.H., Vigano, M.A., Ozato, K., Timmons, P.M., Poirier, F., Rigby, P.W., and Staudt, L.M. (1990). A POU-domain transcription factor in early stem cells and germ cells of the mammalian embryo. *Nature* 345, 686-692.

Rossant, J. (1995). Development of the extraembryonic lineages. *Semin. Dev. Biol.* 6, 237-247.

Rossant, J. and Hopkins, N. (1992). Of fin and fur: mutational analysis of vertebrate embryonic development. [Review] [108 refs]. *Genes. Dev.* 6, 1-13.

Rossant, J. (1986). Development of extra-embryonic cell lineages. In Experimental approaches to mammalian development. J. Rossant and R.A. Pederson, editors. Cambridge University Press, New York, USA. 97-119.

Rossant, J., Gardner, R.L., and Alexandre, H.L. (1978). Investigation of the potency of cells from the postimplantation mouse embryo by blastocyst injection: a preliminary report. *J. Embryol. Exp. Morph.* 48, 239-247.

Rothstein, J.L., Johnson, D., DeLoia, J.A., Skowronski, J., Solter, D., and Knowles, B. (1992). Gene expression during preimplantation mouse development. *Genes. Dev.* 6, 1190-1201.

Rouillon, A., Barbey, R., Patton, E.E., Tyers, M., and Thomas, D. (2000). Feedback-regulated degradation of the transcriptional activator Met4 is triggered by the SCF (Met30) complex. *EMBO. J.* 19, 282-294.

Ruegger, M., Dewey, E., Gray, W.M., Hobbie, L., Turner, J., and Estelle, M. (1998). The TIR1 protein of Arabidopsis functions in auxin response and is related to human SKP2 and yeast grr1p. *Genes. Dev.* 12, 198-207.

Russell, L.B., Hunsicker, P.R., Cacheiro, N.L., Bangham, J.W., Russell, W.L., and Shelby, M.D. (1989). Chlorambucil effectively induces deletion mutations in mouse germ cells. *Proc. Natl. Acad. Sci. USA.* 86, 3704-3708.

Russell, L.D., R.A. Ettlin, A.P. SinhaHikim, and E.D. Clegg. (1990). Histological and histopathological evaluation of the testis. Cache River Press, Clearwater, Fl, USA.

Russell, W.L., Kelly, E.M., Hunsicker, P.R., Bangham, J.W., Maddux, S.C., and Phipps, E.L. (1979). Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. *Proc. Natl. Acad. Sci. USA.* **76**, 5818-5819.

Russell, W. L. (1951). X-ray induced mutations in mice. *Cold Spring Harb. Symp. Quant. Biol.* **16**, 327-336.

Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B., and Erlich, H.A. (1988). Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**, 487-491.

Sakai, T., Furukawa, T., Iwanari, H., Oka, C., Nakano, T., Kawaichi, M., and Honjo, T. (1995). Loss of immunostaining of the RBP-J kappa transcription factor upon F9 cell differentiation induced by retinoic acid. *J. Biochem.* **118**, 621-628.

Salminen, M., Meyer, B.I., and Gruss, P. (1998). Efficient poly A trap approach allows the capture of genes specifically active in differentiated embryonic stem cells and in mouse embryos. *Dev. Dynamics* **212**, 326-333.

Sam, M., Wurst, W., Kluppel, M., Jin, O., Heng, H., and Bernstein, A. (1998). Aquarius, a novel gene isolated by gene trapping with an RNA-dependent RNA polymerase motif. *Dev. Dynamics* **212**, 304-317.

Sambrook, J., E.F. Fritsch, and Maniatis T. (1989). Molecular Cloning. CSH Laboratory Press, Cold Spring Harbor, NY, U.S.A.

Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA.* **74**, 5463-5467.

Sasaki, H. and Hogan, B.L. (1994). HNF-3 beta as a regulator of floor plate development. *Cell* **76**, 103-115.

Sasaki, H. and Hogan, B.L. (1993). Differential expression of multiple fork head related genes during gastrulation and axial pattern formation in the mouse embryo. *Development* **118**, 47-59.

Sasaki, N., Nagaoka, S., Itoh, M., Izawa, M., Konno, H., Carninci, P., Yoshiki, A., Kusakabe, M., Moriuchi, T., Muramatsu, M., Okazaki, Y., and Hayashizaki, Y. (1998). Characterization of gene expression in mouse blastocyst using single-pass sequencing of 3995 clones [published erratum appears in Genomics 1998 Dec 15;54(3):583]. *Genomics* **49**, 167-179.

Savatier, P., Lapillonne, H., van Grunsven, L.A., Rudkin, B.B., and Samarut, J. (1996). Withdrawal of differentiation inhibitory activity/leukemia inhibitory factor up-regulates D-type cyclins and cyclin-dependent kinase inhibitors in mouse embryonic stem cells. *Oncogene* **12**, 309-322.

Savatier, P., Huang, S., Szekely, L., Wiman, K.G., and Samarut, J. (1994). Contrasting patterns of retinoblastoma protein expression in mouse embryonic stem cells and embryonic fibroblasts. *Oncogene* **9**, 809-818.

