THE UNIVERSITY *of* EDINBURGH

GENE TARGETING IN MOUSE EMBRYONAL STEM CELLS

Thesis Submitted for the Degree of Doctor of Philosophy by Stephen C. Pells

University of Edinburgh

1995

DECLARATION

The work described in this thesis has been conducted by myself unless specifically stated otherwise in the text.

S.C.Pells, April 1995

Acknowledgements

This project was a Postgraduate Scholarship in the Faculty of Medicine of the University of Edinburgh, and funded in part by The Melville Trust for the Care and Cure of Cancer, and I am grateful to both these organisations for their financial support.

I was supervised in this project by Dr Martin Hooper and Dr Alan Clarke, to whom I am indebted for their instruction, discussion and encouragement throughout my time in Edinburgh.

The work was conducted in the Department of Pathology in the Faculty of Medicine and I would like to express my thanks to Professor Colin Bird. I would also like to thank Mr Robert Morris and Mrs Audrey Peter for technical advice throughout my time here, and in the case of Audrey, with instruction in the delicate art of caring for ES cells. I very much appreciate Steven Haig's work on the photographs used as figures in this thesis.

The vectors pKiTPNT and pKiD0PNT were made by Dr D. James Williamson, and I would like to thank him for initiating the work on the K-*ras* project. Lorraine Dobbie provided assistance with the injection of blastocysts, and I would also like to thank Dr Charles Patek for his selfless attitude with his reagents when I ran out of mine!

I would also like to thank, for both their friendship and (sometimes!) scientific discussion, all those who have made my time in Edinburgh such an experience-Kate, Lucy, Damian, Sula, Niall, Sandrine, Ann, Jane, Dino and of course, Jen and all of my family.

UNIVERSITY OF EDINBURGH

ABSTRACT OF THESIS (Regulation 3.5.10)

Name of Candidate Stephen Christopher PELLS

Address

Degree PhD *Date* 28.3.95

Title of Thesis Gene targeting in mouse embryonal stem cells

No. of words in the main text of Thesis 40,000

Gene targeting is a method for introducing mutations into specific genes of an organism. To date most work in the field has focused on the creation of insertion or deletion events to generate a null allele in the gene of interest in murine embryonic stem (ES) cells, with the objective of introducing this mutation into the mouse germline. This has facilitated the study of the *in vivo* role of many genes at the level of the organism and provided a ready source of cell lines carrying the desired mutation for further analysis. The generation of more subtle mutations, where function of the gene is modified rather than ablated, is also desirable. This thesis is concerned with "Hit and Run" gene targeting in ES cells, a method designed with the objective of introducing small mutations into the genome. The procedure relies on the integration of a vector carrying the desired mutation, and positive and negative selectable markers. After isolation of an event which has introduced the vector into the target locus by positive selection and genomic screening, a second selection performed upon the negative selectable marker is carried out. This should select for clones which have undergone an intra-chromosomal recombination event which removes the vector sequences and restores the locus. Some of the reverted clones obtained will have retained the mutation.

The hit and run targeting technique relies on a high degree of fidelity in the homologous recombination process, to avoid the introduction of mutations other than the desired mutation into the target locus. However, relatively little is known about the fidelity of gene targeting homologous recombination events. Part of this thesis examines this question in a model gene, the $hprt^{b\text{-}m3}$ null mutant locus. This allele is useful for studying gene targeting reactions because selections either for or against gene activity may be simply achieved by *in vitro* chemical selection. The $hprt^{b\text{-}m3}$ gene was targeted with an insertion-type vector designed to restore *hprt* gene function which also incorporated the positive selectable marker *neo*. The *neo* gene was initially used to select stably-transfected clones. The clones were then studied genotypically and phenotypically to seek events where apparently correct homologous recombination events at the *hprt* locus failed to restore gene function and thus suggested errors in the targeting process. The data indicate that the rate of error in recombination must be low.

Mutations in the kirsten-*ras* proto-oncogene are recognised to be of great significance in neoplasia, but relatively little is known about its function in normal mammalian development. Several features of the gene make it suitable for study by means of hit and run gene targeting: it has an untranslated 5' exon of undefined function, may be activated by point mutations, and contains alternative C-terminal exons. Hit and run experiments designed to modify these features of the gene have been carried out. A hit and run vector designed to delete a 2.5kb region surrounding the promoter/exon 0 region targeted the locus at a low but observable frequency. However, closer examination of the targeted locus revealed a more complicated structure than that predicted, and reversion events under back-selection were found to produce either a variant of the structure predicted for an insertion event or a wild-type structure. It is suggested that a combination of concatenation and extra-chromosomal recombination events resulted in a complex insertion structure which was capable of undergoing multiple excision events.

Hit and run vectors were also used with the objective of introducing point mutations into exons 1 and 4B of the K-*ras* gene. After an unmeasurably low targeting frequency was observed for the first of these vectors, a "Fast-Track" protocol was attempted with both of these vectors which was designed to expand the effective number of potentially targeted clones backselected. This method was shown to enrich slightly for events at the target locus in the case of the construct used to delete exon 0, but was unsuccessful in generating the designed target locus in these experiments. RT-PCR of the modified gene was found to be a good method of screening targeted clones for subtle modifications not detectable by Southern analysis.

As a complement to the targeting studies of splicing variants in the K-*ras* gene, splicing variant usage in the tissues of the developing mouse was also analysed by RT-PCR using primers specific for both variants and comparing quantities of each message. Consistent with other work in the field, this analysis demonstrates that K-*ras* is ubiquitously expressed in the developing mouse. However, previous work has failed to differentiate between splice variants of K-*ras* mRNA. Here it is shown that whilst K-*ras*B has a ubiquitous pattern of expression, the pattern of K-*ras*A expression appears to be more limited, and K-*ras*A is usually expressed at a lower level than K-*ras*B. Little or no temporal variation of relative splice variant usage within tissue types was observed, suggesting that the definition of K-*ras* expression patterns is an early event in organogenesis. Analysis of differentiating ES and EC cells suggests that upregulation of K-*ras*A expression is associated with tissue differentiation.

CONTENTS

CHAPTER 1-
INTRODUCTION

Introduction

1.1 Mammalian Genetics

Much of eucaryotic genetic analysis has been carried out in organisms such as yeast, *Drosophila melanogaster* and *Caenorhabditis elegans* (see, for example, Alberts *et al.*, 1994). These species are of small size and typically have short generation times which permit the relatively simple creation and analysis of a wide variety of mutations. As a model system for the genetic analysis of mammals however, such species are somewhat limited. The size and generation times of mammals, together with the large genome size (approximately $3x10^9$ bp/haploid genome), makes such studies more difficult. In the mouse over 1300 loci are known with naturally-occurring mutations, many of which are for genes which have not as yet been cloned (Green, 1989). This compares poorly with an estimated complement of 10^4-10^5 genes in the mammalian genome (Alberts *et al.*, 1994). Thus many mammalian genes, notably some involved in immune system function or implicated in the development of malignancy, have no known naturally-occurring mutants.

Several thousand genetic diseases are known in humans (McKusick, 1978), most of which have no direct parallel in an experimental animal which may be used as a model in the study of the disease and as a test-bed for potential therapies.

1.1.1 Transgenic Mammals

Transgenic mammals, most usually mouse, but also other species both laboratory and domestic, have been created by the micro-injection of DNA into oocytes or retroviral insertion (reviewed in Palmiter & Brinster, 1986; Jaenisch, 1988) and have proved very helpful in the study of many genes. This technique however is limited by the fact that genetic material may only be added to the genome, not subtracted, at genetic loci

and in copy numbers which may not be predicted. Phenotypes obtained by such procedures may be affected by both the position and the number of copies of the injected sequence which integrate into the genome (reviewed by Palmiter & Brinster, 1986). Local position effects at the site of integration or the lack of introns or enhancer elements within the transgene can cause attenuation of expression to below physiological levels (Jaenisch, 1988). Conversely, integration of multiple tandemly-arrayed copies (a common event) or the presence of cryptic promoter sequences in the transgene can cause unexpectedly high expression in a tissue (e.g. Al-Shawi *et al.*, 1991). Should the transgene integrate into an endogenous gene the results of the experiment may be complicated by additional effects due to loss of activity of the integration locus.

It is often beneficial to generate a null allele for a gene of interest. However, the use of transgenic expression of antisense sequences with the objective of effectively deleting a gene, although successful in some cases where a complete ablation of function is unnecessary to generate a phenotype (Katsuki *et al.*, 1988), is of variable and unpredictable efficiency (Katsuki *et al.*, 1988; Munir *et al.*, 1990).

Reverse genetics of mammals, most notably in the mouse, has been revolutionised in the past decade by the combination of two systems. These are the Embryonal/Embryonic Stem Cell (ES Cell) system, and the Gene Targeting system.

1.2 Embryonal Stem Cells

1.2.1 Early Murine Development (Hogan *et al.*, 1986)

After fertilisation, the egg divides during its journey down the oviduct to produce a morula (ball of cells) within the zona pellucida. As it reaches the 8- to 16-cell stage the cells (a.k.a. blastomeres) flatten and become polarised (compaction). Further division produces a blastocyst, which consists of a layer of cells (the trophoectoderm)

derived from the apical domains of the polarised cells surrounding a fluid-filled cavity (the blastocoel) and the cells of the inner cell mass (ICM), which are derived from the basal domains of the polarised cells. The biastocyst expands and by the 64-cell stage, when the ICM consists of approximately 20 cells, it hatches from the zona pellucida and implants into the uterine epithelium, 4-4.5 days post-coitum (p.c.).