Scherdin, U., Rhodes, K., and Breindl, M. (1990). Transcriptionally active genome regions are preferred targets for retrovirus integration. *J. Virol.* **64**, 907-912.

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

Schnieke, A., Harbers, K., and Jaenisch, R. (1983). Embryonic lethal mutation in mice induced by retrovirus insertion into the alpha 1(I) collagen gene. *Nature* 304, 315-320.

Schöler, H.R. (1991). Octamania: the POU factors in murine development. [Review] [42 refs]. *Trends. Genetics* 7, 323-329.

Schöler, H.R., Ciesiolka, T., and Gruss, P. (1991). A nexus between Oct-4 and E1A: implications for gene regulation in embryonic stem cells. *Cell* 66, 291-304.

Schöler, H.R., Ruppert, S., Suzuki, N., Chowdhury, K., and Gruss, P. (1990). New type of POU domain in germ line-specific protein Oct-4. *Nature* 344, 435-439.

Schöler, H.R., Balling, R., Hatzopoulos, A.K., Suzuki, N., and Gruss, P. (1989a). Octamer binding proteins confer transcriptional activity in early mouse embryogenesis. *EMBO. J.* 8, 2551-2557.

Schöler, H.R., Hatzopoulos, A.K., Balling, R., Suzuki, N., and Gruss, P. (1989b). A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. *EMBO. J.* 8, 2543-2550.

Schuldiner, M., Yanuka, O., Itskovitz-Eldor, J., Melton, D.A., and Benvenisty, N. (2000). Effects of eight growth factors on the differentiation of cells derived from human embryonic stem cells. *Proc. Natl. Acad. Sci. USA.* 97(21), 11307-11312.

Schuler, G.D., Altschul, S.F., and Lipman, D.J. (1991). A workbench for multiple alignment construction and analysis. *Proteins* 9, 180-190.

Schultz, G.A. (1986). Molecular aspects of mammalian oocyte growth and maturation.
In: Experimental approaches to mammalian development. J. Rossant and Pederson. R.A, editors. Cambridge University Press, New York, USA. 195-237.

Schuster-Gossler, K., Bilinski, P., Sado, T., Ferguson-Smith, A., and Gossler, A. (1998). The mouse Gtl2 gene is differentially expressed during embryonic development, encodes multiple alternatively spliced transcripts, and may act as an RNA. *Dev. Dynamics* 212, 214-228.

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

Scott, I.C., Anson-Cartwright, L., Riley, P., Reda, D., and Cross, J.C. (2000). The HAND1 basic helix-loop-helix transcription factor regulates trophoblast differentiation via multiple mechanisms. *Mol. Cell. Biol.* 20, 530-541.

Searle, A. G. (1974). Mutation induction in mice. *Adv. Rad. Biol.* 4, 131-207.

Seol, J.H., Feldman, R.M., Zachariae, W., Shevchenko, A., Correll, C.C., Lyapina, S., Chi, Y., Galova, M., Claypool, J., Sandmeyer, S., Nasmyth, K., and Deshaies, R.J. (1999). Cdc53/cullin and the essential Hrt1 RING-H2 subunit of SCF define a ubiquitin ligase module that activates the E2 enzyme Cdc34. *Genes. Dev.* 13, 1614-1626.

Serafini, T., Colamarino, S.A., Leonardo, E.D., Wang, H., Beddington, R., Skarnes, W.C., and Tessier-Lavigne, M. (1996). Netrin-1 is required for commissural axon guidance in the developing vertebrate nervous system. *Cell* 87, 1001-1014.

Shamblott, M.J., Axelman, J., Littlefield, J.W., Blumenthal, P.D., Huggins, G.R., Cui, Y., Cheng, L., and Gearhart, J.D. (2001). Human embryonic germ cell derivatives express a broad range of developmentally distinct markers and proliferate extensively *in vitro. Proc. Natl. Acad. Sci. USA.* 98(1), 113-118.

Shamblott, M.J., Axelman, J., Wang, S., Bugg, E.M., Littlefield, J.W., Donovan, P.J., Blumenthal, P.D., Huggins, G.R., and Gearhart, J.D. (1998). Derivation of pluripotent stem cells from cultured human primordial germ cells [published erratum appears in Proc Natl Acad Sci U S A 1999 96(3):1162]. *Proc. Natl. Acad. Sci. USA.* **95**, 13726-13731.

Shawlot, W., Wakamiya, M., Kwan, K.M., Kania, A., Jessell, T.M., and Behringer, R.R. (1999). Lim1 is required in both primitive streak-derived tissues and visceral endoderm for head formation in the mouse. *Development* **126**, 4925-4932.

Shawlot, W. and Behringer, R.R. (1995). Requirement for Lim1 in head-organizer function. *Nature* **374**, 425-430.

Shedlovsky, A., McDonald, J.D., Symula, D., and Dove, W.F. (1993). Mouse models of human phenylketonuria. *Genetics* **134**, 1205-1210.

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. Natl. Acad. Sci. USA.* **85**, 180-184.

Shen, M.M., Wang, H., and Leder, P. (1997). A differential display strategy identifies Cryptic, a novel EGF-related gene expressed in the axial and lateral mesoderm during mouse gastrulation. *Development* **124**, 429-442.