Just before implantation the primitive endoderm, a layer of epithelial cells, differentiates from and surrounds the ICM. These cells go on to form the yolk sac of the embryo. The core of the ICM forms a layer of cells called the epiblast and this forms the ectoderm, endoderm and mesoderm. The ICM therefore gives rise to the entire foetus.

1.2.2 The Germline

The primordial germ cells (PGCs) originate from extra-embryonic tissue (Ginsburg *et al.* 1990) and migrate from the extraembryonic mesoderm to the gonads by day 12.5-13.5 p.c. (Ginsburg *et al.*, 1990; Fox *et al.*, 1981; Stinnakre *et al.*, 1981). They generate the germ cells in the adult embryo. However, they may only be cultured *in vitro* for a limited period before reaching mitotic arrest (Matsui *et al.*, 1991) and appear to be incapable of contributing to tissues in a chimaera following direct explantation from an embryo (Stewart *et al.*, 1994). Embryonic Germline (EG) cells are derived from PGCs by inducing them to divide in culture for long periods in the presence of multiple growth factors (Matsui *et al.*, 1992; Resnick *et al.*, 1992), and have recently been shown to be capable of contributing to the somatic tissues and germline of chimaeras (Stewart *et al.*, 1994; Matsui *et al.*, 1992).

1.2.3 Teratomas: Tumours of the Germline

Teratomas and teratocarcinomas are tumours which occur most commonly in the gonads, but they are also observed at other sites. They are defined by a confused

appearance due to the presence of multiple differentiated cell types usually derived from all three germ layers, which differentiate from a malignant developmentally pluripotent cell population: the embryonal carcinoma (EC) cells (Solter & Damjanov, 1979; Martin, 1980; Mintz & Fleischman, 1981; Silver *et al.*, 1983; Gardner, 1983). If all the EC cells differentiate the result is a benign teratoma; any malignancy is due entirely to the remaining EC cells. In mouse strain 129 spontaneous testicular teratomas are observed frequently (Stevens, 1983) and they can also be induced in a histocompatible recipient by grafting of the male genital ridges from a 129-strain embryo. Such transplanted tumours are therefore derived from PGCs (see previous section). Grafts from early stage embryos of many mouse strains can also produce teratomas in the recipient, but these are derived from the epiblast. This has been interpreted as meaning that EC cells behave malignantly by virtue of their local environment (at least initially), rather than being triggered to do so by an oncogenic stimulus at the genetic level. The fact that EC cells may originate either from the epiblast or the PGCs (reviewed in Hooper, 1992) suggests a close relationship between all of these cell types.

Chimaeras may be made with EC cells by either injecting a blastocyst with EC cells or aggregating EC cells with morulae (Bradley, 1987). They have a variety of possible fates: (1) failure to integrate; (2) colonisation of the extra-embryonic tissues; (3) proliferation resulting in the death of the host embryo; (4) integration into the embryo to produce tumours; (5) integration into the embryo and contributing to the somatic tissues or (6) integration and contribution to the functional germline cells of the embryo. Depending on the EC cell line used to make the chimaera, the relative frequency of each of the above possibilities can vary (reviewed in Papaioannou & Rossant, 1983). Importantly however, possibility (6), that is colonisation of the germline by EC cells in chimaeras, is very rare (Papaioannou & Rossant, 1983; Stewart & Mintz, 1981, 1982). This is likely to be due to the fact that selection either

in the primary tumour or as the cell undergoes adaptation to culture conditions results in a specific, germline-incompetent, subpopulation outgrowing the truly pluripotent EC cells. Despite their relatively homogenous appearance, different EC lines vary in their required culture conditions and the extent to which they will differentiate. Few have a normal diploid karyotype. (Table 1, Appendix of Silver *et al.*, 1983).

1.2.4 Embryonal Stem Cells

The apparent similarity between pluripotent EC cells and the cells of the epiblast and PGCs suggested that cell lines derived directly from the ICM might also be expected to be pluripotent. Embryonal (or embryonic) stem cells (ES cells) were first isolated from blastocysts by isolation and disruption of the ICM followed by subculture of the cells on feeder cell layers (Evans and Kaufman, 1981) or in EC cell-conditioned medium (Martin, 1981) to prevent differentiation. ES cells are pluripotent, and like EC cells may be induced to differentiate into a variety of different lineages *in vitro* (reviewed by Hooper, 1992). Injection of ES cells into blastocysts creates chimaeric progeny which may include ES cell-derived tissues in all three germ layers including the germline and the extraembryonic membranes (Bradley *et al.* 1984; Beddington & Robertson, 1989). Levels of chimaerism can be extremely high with ES cell-derived chimaeras. Because ES cells also contribute to the germline of a chimaera, it is possible to generate mice derived from ES cells which have undergone manipulation *in vitro* (Gossler *et al.*, 1986; Robertson *et al.*, 1986).

Most ES cell lines described have a 40, XY karyotype (Robertson *et al.*, 1983) and are therefore male as opposed to the majority of EC cell lines which are XO. This is of benefit because male animals have a higher reproductive potential than females and so ES cell-derived germ cells present at a low frequency will be easier to detect.

Pluripotency is obviously a defining characteristic of ES cells, and it is important that it be preserved while the cells are manipulated *in vitro*. Differentiation was originally prevented in ES (and EC) cells by co-culturing the ES/EC cells on layers of mitotically inactivated mouse embryo fibroblasts (Martin & Evans, 1975). ES cells may also be cultured on permanent fibroblast lines such as STO fibroblasts. Subsequently medium conditioned by the Buffalo Rat Liver (BRL) epithelial cell line (Smith & Hooper, 1987) or human bladder carcinoma 5637 cells (Williams *et al.*, 1988) was shown to include an efficient inhibitor of differentiation which permitted the continuous culture of ES cells in the absence of feeders. Isolation of the anti-differentiating activity showed that it is identical to the leukaemia inhibitory factor, LIF (Williams *et al.*, 1988; Smith *et al.*, 1988; Moreau *et al.*, 1988), so-called because *in vitro* it inhibits the M1 myeloid leukaemia cell line. More recently, ciliary neurotrophic factor (CNTF) has also been shown to maintain the pluripotency of ES cells *in vitro* (Ip *et al.*, 1992; Wolf *et al.*, 1994), as has human oncostatin M (OSM) (Gearing & Bruce, 1992).

The ability of the ES cells to contribute to the germline of the developing mouse thus provides a means of creating a mouse strain carrying a desired modification if such a change can be introduced into the ES cell.

1.3 Homologous Recombination in Mammalian Cells

Homologous recombination (HR) is recombination between DNA molecules dependent on homology between the recombining sequences and not on the specific sequence *per se*. In addition to procaryotic organisms, it also occurs in eucaryotic organisms during meiosis, mitosis and DNA repair (reviewed in Bollag *et al.*, 1989; Hooper, 1992). Evolutionarily HR functions to generate genetic variation and to rescue genetic information which may be lost in DNA damaging events by repairing a DNA molecule using information from the sister DNA molecule. Mammalian cells

may also undergo other types of recombination such as nonhomologous recombination and specific recombination types such as occurs in developing lymphocytes during immunoglobulin gene rearrangement.

HR occurs in mammalian cells between introduced free DNA molecules (Extra-Chromosomal Recombination, ECR), between homologous sequences in the chromosomes (Intra-Chromosomal Recombination [ICR] between homologous regions on a chromosome, and inter-chromosomal recombination, between homologous sequences on sister chromatids) and between an introduced sequence and a chromosomal sequence (gene targeting, considered in more depth in the next section) (reviewed in Bollag *et al.*, 1989; Hooper, 1992).

Whilst there are variations which attempt to account for certain specific observations, the models of recombination all invoke a free end of a single-strand of DNA which invades into a homologous double-stranded DNA molecule to produce a heteroduplex structure (so called because it involves a single strand from two different DNA molecules and thus mismatches in base-pairing are possible). The pairing of the strands releases a single strand of uncut sequence from the recipient molecule in a structure called a D-loop in response to repair synthesis at the end of the invading strand forcing it from the recipient duplex. The D-loop may then hybridise with the remaining strand of the donor molecule. This forms a strand exchange structure called a Holliday Junction (Holliday, 1964). The two major models suggest either a single-stranded nick (Meselson & Radding, 1975) or a double-stranded break (Orr-Weaver *et al.*, 1981) in the initiating strand. Depending on how the Holliday junctions are resolved (see figures 1-1 and 1-2), crossover, gene conversion or a combination of possibilities may be the result. Schematic diagrams of the Meselson-Radding and double-strand break-repair models of recombination are shown in figures 1-1 and 1-2.

Figure 1-1: Meselson-Radding Model of Homologous Recombination

Each line represents a single strand of DNA with the 3' end marked by an arrow. Newly-synthesised sequence is drawn as a fine line. Recombination is initiated by a single-stranded nick in one of the recombining molecules. The 3' end of the nick acts as a primer for repair synthesis and the newly-synthesised strand displaces a single-stranded D-loop from the other homologue as it invades (1). The D-loop is degraded, and by a combination of this degradation of the recipient and synthesis on the donor an asymmetric heteroduplex (a heteroduplex on only one of the homologues) forms (2). Ligation produces a Holliday Junction which may move along the paired molecules by branch migration (3). As shown the Holliday Junction has two crossed and two uncrossed strands. If the crossed strands are cut (4), the structure resolves to give two double stranded molecules, each consisting of a strand which is identical to the parent and one of which has undergone a gene conversion event without a crossover. The size of the region having undergone gene conversion differs between the molecules (it is smaller in the "lower/grey" case) because the newly-synthesised sequence was copied from the donor homologue. The Holliday Junction may alternatively isomerise, so that the crossed strands become uncrossed, and vice-versa (5). When this structure resolves, it produces crossover products with a segment that has undergone gene conversion (6).

(Adapted with permission from Hooper, 1992).

Figure 1-1

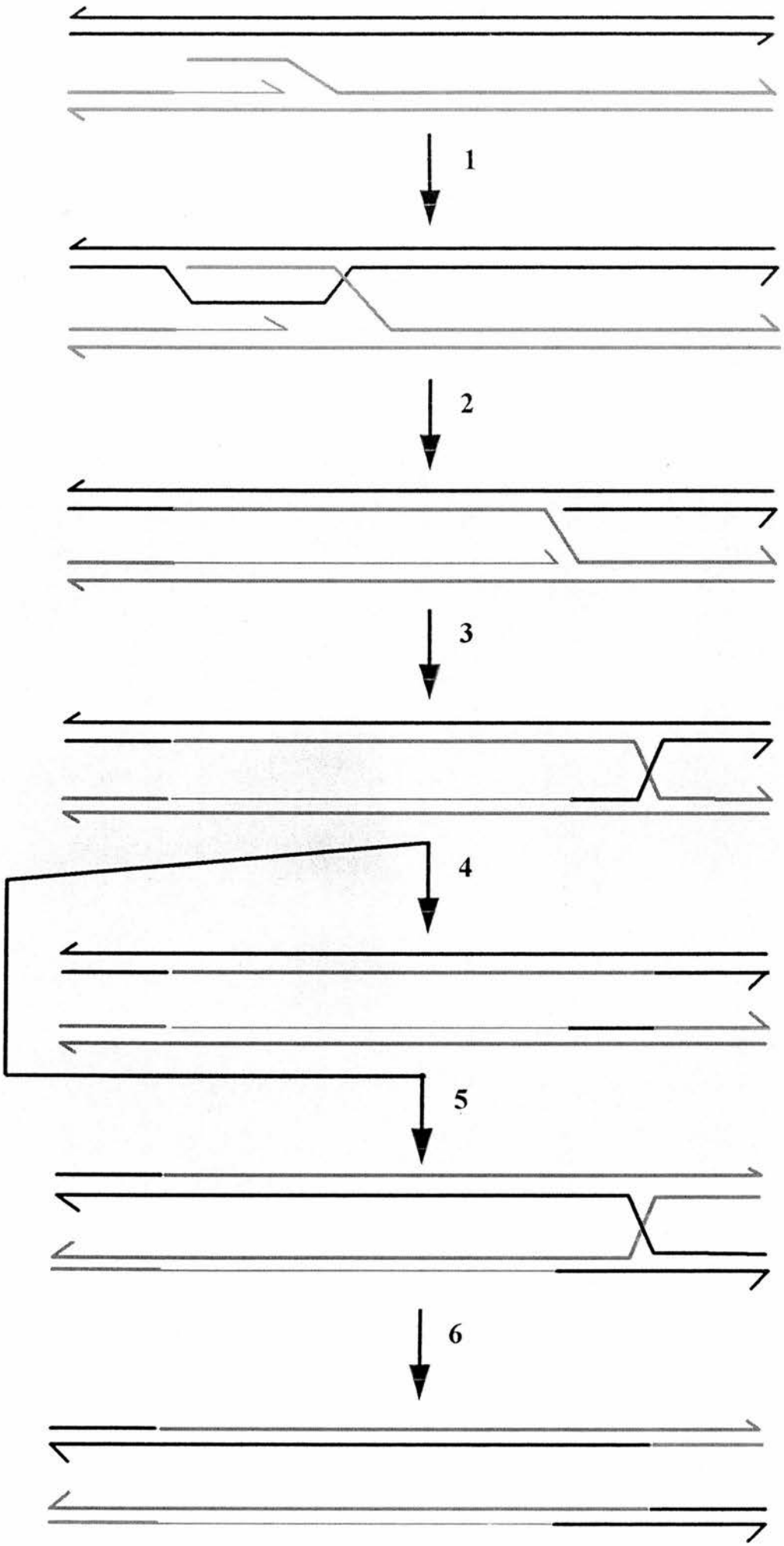
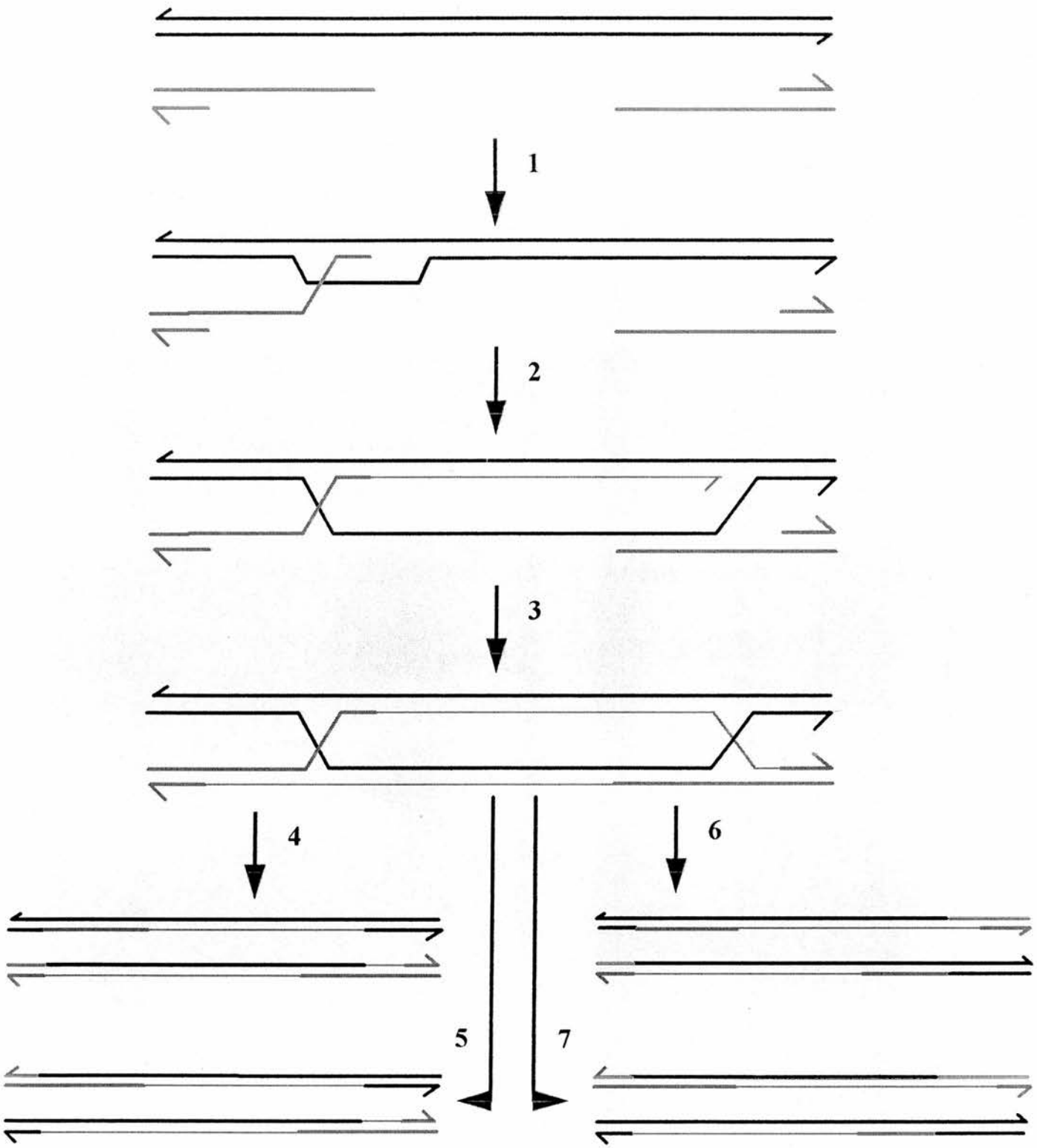


Figure 1-2: Double-Strand Break-Repair Model of Homologous Recombination

Recombination is initiated by a double-stranded break in one of the homologues, which is followed by exonuclease digestion to leave a gap flanked on either side by 3' overhanging ends. A 3' end can invade the other duplex, displacing a D-loop (1). As repair synthesis enlarges the D-loop, this free strand may anneal to the free 3' end at the other end of the gap (2). This strand then acts as a template for repair synthesis, thus repairing the gap (3). Ligation results in the formation of two Holliday Junctions, both of which may branch migrate. Depending on whether neither (4), the left (5), the right (6), or both (7) Holliday Junctions undergo isomerisation prior to resolution, there are several possibilities for the final structure. Resolutions 5 and 6 lead to a crossover event, whereas 4 and 7 do not.