Sheng, Z., Knowlton, K., Chen, J., Hoshijima, M., Brown, J.H., and Chien, K.R. (1997). Cardiotrophin 1 (CT-1) inhibition of cardiac myocyte apoptosis via a mitogen-activated protein kinase-dependent pathway. Divergence from downstream CT-1 signals for myocardial cell hypertrophy. *J. Biol. Chem.* **272**, 5783-5791.

Shih, C.C., Stoye, J.P., and Coffin, J.M. (1988). Highly preferred targets for retrovirus integration. *Cell* **53**, 531-537.

Shirai, M., Miyashita, A., Ishii, N., Itoh, Y., Satokata, I., Watanabe, Y.G., and Kuwano, R. (1996). A gene trap strategy for identifying the gene expressed in the embryonic nervous system. *Zool. Sci.* **13**, 277-283.

Singh, G., Supp, D.M., Schreiner, C., McNeish, J., Merker, H.J., Copeland, N.G., Jenkins, N.A., Potter, S.S., and Scott, W. (1991). legless insertional mutation: morphological, molecular, and genetic characterization. *Genes. Dev.* **5**, 2245-2255.

Sirard, C., de la Pompa, J.L., Elia, A., Itie, A., Mirtsos, C., Cheung, A., Hahn, S., Wakeham, A., Schwartz, L., Kern, S.E., Rossant, J., and Mak, T.W. (1998). The tumor suppressor gene Smad4/Dpc4 is required for gastrulation and later for anterior development of the mouse embryo. *Genes. Dev.* **12**, 107-119.

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

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. (1990). Entrapment vectors: a new tool for mammalian genetics. [Review] [30 refs]. *Bio/Technology* **8**, 827-831.

Skowyra, D., Koepp, D.M., Kamura, T., Conrad, M.N., Conaway, R.C., Conaway, J.W., Elledge, S.J., and Harper, J.W. (1999). Reconstitution of G1 cyclin ubiquitination with complexes containing SCFGrr1 and Rbx1. *Science* **284**, 662-665.

Skowyra, D., Craig, K.L., Tyers, M., Elledge, S.J., and Harper, J.W. (1997). F-box proteins are receptors that recruit phosphorylated substrates to the SCF ubiquitin-ligase complex. *Cell* **91**, 209-219.

Smith, A.G. (1992). Mouse embryo stem cells: their identification, propagation and manipulation. [Review] [118 refs]. *Semin. Cell. Biol.* **3**, 385-399.

Smith, A.G., Nichols, J., Robertson, M., and Rathjen, P.D. (1992). Differentiation inhibiting activity (DIA/LIF) and mouse development. [Review] [87 refs]. *Dev. Biol.* **151**, 339-351.

Smith, A. G. (1991). Culture and differentiation of embryonic stem cells. *J. Tiss. Cult. Meth.* **13**, 89-94.

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 differentiation by purified polypeptides. *Nature* **336**, 688-690.

Smith, A.G. ard Hooper, M.L. (1987). Buffalo rat liver cells produce a diffusible activity which inhibits the differentiation of murine embryonal carcinoma and embryonic stem cells. *Dev. Biol.* **121**, 1-9.

Snow, M.H. and Tam, P.P. (1979). Is compensatory growth a complicating factor in mouse teratology? *Nature* **279**, 555-557.

Snow, M. H. (1977). Gastrulation in the mouse: growth and regionalization of the epiblast. *J. Embryol. Exp. Morph.* **42**, 293-303.

Snow, M.H., Aitken, J., and Ansell, J.D. (1976). Role of the inner cell mass in controlling implantation in the mouse. *J. Reprod. Fertil.* **48**, 403-404.

Soares, M.B., Bonaldo, M.F., Jelene, P., Su, L., Lawton, L., and Efstratiadis, A. (1994). Construction and characterization of a normalized cDNA library. *Proc. Natl. Acad. Sci. USA.* **91**, 9228-9232.

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.

Solter, D., Skreb, N., and Damjanov, I. (1970). Extrauterine growth of mouse egg-cylinders results in malignant teratoma. *Nature* **227**, 503-504.

Soria, B., Roche, E., Berna, G., Leon-Quinto, T., Reig, J.A., and Martin, F. (2000). Insulin-secreting cells derived from embryonic stem cells normalize glycemia in streptozotocin-induced diabetic mice. *Diabetes* **49**, 157-162.

Soriano, P., Gridley, T., and Jaenisch, R. (1987). Retroviruses and insertional mutagenesis in mice: proviral integration at the Mov 34 locus leads to early embryonic death. *Genes. Dev.* **1**, 366-375.

Southern, E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* **98**, 503-517.

Spyropoulos, D.D. and Capecchi, M.R. (1994). Targeted disruption of the even-skipped gene, evx1, causes early postimplantation lethality of the mouse conceptus. *Genes. Dev.* **8**, 1949-1961.

Stahl, N., Farruggella, T.J., Boulton, T.G., Zhong, Z., Darnell, J.E., Jr., and Yancopoulos, G.D. (1995). Choice of STATs and other substrates specified by modular tyrosine-based motifs in cytokine receptors. *Science* **267**, 1349-1353.