(Adapted with permission from Hooper, 1992).

Figure 1-2



The outcome (reviewed in Bollag *et al.*, 1989; Hooper, 1992) of a recombination event may be either a reciprocal exchange between the DNA molecules (crossover), a non-reciprocal exchange (gene conversion), a gene conversion with an associated crossover or a non-conservative recombination event. These possibilities are shown in figure 1-3.

1.3.1 Extrachromosomal Recombination

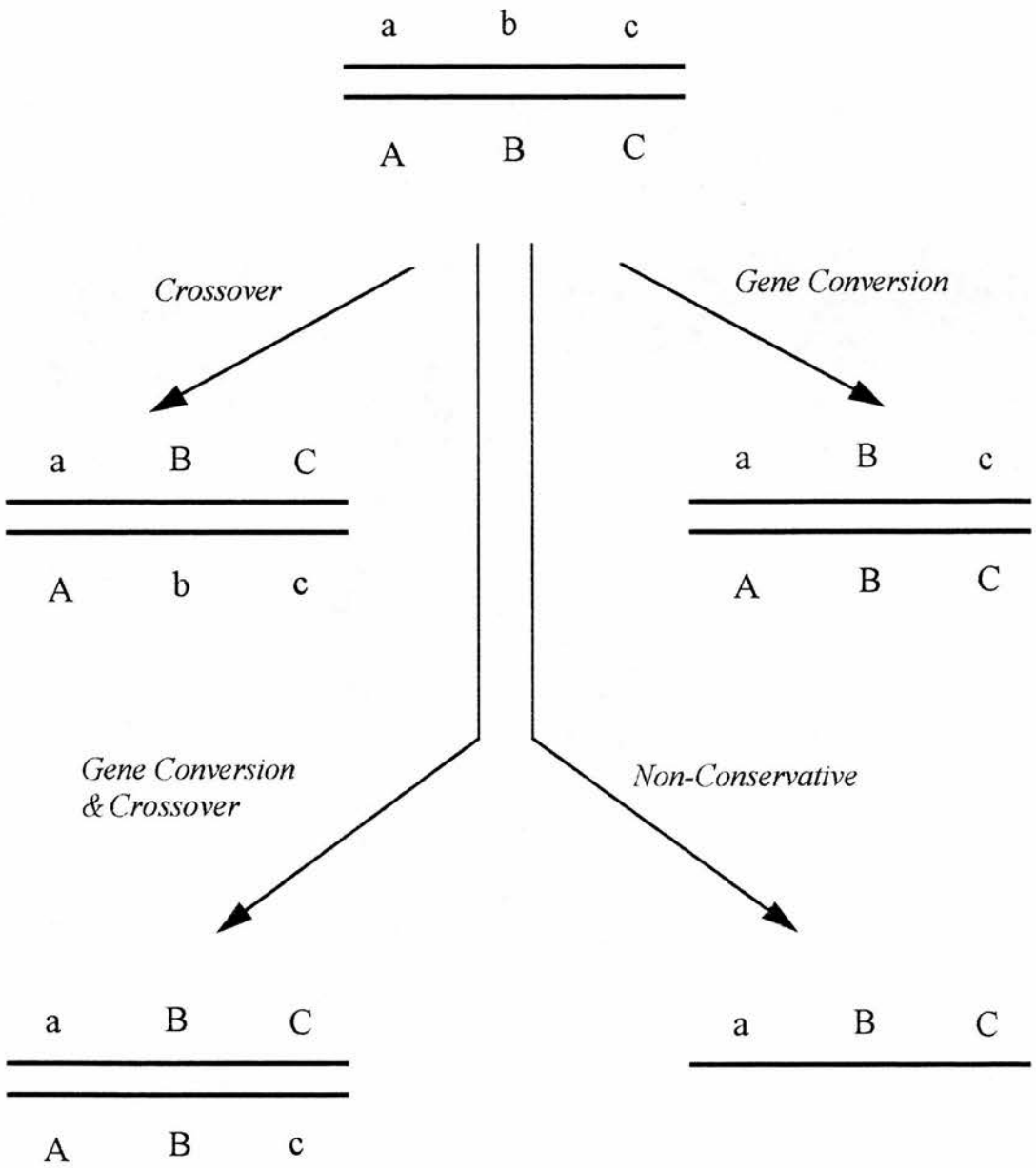
ECR has been extensively studied and has provided much information on the mechanism and requirements for recombination in mammalian cells. It should be noted however, that there are important differences between ECR and recombination involving the chromosomes (see below).

ECR studies (reviewed in Bollag *et al.*, 1989) have typically involved the introduction of a defective selectable system into the cultured cells, followed by selection as an assay for recombination. Such systems are typically either defective viral genomes, which produce a functional virus upon recombination and are monitored by the appearance of a plaque of lysed cells, or a selectable marker such as defective *neo* genes which confer resistance to the aminoglycoside G418 upon recombination to produce a functional gene. ECR may be assayed either by direct selection for the recombined marker, which relies on subsequent integration of the plasmid into the genome, or by harvesting of the low molecular weight DNA and transformation of recombination-incompetent bacteria followed by selection for recombination-dependent and recombination-independent markers. ECR frequencies of 1-20% have been measured in the former systems, and 10^{-2}-10^{-4} in the latter. Both intra- and inter-molecular recombination events have been observed. It has been suggested that almost every transfected DNA molecule may undergo at least one ECR event (Folger *et al.*, 1985). The mechanism of ECR may employ part of the molecular machinery required for integration of the molecule into the genome, which

Figure 1-3: The Possible Results of a Recombination Event

A and a, B and b, C and c represent different alleles at a genetic locus undergoing recombination. Each line represents a double stranded DNA molecule. Crossover results in the translocation of information, whereas gene conversion may result in the loss of genetic information. Non-conservative recombination results in the destruction of one of the recombining molecules, and hence always results in the loss of information.

Figure 1-3


a b c
A B C
Crossover
Gene Conversion
a B C
A b c
a B c
A B C
Gene Conversion
& Crossover
Non-Conservative
a B C
A B c
a B C

is why those assays for ECR reliant on subsequent stable transformation of the recombined plasmid give higher HR frequencies. In Rat-20 cells which had undergone an ECR event which reconstructed the *aprt* gene 50% of the cells went on to stably integrate the gene into the genome, in contrast to only 5% of cells transfected with wild-type *aprt* (measured in a transient expression assay) (Wong and Capecchi 1986).

Recombination is modulated temporally: DNA molecules must be introduced into the cell within an hour of each other for recombination to occur (Folger *et al.*, 1985), suggesting that within a relatively short time they become refractory to recombination, perhaps by assembly into chromatin. ECR in the *aprt* system was shown to occur preferentially (15-fold higher) during early to mid-S phase than in early G_1 phase of the cell cycle (Wong & Capecchi, 1987). Transcription stimulates recombination six-fold in mammalian cells (Nickoloff & Reynolds, 1990), also suggesting a role for the "activity" of DNA in recombination.

The homology requirements for HR are an important consideration. They differ between different types of recombination. For ECR the recombination frequency shows a biphasic relationship increasing with the length of homology, interpreted as there being two systems in mammalian cells capable of accomplishing ECR, with one operating on shorter lengths of homology than the other (Rubnitz & Subramani, 1984; Ayares *et al.*, 1986). This biphasic relationship was not however observed with replicating substrates (Ayares *et al.*, 1986). The transition occurs in the range of about 150-400 base pairs, with recombination still measurable with lengths of homology as short as 14 bp (Rubnitz & Subramani, 1984). 19% nucleotide mismatches reduce the frequency of ECR 3- to 15-fold compared to perfectly homologous sequences, which is a much smaller attenuation than the reduction in ICR under such conditions of over 1000-fold (Waldman & Liskay, 1987).

As expected from the yeast paradigm (Orr-Weaver *et al.*, 1981), double-stranded breaks within or near the region of homology increase the frequency of ECR some 10-fold (reviewed in Bollag *et al,* 1989). A break some distance from the homology has a smaller effect, but small heterologies do not significantly affect the recombination frequency. (Wake *et al.*, 1985).

1.3.2 Intrachromosomal Recombination

Intrachromosomal Recombination (ICR) occurs at lower frequencies (typically 10^{-6}-10^{-5} events/cell/generation) than ECR in mammalian cells (reviewed in Bollag *et al,* 1989) and requires longer lengths of homology than ECR: a minimum of between 134 and 232 bp (Waldman & Liskay, 1988). The recombination frequency increases in a linear fashion as the length of homology increases from 292-2000 bp (Liskay *et al.*, 1987). Regions of heterology reduce the frequency of ICR, but an adjacent region of perfect homology ameliorates this effect suggesting that initiation rather than propagation of the event is mainly affected and that the recombination efficiency is determined by the maximum length of perfect homology (Waldman & Liskay, 1988). ICR shows very high fidelity of recombination (Stachelek & Liskay, 1988). As is the case with ECR, the rate of ICR is enhanced by transcriptional activity of the recombining loci (Nickoloff, 1992).

1.3.3 Gene Targeting

The observations considered above have encouraged the use of homologous recombination as a strategy for modifying the mammalian genome. HR between introduced plasmid DNA sequences and homologous chromosomal sequences is called gene targeting.

Initial experiments, as with ECR and ICR, relied on the reconstruction of a crippled selectable marker as a method of selection for HR events and targeting frequencies of

10^{-3} to 10^{-7} with ratios of homologous to non-homologous events of 10^{-2} to 10^{-5} were observed (reviewed in Bollag *et al.*, 1989; Hooper, 1992).

Based on the finding that linearisation of the targeting vector within the region of homology increases the targeting frequency by up to 1000-fold in fungi (reviewed in Orr-Weaver & Szostak, 1985) and also stimulates ECR in mammalian cells (see section 1.1.6.1), mammalian gene targeting vectors are also linearised. Targeting constructs carrying a double-stranded break or gap in the homology undergo HR between 33 and 140 times as efficiently as uncut (supercoiled) plasmids (Valencius & Smithies, 1991a; Jasin & Berg, 1988) and the presence of a gap instead of a break does not significantly reduce the targeting frequency because the gap is repaired by gap repair, using information from the target gene (see figure 1-2). Heterologies close to the gap are often lost, suggesting that the gap may be enlarged during repair and/or that the free ends of DNA may not always directly participate in HR events; therefore perhaps the improvement is due to the greater steric freedom of linear DNA molecules than supercoiled ones (Shulman *et al.*, 1990; Valencius & Smithies, 1991a; Hasty *et al.*, 1992; Kumar & Simons, 1993; Deng *et al.*, 1993).

Copy number of either target sequence or vector is irrelevant to targeting frequency (Thomas *et al.*, 1986; Zheng & Wilson, 1990) and this has generally been taken to imply that the step of homologous sequence location by the targeting vector is not rate-limiting in a gene targeting reaction.

The first endogenous gene to be modified by gene targeting in mammalian cells was the β-globin gene, in human cells (Smithies *et al.*, 1985). Currently most gene targeting is carried out on murine ES cells with the objective of creating mice carrying specific mutations.

1.3.3.1 Gene Targeting Strategies

Two major systems are available when considering the strategy of a gene targeting experiment, which utilise different recombination possibilities observed in initial targeting experiments. Insertion (O-type) vectors are linearised within the region of homology and integrate in their entirety into the genome by means of a single crossover event at the site of linearisation. They thus result in a duplication of the homology region within the genome, separated by the vector sequences and selectable marker. Replacement (Ω-type) vectors are linearised outwith the region of homology and result in the replacement of a length of the genomic sequence with vector sequence comprising the homology which is divided into two lengths by intervening sequence, usually a selectable marker . This is usually assumed to occur by means of two separate crossovers although the observations that both types of vector can have similar efficiencies (Thomas & Capecchi, 1987) and that the length of homology on the small arm of the vector does not effect its targeting frequency provided it is above a minimum of less than 470 bp (Hasty *et al.*, 1991a) suggested that in fact only one crossover is necessary. Branch migration of the crossover through the vector to resolve at the other region of homology may be the mechanism (Ellis & Bernstein, 1989). These two strategies are shown in figure 1-4 parts (a) and (b).

There is some debate about the relative advantages of these two strategies both in terms of their targeting efficiencies and the nature of the structure they are likely to produce upon integration. Initial reports suggested that replacement and insertion vectors targeted the locus at the same frequency (Thomas & Capecchi, 1987, Deng & Capecchi, 1992), but some workers have found targeting an order of magnitude (Hasty *et al.*, 1991b) or greater (Dickinson *et al.*, 1993; Rudolph *et al.*, 1993a) more efficient with insertion vectors. Furthermore, it has been argued that replacement

Figure 1-4: Gene Targeting Strategies

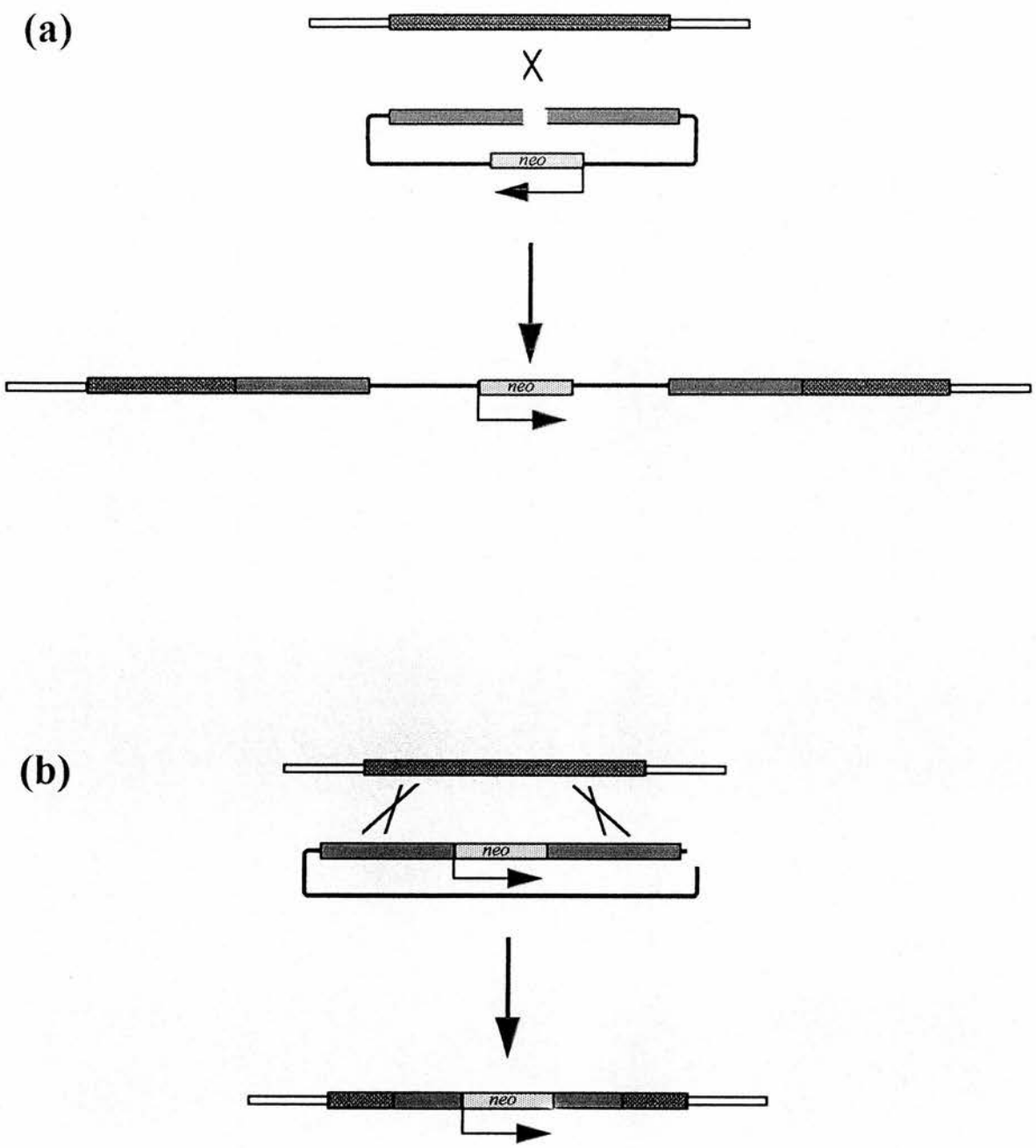
(a) Gene Targeting by Insertion

The bars represent genomic sequence, and lines represent vector (bacterial plasmid) sequence. A region of homology (light grey) to the genomic locus of interest (stippled grey) is linearised internally, so that recombination results in a locus containing a duplication of the region of homology interrupted by the vector sequence. Selection for a positive integration event using the *neo* gene, which confers resistance to G418 (orientation indicated by an arrow) and Southern analysis are used to identify the correctly-targeted clone.

(b) Gene Targeting by Replacement

The homology (light grey) to the target locus (stippled grey) is interrupted with the selectable marker (*neo*; orientation indicated by an arrow) to create a mutation. The vector is linearised outside the region of homology. A double-crossover event results in an exchange of vector sequence with the chromosomal sequence.

Figure 1-4



vectors can frequently recombine in an unexpected fashion (Hasty *et al.*, 1991b; Zhang *et al.*, 1994), but this may be related to the lengths of homology on each arm of the vector, and provided each arm is over 1kb in size the replacement vector will recombine as predicted (Thomas *et al.*, 1992).

1.3.3.2 Homology Requirements for Gene Targeting

The length of homology required for efficient gene targeting events has been studied to a limited extent, mostly in selectable genes such as *hprt* which provide a simple means of isolating and characterising homologous recombinants (See section 1.1.8). In common with other types of HR event in mammalian cells (see above), the targeting frequency of the *hprt* gene shows a dependence on the length of homology but there is some doubt about the exact relationship, since differences have been observed between experiments even in the same system. At the murine *hprt* locus in ES cells, an approximately linear relationship between targeting frequency and length of homology was observed over the homology range 1.3-6.8kb (Hasty *et al.*, 1991a), but an exponential relationship was observed by other workers (Thomas & Capecchi, 1987; Deng & Capecchi, 1992). There was also an exponential increase in targeting frequency at the *aprt* locus in Chinese hamster ovary (CHO) cells between 896 bp and 3275 bp (Scheerer & Adair, 1994). A greater than linear relationship between length of homology and targeting frequency was observed at the μ immunoglobulin locus in hybridomas, where an increase in targeting frequency of 4-fold was observed with an increase in homology from 4.3-9.5kb (Shulman *et al.*, 1990). Concordant with these observations, an increase in targeting frequency of 1.7-fold was seen at the unselectable *cftr* locus in murine ES cells when the homology was increased in size from 3.5kb to 4.3kb (Dickinson *et al.*, 1993).