Stebbins, C.E., Kaelin, W.G., Jr., and Pavletich, N.P. (1999). Structure of the VHL-ElonginC-ElonginB complex: implications for VHL tumor suppressor function. *Science* **284**, 455-461.

Stevens, L.C. (1983). The origin and development of testicular, ovarian and embryo-derived teratomas. In Teratocarcinoma stem cells. Silver L.M., Martin G.R., and Strickland S, editors. CSH Press, Cold Spring Harbor, USA. 23-36.

Stevens, L.C. and Varnum, D.S. (1974). The development of teratomas from parthenogenetically activated ovarian mouse eggs. *Dev. Biol.* **37**, 369-380.

Stevens, L.C. (1970). The development of transplantable teratocarcinomas from intratesticular grafts of pre- and postimplantation mouse embryos. *Dev. Biol.* **21**, 364-382.

Stevens, L. C. and Little C.C. (1954). Spontaneous testicular teratomas in an inbred strain of mice. *Proc. Natl. Acad. Sci. USA.* **40**, 1080-1087.

Stewart, C.L., Gadi, I., and Bhatt, H. (1994). Stem cells from primordial germ cells can reenter the germ line. *Dev. Biol.* **161**, 626-628.

Stewart, C.L. (1982). Formation of viable chimaeras by aggregation between teratocarcinomas and preimplantation mouse embryos. *J. Embryol. Exp. Morph.* **67**, 167-179.

Stewart, T.A. and Mintz, B. (1982). Recurrent germ-line transmission of the teratocarcinoma genome from the METT-1 culture line to progeny *in vivo. J. Exptl. Zool.* **224**, 465-469.

Stoye, J.P., Fenner, S., Greenoak, G.E., Moran, C., and Coffin, J.M. (1988). Role of endogenous retroviruses as mutagens: the hairless mutation of mice. *Cell* **54**, 383-391.

Stoykova, A., Chowdhury, K., Bonaldo, P., Torres, M., and Gruss, P. (1998). Gene trap expression and mutational analysis for genes involved in the development of the mammalian nervous system. *Dev. Dynamics* **212**, 198-213.

Strickland, S. and Richards, W.G. (1992). Invasion of the trophoblasts. [Review] [28 refs]. *Cell* **71**, 355-357.

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

Surani, M.A., Barton, S.C., and Norris, M.L. (1984). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* **308**, 548-550.

Surani, M.A. and Barton, S.C. (1977). Trophoblastic vesicles of preimplantation blastocysts can enter into quiescence in the absence of inner cell mass. *J. Embryol. Exp. Morph.* **39**, 273-277.

Sutterluty, H., Chatelain, E., Marti, A., Wirbelauer, C., Senften, M., Muller, U., and Krek, W. (1999). p45SKP2 promotes p27Kip1 degradation and induces S phase in quiescent cells. *Nature Cell Biol.* **1**, 207-214.

Sutton, J., Costa, R., Klug, M., Field, L., Xu, D., Largaespada, DA, Fletcher, C.F., Jenkins, N.A., Copeland, N.G., Klemsz, M., and Hromas, R. (1996). Genesis, a winged helix transcriptional repressor with expression restricted to embryonic stem cells. *J. Biol. Chem.* **271**, 23126-23133.

Suzuki, Y., Yoshitomo-Nakagawa, K., Maruyama, K., Suyama, A., and Sugano, S. (1997). Construction and characterization of a full length-enriched and a 5'-end-enriched cDNA library. *Gene* **200**, 149-156.

Sweeney, C., Murphy, M., Kubelka, M., Ravnik, S.E., Hawkins, C.F., Wolgemuth, D.J., and Carrington, M. (1996). A distinct cyclin A is expressed in germ cells in the mouse. *Development* **122**, 53-64.

Szyf, M., Tanigawa, G., and McCarthy, P.L., Jr. (1990). A DNA signal from the Thy-1 gene defines de novo methylation patterns in embryonic stem cells. *Mol. Cell. Biol.* **10**, 4396-4400.

Tada, M., Tada, T., Lefebvre, L., Barton, S.C., and Surani, M.A. (1997). Embryonic germ cells induce epigenetic reprogramming of somatic nucleus in hybrid cells. *EMBO. J.* **16**, 6510-6520.

Tada, T., Tada, M., Hilton, K., Barton, S.C., Sado, T., Takagi, N., and Surani, M.A. (1998). Epigenotype switching of imprintable loci in embryonic germ cells. *Dev. Genes. Evol.* **207**, 551-561.

Takeda, H., Yamakuchi, H., Ihara, N., Hara, K., Watanabe, T., Sugimoto, Y., Oshiro, T., Kishine, H., Kano, Y., and Kohno, K. (1998). Construction of a bovine yeast artificial chromosome (YAC) library. *Animal. Genet.* **29**, 216-219.

Takeuchi, T. (1997). A gene trap approach to identify genes that control development. [Review] [31 refs]. *Dev. Growth & Differ.* **39**, 127-134.

Takeuchi, T., Yamazaki, Y., Katoh-Fukui, 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.

Tam, P.P. and Behringer, R.R. (1997). Mouse gastrulation: the formation of a mammalian body plan. [Review] [252 refs]. *Mech. Dev.* **68**, 3-25.