Sizeable interruptions in homology are largely irrelevant to the targeting frequency (Mombaerts *et al.*, 1991; Zhang *et al.*, 1994), permitting the creation of relatively large insertion or deletion mutations.

The maximum size of homology resulting in an increase in targeting efficiency is an important factor, balancing ease of construction of the targeting vector with high targeting frequency (see above). Based on the initial finding that a relatively small (2-fold) increase in homology produced a large improvement (20-fold) in targeting efficiency (Thomas & Capecchi, 1987), it was suggested that the use of very large pieces of homology (greater than 30kb) would be of benefit (Bollag *et al.*, 1989). In experiments studying the relationship between length of homology and targeting frequency (see above) saturation of the mammalian HR apparatus appeared to occur in the region of 14 kb (Deng & Capecchi, 1992). With an isogenic construct incorporating 17kb of homologous sequence, 78% of integration events were by homologous recombination (Te Riele *et al.*, 1992), which suggests that under optimised conditions, HR rather than random integration can be the major event in mammalian cells, as it is in yeast (Rothstein, 1991). Several workers have reported much higher targeting frequencies than were first observed in gene targeting experiments when using vectors with upwards of about 10kb of homology (Rudolph *et al.*, 1993a; Deng & Capecchi, 1992).

A related issue to the length of homology in a targeting vector is the extent to which small mismatches in the homology can reduce the efficiency by reducing the size of the maximum length of uninterrupted homology (see section 1.1.6.2). Several studies have shown improvements in targeting frequency of between 4- or 5- and 25-fold by the use of isogenic constructs (Te Riele *et al.*, 1992; Van Deursen & Wieringa, 1992; Deng & Capecchi, 1992) over the use of non-isogenic DNA which had small heterologies of approximately 0.7-2% (Te Riele *et al.*, 1992; Van Deursen &

Wieringa, 1992). A *Ren-1D* targeting vector which was also greater than 95% homologous to the *Ren-2* gene was observed to target only *Ren-1D* (Miller *et al.*, 1992). However, individual mismatches separated from the site of linearisation by 2.4kb and 3.2kb of isogenic sequence did not adversely affect targeting frequency, suggesting that the minimum length of perfect homology required for efficient targeting is smaller than this (Dickinson *et al.*, 1993).

1.3.3.3 Fidelity of Gene Targeting

When introducing subtle mutations into the genome, the fidelity of gene targeting by HR is an important consideration because the effect of the intended change may be complicated by the presence of subsidiary mutations introduced by the HR process. Unfortunately this is an area where relatively little information exists. Some early results suggested that targeted HR may introduce mutations at the target locus (Thomas & Capecchi, 1986; Doetschman *et al.*, 1988; Thompson *et al.*, 1989; Brinster *et al.*, 1989), in contrast to ICR which is known to be very accurate (Stachelek & Liskay, 1988). Thomas & Capecchi observed that in targeting experiments designed to correct a deficient *neo* gene by HR, 50% of the recovered clones had a second mutation rather than a correction at the site of the first mutation. Since not all introduced mutations are likely to result in restored *neo* function the actual mutation rate was probably much higher. Doetschman *et al.* and Thompson *et al.* reported that deletions were observed in the *hprt* gene after targeting. Brinster *et al.*, using pronuclear injection as their method of introducing DNA, found 15 new mutations introduced into a 1.5kb length of sequence. However, to date only one systematic high-resolution study of the fidelity of gene targeting reactions has been carried out (Zheng *et al.*, 1991). In this study though, chemical cleavage of sequence mismatches surrounding the site of integration of the targeting vector showed only

two point mutations in 80kb of sequence from 44 studied clones, and thus it was concluded that gene targeting can be extremely accurate.

1.3.3.4 Enriching and Screening for Targeting Events

In the majority of gene targeting experiments, integration of the targeting construct occurs rarely compared to the number of electroporated cells (typically around 10^{-5}) and of these integration events the proportion of legitimate HR events may vary widely (over 3 orders of magnitude) between genes (reviewed in Camerini-Otero & Kucherlapati, 1990). The reasons for such wide variability are unknown, and conclusions are difficult to draw because of the complication of the picture by the very different strategies, characteristics of homology used etc, in different targeting experiments. The organisation into chromatin of genes, or their location in "hot-spots" for recombination, or regions where it is suppressed, may be a significant factor. Whilst transcription of genes may improve targeting efficiencies (Mansour *et al.*, 1988), it is not essential for a gene to be "targetable" by HR (Smithies *et al.*, 1985; Johnson *et al.*, 1989; Jeanotte *et al.*, 1991).

Therefore selective methods and/or rapid screening techniques are required in order to isolate the desired clones. Some initial targeting experiments in ES cells were carried out on the *hprt* locus, permitting direct selection of targeted recombinants (Thomas & Capecchi, 1987; Doetschman *et al.*, 1987, 1988; Thompson *et al.*, 1989), but clearly methods applicable to unselectable genes are required.

Expression of a positive selectable marker is used to isolate those clones derived from cells which have stably integrated the targeting plasmid into the genome. Usually this is *neo*, which confers resistance *in vitro* to the synthetic aminoglycoside G418 (Southern & Berg, 1982), but others such as *hyg*, conferring resistance to hygromycin B, and *hprt* minigenes (Selfridge *et al.*, 1992; Detlott *et al.*, 1994) are

also available. Multiple selections to enrich for HR events at the expense of non-homologous events have been applied and this has increased the efficiency of selection to such an extent that isolation of targeted clones is often a relatively simple matter.

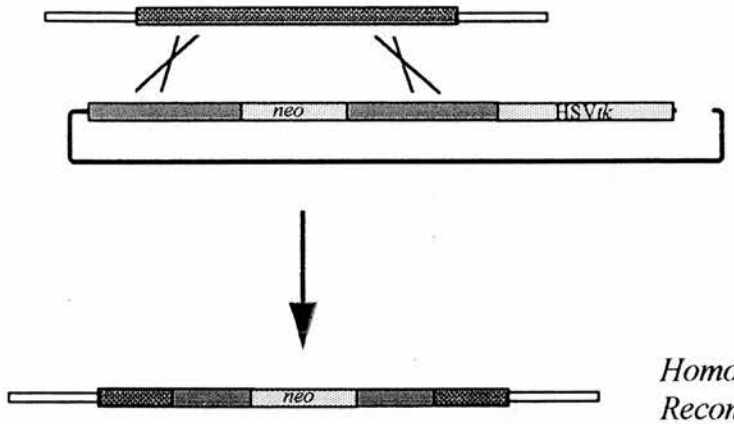
The first of these is the positive-negative selection (PNS), used with replacement vectors (Mansour *et al.*, 1988), and this has become the standard method for the production of gene knock-outs in mice over the past few years. The strategy is shown in figure 1-5. Initially PNS was reported as giving enrichments of 2000-fold for HR events over illegitimate events (Mansour *et al.*, 1988), but this appears to be the exception rather than the rule and more normally enrichments of about 2-20-fold are observed (Zijlstra *et al.*, 1989; Mombaerts *et al.*, 1991; Rudolph *et al.*, 1993a; Hanson & Sedivy, 1995). This is likely to be due in part to exonuclease degradation of the negative selectable marker prior to integration of the plasmid into the genome. Such reductions in enrichment may be ameliorated by the presence of additional sequence or hairpin-shaped oligonucleotide linkers on the free ends of the construct to protect the negative selectable gene (Bernetgrandaud *et al.*, 1992; Horie *et al*, 1994; Horie & Shimada, 1994). The diphtheria toxin A chain has also been used as the negative selectable marker in the PNS procedure, where a 9- to 29-fold enrichment for HR events was observed (McCarrick *et al.*, 1993).

An alternative strategy for enrichment of homologous recombination events is to use a selectable marker gene in the construct lacking either a functional promoter or polyadenylation sequence or requiring the trapping of an enhancer (Doetschman *et al.*, 1988; Dorin *et al.*, 1989; Jasin & Berg, 1988; Sedivy & Sharp, 1989; Schwartzeberg *et al.*, 1989; Joyner *et al.*, 1989; Jeanotte *et al.*, 1991). Such an approach enriches for correctly-targeted events by preventing expression of the positive selectable marker when an illegitimate recombination event occurs: it is

Figure 1-5: Positive-Negative Selection

The strategy is a refinement of targeting by replacement (figure 1-4). Another selectable marker (HSV*tk*) is situated outside the region of homology. This selectable marker permits negative selection with ganciclovir (GANC). A faithful homologous recombination event results in the loss of this marker, so that a correctly-targeted clone is $G418^r$, because of the *neo* insertion, and $GANC^r$, because the HSV*tk* gene is not incorporated (upper structure). A random insertion of the vector into the genome is most likely to occur via the free ends of the structure however, so that the HSV*tk* gene is also likely to become integrated (lower structure). These events may be eliminated by virtue of their susceptibility to GANC, thus enriching for correctly-targeted events.

Figure 1-5



Homologous Recombination:

G418^{r} GANC^{r}

Illegitimate Recombination:

G418^{r} GANC^{s}

relatively improbable that the construct will integrate adjacent to a sequence capable of activating expression of the *neo* gene, so that it remains silent and the cell dies.

Different selections vary in efficacy, for example *hyg* produces 200-fold fewer colonies than wild-type *neo*; a weaker mutant *neo* was used to improve the targeting efficiency of the *c-myc* gene, which is expressed at a relatively low level (Hanson & Sedivy, 1995). Selectable markers should therefore be selected according to their biological properties and those of the gene to be modified. Wild-type *neo* is such a strong selection that it may be used without a promoter to target genes with extremely low levels of expression (Jeanotte *et al.*, 1991; DeGregori *et al.*, 1994).

Screening of the resultant clones by Southern analysis or PCR is then used to identify those clones which have undergone an HR event at the desired locus. PCR across the junction of the insertion of the vector (Kim & Smithies, 1988) provides one of the most sensitive methods of detection of targeted clones, permitting the use of pools of clones to screen very large numbers, but is prone to error (Frohman & Martin, 1990; Kim *et al.*, 1991) and should therefore be used in conjunction with a confirmatory screen carried out by Southern analysis. A similar method of screening large cell numbers in groups is to enrich for the desired cells by sib-selection (Cavalli-Sforza & Lederberg, 1956) coupled to PCR.

More recently the detection of homologous recombinants has been achieved by assaying for secretion of a protein which will only occur in recombinants (Itzhaki & Porter, 1991; Smith & Kalagerakis, 1991), either using an inducible promoter or modifying an allelic variant of the targeted secreted protein.

1.3.3.5 Strategies for Site-Directed Mutagenesis in the Mouse

The discussion above pertains to gene targeting in general. Most targeting experiments described to date have had the objective of introducing a large mutation into a gene in ES cells to create a null allele, with the aim of generating a "knockout" mouse as an aid to study of the gene's function *in vivo* and as a source of cell lines carrying defined mutations. Phenotypes obtained with knockout mice have varied from having no observable effect on health and viability to being lethal in early embryonic life, and different targeted mutations in the same gene have produced different phenotypes. In the case of the cystic fibrosis transmembrane receptor *cftr* a complete knockout (Snouwaert *et al.*, 1992; Colledge *et al.*, 1992; O'Neal *et al.*, 1993) has a perinatal lethal phenotype, but a mutation achieved by insertion can be spliced out in a minority of transcripts and the reduced level of expression provides a model of human cystic fibrosis (Dorin *et al.*, 1992). Similarly, mutations in the *Hox* 2.6 gene have subtly different phenotypes depending on the site of interruption of the gene (Ramirez-Solis *et al.*, 1993), and a knockout of the N-*myc* proto-oncogene is lethal at day 10.5 of embryogenesis (Stanton *et al.*, 1992; Charron *et al.*, 1992; Sawai *et al,* 1993) whereas a "leaky" mutation permits survival to birth (Moens *et al.*, 1992). Some mutations have proved to be disappointing in that they produce a phenotype which gives little clue as to the function of the gene, or do not model the human disease for which they were designed.

The introduction of more subtle mutations such as missense mutations to modify specific functions of a protein is therefore desirable and likely to become more important. Such small modifications to the genome present problems in terms of their creation without other changes being co-introduced, and in terms of their detection.

A point mutation may introduce a selectable phenotype as is sometimes the case with conventional gene inactivation; for example introduction of a point mutation into

RNA polymerase II conferred resistance to α-amanitin (Steeg *et al.*, 1990), but this is likely to be the exception rather than the rule. There are several strategies now available for such experiments, including microinjection of a targeting construct with no selectable marker (Zimmer & Gruss, 1989) or coelectroporation of the selectable marker (Shulman *et al.*, 1990; Davis *et al.*, 1992), replacement where the selectable marker is located in a nominally "neutral" part of the homology (Rubinstein *et al.*, 1993), "plug & socket" targeting (Detlott *et al.*, 1994; figure 1-6), "tag and exchange"/double replacement targeting (Askew *et al.*, 1993; Stacey *et al.*, 1994; Wu *et al.*, 1994; figure 1-7), site-specific recombinase systems (Gu *et al.*, 1993; Fiering *et al.*, 1993; Jung *et al.*, 1993) and "hit and run"/"in-out" targeting (Hasty *et al.*, 1991c; Valencius & Smithies, 1991b; figure 1-8). They have slightly different characteristics, so that selection of the technique can be chosen according to the objectives of the experiment. The technique of Rubinstein *et al* is essentially a standard replacement/PNS-type strategy. However, instead of incorporating a deletion or insertional mutation caused by the *neo* gene interrupting exonic sequence into the homology, they inserted a point mutation into the proopiomelanocortin (POMC) gene, and located the *neo* cassette downstream of the polyadenylation site. Although they showed that the POMC gene was not activated in reverse by the PGK promoter, as may have been expected (Johnson & Friedman, 1990), this system obviously relies on the fact that a suitably innocuous region of the genome will be available. Especially for large genes, and occasions where a mutation is to be introduced into an exon which is some distance from sequences outwith the gene, an intronic site would have to be selected. This may be problematic due to the interruption of enhancer-type sequences or the introduction of cryptic splice-acceptor sites within the selectable marker. For example, an insertional disruption of the mouse major adult β-globin gene created a much more severe phenotype than a deletion at the locus, and it was suggested that this was due to the presence of the additional promoter driving the

Figure 1-6: "Plug and Socket" Gene Targeting

The targeting is carried out in *hprt*$^-$ cells, to permit selection for *hprt* in the second step. The regions of homology (medium grey and white) are interrupted by a *neo* gene, to permit positive selection for a correctly-targeted clone, and a defective *hprt* gene. Vector 1 is targeted into the locus by replacement using *neo* selection, to create the "socket". This clone, which is G418^r, HAT^s, is then targeted with vector 2 (the "plug"), carrying a subtle mutation (light grey), and another defective *hprt* gene. Since there are different deletions in the *hprt* genes in both vectors 1 and 2 (hatched regions), the only HAT^r clones resulting are those which have correctly recombined to give a functional *hprt* gene at the target locus, and in the process introduced a subtle mutation into the target gene. Screening of HAT^r clones is then carried out to identify those carrying the linked mutation.

Figure 1-6

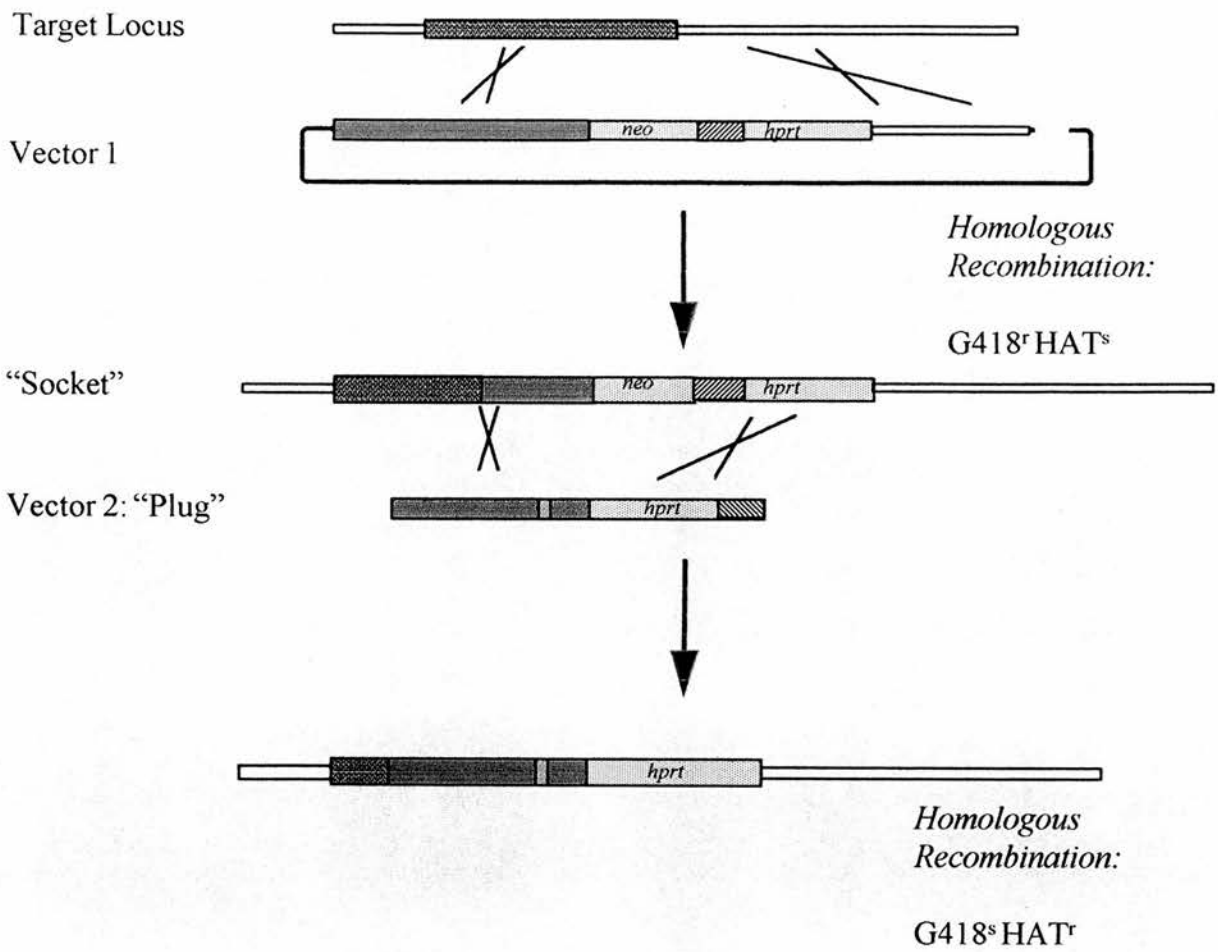

Target Locus
Vector 1
neo
hprt
Homologous Recombination:
G418r HATs
"Socket"
neo
hprt
Vector 2: "Plug"
hprt
hprt
Homologous Recombination:
G418s HATr