Tam, P.P. and Zhou, S.X. (1996). The allocation of epiblast cells to ectodermal and germ-line lineages is influenced by the position of the cells in the gastrulating mouse embryo. *Dev. Biol.* **178**, 124-132.

Tam, P.P. and Beddington, R.S. (1992). Establishment and organization of germ layers in the gastrulating mouse embryo. [Review] [33 refs]. *Ciba Foundation Symposium* **165**, 27-41; discussion 42-49.

Tam, P.P. and Meier, S. (1982). The establishment of a somitomeric pattern in the mesoderm of the gastrulating mouse embryo. *Am. J. Anat.* **164**, 209-225.

Tam, P.P. and Snow, M.H. (1981). Proliferation and migration of primordial germ cells during compensatory growth in mouse embryos. *J. Embryol. Exp. Morph.* **64**, 133-147.

Tanaka, S., Kunath, T., Hadjantonakis, A.K., Nagy, A., and Rossant, J. (1998). Promotion of trophoblast stem cell proliferation by FGF4. *Science* **282**, 2072-2075.

Tao, W. and Lai, E. (1992). Telencephalon-restricted expression of BF-1, a new member of the HNF-3/fork head gene family, in the developing rat brain. *Neuron* **8**, 957-966.

Thomas, K.R. and Capecchi, M.R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51**, 503-512.

Thomas, P.Q., Brown, A., and Beddington, R.S. (1998). Hex: a homeobox gene revealing peri-implantation asymmetry in the mouse embryo and an early transient marker of endothelial cell precursors. *Development* 125, 85-94.

Thomas, P. Q., J. Brickman, H. Popperl, R. Krumlauf, and Beddington. R. S. P. (1997). Axis duplication and anterior identity in the mouse embryo. *Cold Spring Harb. Symp. Quant. Biol.* 62.:115-125.

Thomas, P. and Beddington, R. (1996). Anterior primitive endoderm may be responsible for patterning the anterior neural plate in the mouse embryo. *Curr. Biol.* 6, 1487-1496.

Thompson, S., Clarke, A.R., Pow, A.M., Hooper, M.L., and Melton, D.W. (1989). Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. *Cell* 56, 313-321.

Thomson, J.A. and Odorico, J.S. (2000). Human embryonic stem cell and embryonic germ cell lines. *Trends. Biotech.* 18, 53-57.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts [see comments] [published erratum appears in Science 1998 Dec 4; 282(5395):1827]. *Science* 282, 1145-1147.

Thomson, J.A., Kalishman, J., Golos, T.G., Durning, M., Harris, C.P., and Hearn, J.P. (1996). Pluripotent cell lines derived from common marmoset (Callithrix jacchus) blastocysts. *Biol. Reprod.* 55, 254-259.

Thomson, J.A., Kalishman, J., Golos, T.G., Durning, M., Harris, C.P., Becker, R.A., and Hearn, J.P. (1995). Isolation of a primate embryonic stem cell line. *Proc. Natl. Acad. Sci. USA.* 92, 7844-7848.

Tilghman, S.M. (1999). The sins of the fathers and mothers: genomic imprinting in mammalian development. [Review] [68 refs]. *Cell* 96, 185-193.

Torres, M., Stoykova, A., Huber, O., Chowdhury, K., Bonaldo, P., Mansouri, A., Butz, S., Kemler, R., and Gruss, P. (1997). An alpha-E-catenin gene trap mutation defines its function in preimplantation development. *Proc. Natl. Acad. Sci. 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. Res.* 7, 293-298.

Tsvetkov, L.M., Yeh, K.H., Lee, S.J., Sun, H., and Zhang, H. (1999). p27(Kip1) ubiquitination and degradation is regulated by the SCF(Skp2) complex through phosphorylated Thr187 in p27. *Curr. Biol.* 9, 661-664.

Tyers, M. and Jorgensen, P. (2000). Proteolysis and the cell cycle: with this RING I do thee destroy. [Review] [88 refs]. *Curr. Opin. Genet. Devel.* 10, 54-64.

Valancius, V. and Smithies, O. (1991). Testing an "in-out" targeting procedure for making subtle genomic modifications in mouse embryonic stem cells. *Mol. Cell. Biol.* 11, 1402-1408.

Van Deursen, J., Fornerod, M., Van Rees, B., and Grosveld, G. (1995). Cre-mediated site-specific translocation between nonhomologous mouse chromosomes. *Proc. Natl. Acad. Sci. USA.* 92, 7376-7380.

Varlet, I., Collignon, J., and Robertson, E.J. (1997). nodal expression in the primitive endoderm is required for specification of the anterior axis during mouse gastrulation. *Development* 124, 1033-1044.

Vestweber, D., Gossler, A., Boller, K., and Kemler, R. (1987). Expression and distribution of cell adhesion molecule uvomorulin in mouse preimplantation embryos. *Dev. Biol.* 124, 451-456.

Vijaya, S., Steffen, D.L., and Robinson, H.L. (1986). Acceptor sites for retroviral integrations map near DNase I-hypersensitive sites in chromatin. *J. Virol.* 60, 683-692.

Vitaterna, M.H., King, D.P., Chang, A.M., Kornhauser, J.M., Lowrey, P.L., McDonald, J.D., Dove, W.F., Pinto, L.H., Turek, F.W., and Takahashi, J.S. (1994). Mutagenesis and mapping of a mouse gene, Clock, essential for circadian behavior. *Science* 264, 719-725.

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. *Genes. Dev.* 6, 919-927.

von Melchner, H. and Ruley, H.E. (1989). Identification of cellular promoters by using a retrovirus promoter trap. *J. Virol.* 63, 3227-3233.

Voss, A.K., Thomas, T., Petrou, P., Anastassiadis, K., Scholer, H., and Gruss, P. (2000). Taube nuss is a novel gene essential for the survival of pluripotent cells of early mouse embryos. *Development* 127(24), 5449-5461.

Voss, A.K., Thomas, T., and Gruss, P. (1998a). Efficiency assessment of the gene trap approach. *Dev. Dynamics* 212, 171-180.

Voss, A.K., Thomas, T., and Gruss, P. (1998b). Compensation for a gene trap mutation in the murine microtubule-associated protein 4 locus by alternative polyadenylation and alternative splicing. *Dev. Dynamics* 212, 258-266.

Wagner, E.F., Stewart, T.A., and Mintz, B. (1981). The human beta-globin gene and a functional viral thymidine kinase gene in developing mice. *Proc. Natl. Acad. Sci. USA.* 78, 5016-5020.

Wagner, T.E., Hoppe, P.C., Jollick, J.D., Scholl, D.R., Hodinka, R.L., and Gault, J.B. (1981). Microinjection of a rabbit beta-globin gene into zygotes and its subsequent expression in adult mice and their offspring. *Proc. Natl. Acad. Sci. USA.* 78, 6376-6380.

Wakayama, T., Perry, A.C., Zuccotti, M., Johnson, K.R., and Yanagimachi, R. (1998). Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. *Nature* 394, 369-374.

Waldrip, W.R., Bikoff, E.K., Hoodless, P.A., Wrana, J.L., Robertson, and E.J. (1998). Smad2 signaling in extraembryonic tissues determines anterior-posterior polarity of the early mouse embryo. *Cell* 92, 797-808.

Waltzer, L., Logeat, F., Brou, C., Israel, A., Sergeant, A., and Manet, E. (1994). The human J kappa recombination signal sequence binding protein (RBP-J kappa) targets the Epstein-Barr virus EBNA2 protein to its DNA responsive elements. *EMBO. J.* 13, 5633-5638.

Wassarman, P.M., Bleil, J.D., Florman, H.M., Greve, J.M., Roller, R.J., and Salzmann, G.S. (1986). Nature of the mouse egg's receptor for sperm. [Review] [73 refs]. *Adv. Exp. Med. Biol.* 207, 55-77.

Wassarman, P.M., J.M. Greve, R.M. Perona, R.J. Roller, and Salzmann G.S. (1984). How mouse cells put cn and take off their extracellular coat. In Molecular biology of development. E. Davidson and R. Firtel, editors. AR Liss, New York, USA. 213-225.

Watt, F.M. and Hogan, B.L. (2000). Out of Eden: stem cells and their niches. [Review] [34 refs]. Science 287, 1427-1430.

Weber, F., de Villiers, J., and Schaffner, W. (1984). An SV40 "enhancer trap" incorporates exogenous enhancers or generates enhancers from its own sequences. Cell 36, 983-992.

Weiher, H., Noda, T., Gray, D.A., Sharpe, A.H., and Jaenisch, R. (1990). Transgenic mouse model of kidney disease: insertional inactivation of ubiquitously expressed gene leads to nephrotic syndrome. Cell 62, 425-434.

Weinstein, D.C., Ruiz, I., Altaba A, Chen, W.S., Hoodless, P., Prezioso, V.R., Jessell, T.M., and Darnell, J.E., Jr. (1994). The winged-helix transcription factor HNF-3 beta is required for notochord development in the mouse embryo. Cell 78, 575-588.

Weissman, I.L. (2000). Stem cells: units of development, units of regeneration, and units in evolution. [Review] [110 refs]. Cell 100, 157-168.

Wenk, J., Andrews, P.W., Casper, J., Hata, J., Pera, M.F., von Keitz, A., Damjanov, I., and Fenderson, B.A. (1994). Glycolipids of germ cell tumors: extended globo-series glycolipids are a hallmark of human embryonal carcinoma cells. Int. J. Cancer 58, 108-115.

Whittingham, D.G. (1971). Culture of mouse ova. J. Reprod. Fertil. -Supp. 14, 7-21.

Wiley, L.M., Wu, J.X., Harari, I., and Adamson, E.D. (1992). Epidermal growth factor receptor mRNA and protein increase after the four-cell preimplantation stage in murine development. Dev. Biol. 149, 247-260.

Wiley, L.M., Spindle, A.I., and Pedersen, R.A. (1978). Morphology of isolated mouse inner cell masses developing in vitro. Dev. Biol. 63, 1-10.

Wilkinson, D.G., Bhatt, S., and Herrmann, B.G. (1990). Expression pattern of the mouse T gene and its role in mesoderm formation. Nature 343, 657-659.

Willems, A.R., Lanker, S., Patton, E.E., Craig, K.L., Nason, T.F., Mathias, N., Kobayashi, R., Wittenberg, C., and Tyers, M. (1996). Cdc53 targets phosphorylated G1 cyclins for degradation by the ubiquitin proteolytic pathway. Cell 86, 453-463.

Williams, D.E., Eisenman, J., Baird, A., Rauch, C., Van Ness, K., March, C.J., Park, L.S., Martin, U., Mochizuki, D.Y., Boswell, H.S., and et al. (1990). Identification of a ligand for the c-kit proto-oncogene. Cell 63, 167-174.

Williams, R.L., Hilton, D.J., Pease, S., Willson, 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.

Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J., and Campbell, K.H. (1997). Viable offspring derived from fetal and adult mammalian cells [see comments] [published erratum appears in Nature 1997 Mar 13; 386(6621):200]. Nature 385, 810-813.

Wilson, C., Bellen, H.J., and Gehring, W.J. (1990). Position effects on eukaryotic gene expression. [Review] [170 refs]. Annu. Rev. Cell. Biol. 6, 679-714.

Wilson, C., Pearson, R.K., Bellen, H.J., O'Kane, C.J., Grossniklaus, U., and Gehring, W.J. (1989). P-element-mediated enhancer detection: an efficient method for isolating and characterizing developmentally regulated genes in Drosophila. *Genes. Dev.* 3, 1301-1313.

Winnier, G., Blessing, M., Labosky, P.A., and Hogan, B.L. (1995). Bone morphogenetic protein-4 is required for mesoderm formation and patterning in the mouse. *Genes. Dev.* 9, 2105-2116.

Winston, J.T., Koepp, D.M., Zhu, C., Elledge, S.J., and Harper, J.W. (1999a). A family of mammalian F-box proteins. *Curr. Biol.* 9, 1180-1182.

Winston, J.T., Strack, P., Beer-Romero, P., Chu, C.Y., Elledge, S.J., and Harper, J.W. (1999b). The SCFbeta-TRCP-ubiquitin ligase complex associates specifically with phosphorylated destruction motifs in IkappaBalpha and beta-catenin and stimulates IkappaBalpha ubiquitination *in vitro* [published erratum appears in Genes Dev 1999 Apr 15;13(8):1050]. *Genes. Dev.* 13, 270-283.

Withers-Ward, E.S., Kitamura, Y., Barnes, J.P., and Coffin, J.M. (1994). Distribution of targets for avian retrovirus DNA integration *in vivo. Genes. Dev.* 8, 1473-1487.

Wobus, A. M., J. Rohwedel, V. Maltsev, and Hescheler J. (1994). *In vitro* differentiation of embryonic stem cells into cardiomyocytes or skeletal muscle cells is specifically modulated by retinoic acid. *Roux's Arch. Dev. Biol.* 204, 36-45

Wolf, E., Kramer, R., Polejaeva, I., Thoenen, H., and Brem, G. (1994). Efficient generation of chimaeric mice using embryonic stem cells after long-term culture in the presence of ciliary neurotrophic factor. *Trans. Res.* 3, 152-158.

Wolpert, L., R. Beddington, J. Brockes, T. Jessell, P. Lawrence, and Meyerowitz. E. (1998). Principles of development. Lawrence. E., editor. Oxford University Press, Oxford, UK.

Woychik, R.P., Maas, R.L., Zeller, R., Vogt, T.F., and Leder, P. (1990). 'Formins': proteins deduced from the alternative transcripts of the limb deformity gene. *Nature* 346, 850-853.

Woychik, R.P., Stewart, T.A., Davis, L.G., D'Eustachio, P., and Leder, P. (1985). An inherited limb deformity created by insertional mutagenesis in a transgenic mouse. *Nature* 318, 36-40.

Wu, H., Liu, X., and Jaenisch, R. (1994). Double replacement: strategy for efficient introduction of subtle mutations into the murine Col1a-1 gene by homologous recombination in embryonic stem cells. *Proc. Natl. Acad. Sci. USA.* 91, 2819-2823.

Wurst, W., Rossant, J., Prideaux, V., Kownacka, M., Joyner, A., Hill, D.P., Guillemot, F., Gasca, S., Cado, D., Auerbach, A., and et al. (1995). A large-scale gene-trap screen for insertional mutations in developmentally regulated genes in mice. *Genetics* 139, 889-899.

Xiong, J.W., Battaglino, R., Leahy, A., and Stuhlmann, H. (1998). Large-scale screening for developmental genes in embryonic stem cells and embryoid bodies using retroviral entrapment vectors. *Dev. Dynamics* 212, 181-197.

Xu, C., Inokuma, M.S., Denham, J., Golds, K., Kundu, P., Gold, J.D., Carpenter, M.K. (2001). Growth of undifferentiated human embryonic stem cells on defined matrices with conditioned medium. [Abstract]. *Proceedings of Keystone symposium, Colorado, USA.* Pluripotent stem cells: Biology and applications. Abstract #133.

Yamashita, J., Itoh, H., Hirashima, M., Ogawa, M., Nishikawa, S., Yurugi, T., Naito, M., and Nakao, K. (2000). Flk1-positive cells derived from embryonic stem cells serve as vascular progenitors. *Nature* 408(6808), 92-96.

Yang, R., Morosetti, R., and Koeffler, H.P. (1997). Characterization of a second human cyclin A that is highly expressed in testis and in several leukemic cell lines. *Cancer. Res.* **57**, 913-920.

Yaron, A., Hatzubai, A., Davis, M., Lavon, I., Amit, S., Manning, A.M., Andersen, J.S., Mann, M., Mercurio, F., and Ben-Neriah, Y. (1998). Identification of the receptor component of the IkappaBalpha-ubiquitin ligase. *Nature* **396**, 590-594.

Yenofsky, R.L., Fine, M., and Pellow, J.W. (1990). A mutant neomycin phosphotransferase II gene reduces the resistance of transformants to antibiotic selection pressure. *Proc. Natl. Acad. Sci. USA.* **87**, 3435-3439.

Yeo, T.T., Yang, T., Massa, S.M., Zhang, J.S., Honkaniemi, J., Butcher, L.L., and Longo, F.M. (1997). Deficient LAR expression decreases basal forebrain cholinergic neuronal size and hippocampal cholinergic innervation. *J. Neuroscience Research* **47**, 348-360.

Yeom, Y.I., Fuhrmann, G., Ovitt, C.E., Brehm, A., Ohbo, K., Gross, M, Hubner, K., and Scholer, H.R. (1996). Germline regulatory element of Oct-4 specific for the totipotent cycle of embryonal cells. *Development* **122**, 881-894.

Yoshida, K., Chambers, I., Nichols, J., Smith, A., Saito, M., Yasukawa, K., Shoyab, M., Taga, T., and Kishimoto, T. (1994). Maintenance of the pluripotential phenotype of embryonic stem cells through direct activation of gp130 signalling pathways. *Mech. Dev.* **45**, 163-171.

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.

Yoshinaga, K. and Adams, C.E. (1966). Delayed implantation in the spayed, progesterone treated adult mouse. *J. Reprod. Fertil.* **12**, 593-595.

Yuan, H., Corbi, N., Basilico, C., and Dailey, L. (1995). Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes. Dev.* **9**, 2635-2645.

Zachariae, W. and Nasmyth, K. (1999). Whose end is destruction: cell division and the anaphase-promoting complex. [Review] [165 refs]. *Genes. Dev.* **13**, 2039-2058.

Zambrowicz, B.P., Friedrich, G.A., Buxton, E.C., Lilleberg, S.L., Person, C., and Sands, A.T. (1998). Disruption and sequence identification of 2,000 genes in mouse embryonic stem cells. *Nature* **392**, 608-611.

Zernicka-Goetz, M., Pines, J., McLean Hunter, S., Dixon, J.P., Siemering, K.R., Haseloff, J., and Evans, M.J. (1997). Following cell fate in the living mouse embryo. *Development* **124**, 1133-1137.

Zhang, H., Kobayashi, R., Galaktionov, K., and Beach, D. (1995). p19Skp1 and p45Skp2 are essential elements of the cyclin A-CDK2 S phase kinase. *Cell* **82**, 915-925.

Zhang, J.G., Farley, A., Nicholson, S.E., Willson, T.A., Zugaro, L.M., Simpson, R.J., Moritz, R.L., Cary, D., Richardson, R., Hausmann, G., Kile, B.J., Kent, S.B., Alexander, W.S., Metcalf, D., Hilton, D.J., Nicola, N.A., and Baca, M. (1999). The conserved SOCS box motif in suppressors of cytokine signaling binds to elongins B and C and may couple bound proteins to proteasomal degradation. *Proc. Natl. Acad. Sci. USA.* **96**, 2071-2076.

Zheng, X.H. and Hughes, S.H. (1999). An avian sarcoma/leukosis virus-based gene trap vector for mammalian cells. *J. Virol.* **73**, 6946-6952.

Zhou, P. and Howley, P.M. (1998). Ubiquitination and degradation of the substrate recognition subunits of SCF ubiquitin-protein ligases. *Mol. Cell.* **2**, 571-580.

Zhou, X., Sasaki, H., Lowe, L., Hogan, B.L., and Kuehn, M.R. (1993). Nodal is a novel TGF-beta-like gene expressed in the mouse node during gastrulation. *Nature* **361**, 543-547.

Zimber-Strobl, U., Strobl, L.J., Meitinger, C., Hinrichs, R., Sakai, T., Furukawa, T., Honjo, T., and Bornkamm, G.W. (1994). Epstein-Barr virus nuclear antigen 2 exerts its transactivating function through interaction with recombination signal binding protein RBP-J kappa, the homologue of Drosophila Suppressor of Hairless. *EMBO. J.* **13**, 4973-4982.

Zsebo, K.M., Wypych, J., McNiece, I.K., Lu, H.S., Smith, K.A., Karkare, S.B., Sachdev, R.K., Yuschenkoff, V.N., Birkett, N.C., Williams, L.R. (1990). Identification, purification, and biological characterization of hematopoietic stem cell factor from buffalo rat liver-conditioned medium. *Cell* **63**, 195-201.