Figure 1-7: Double Replacement Gene Targeting

This strategy also uses two vectors, like plug and socket targeting. In the figure an *hprt* minigene is used for both positive and negative selection, but *neo* for positive selection and HSV*tk* for negative selection have also been successfully employed (see text). In the first step, PNS targeting is used to generate a HAT^r knockout clone. Here a selection based on hypoxanthine and azaserine (which does not inhibit thymidylate synthetase) would be used to avoid the thymidine present in HAT medium rescuing the HSV*tk* selection. The targeted clone is then re-targeted with a second vector carrying homology identical to the endogenous locus except for the desired mutation (light grey square). The double-targeted clone is once again HAT^s, and is then screened to check for presence of the designed mutation.

Figure 1-7

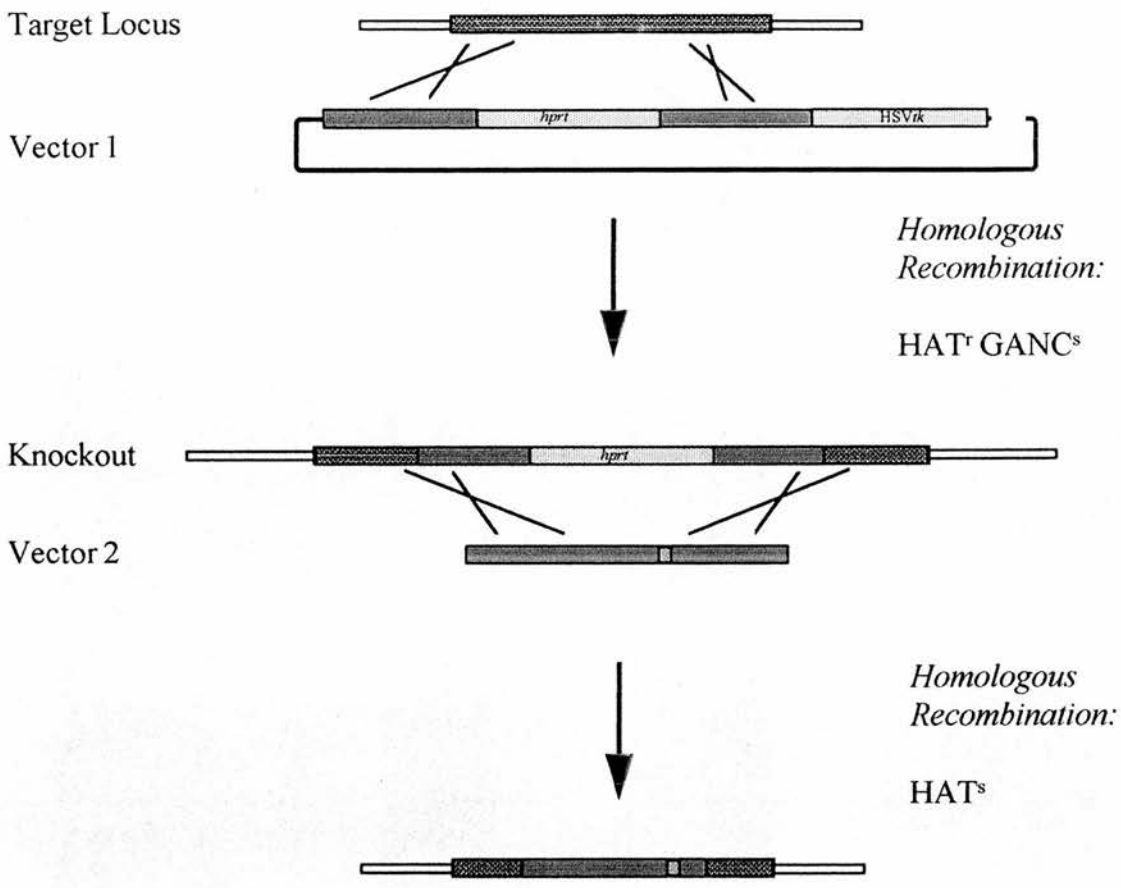
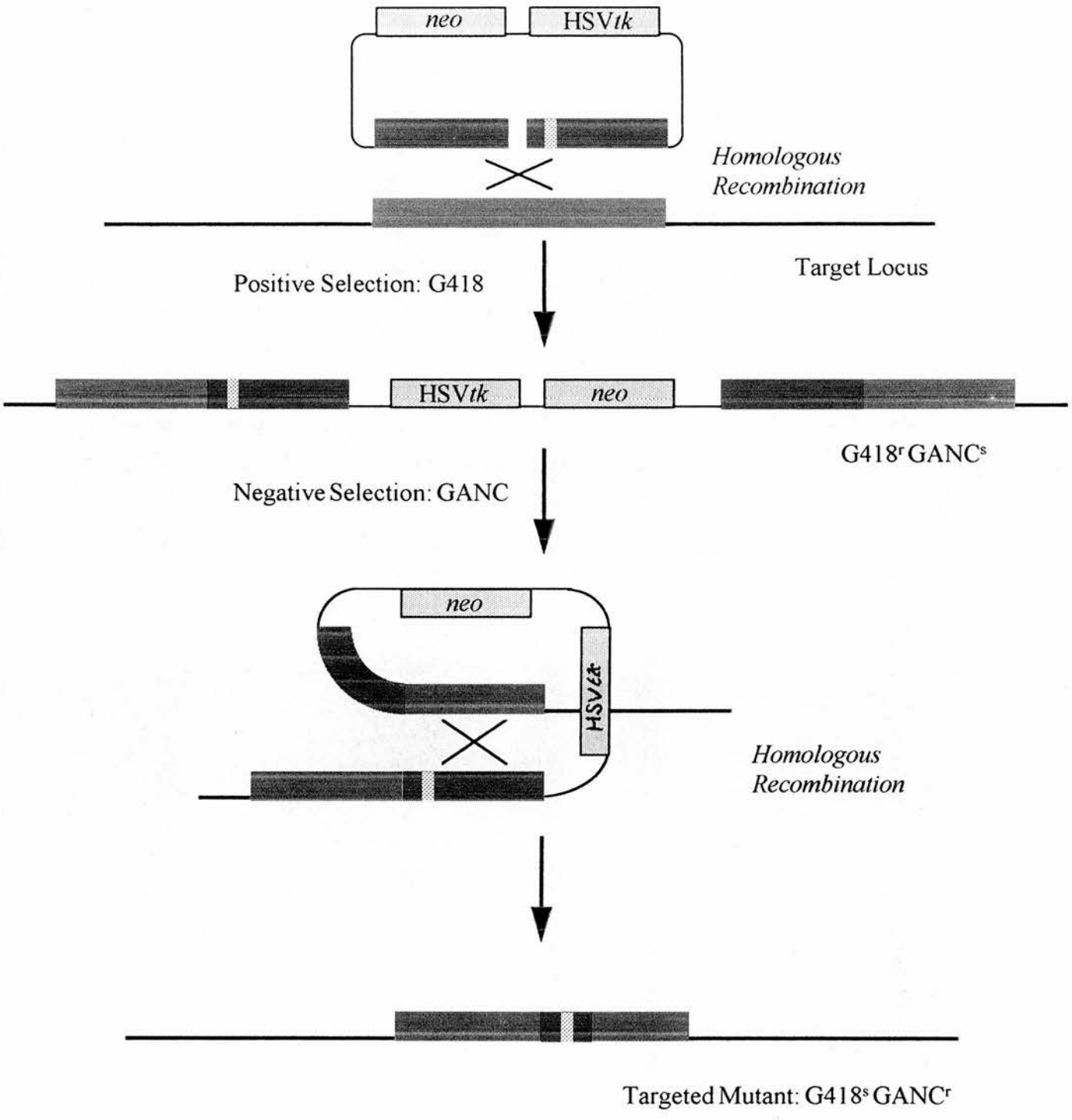

Target Locus
Vector 1
hprt
HSVtk
Homologous Recombination:
HAT^r GANC^s
Knockout
Vector 2
hprt
Homologous Recombination:
HAT^s

Figure 1-8: Hit and Run Gene Targeting

An insertion vector carrying both positive and negative selectable markers (*neo* and HSV*tk* respectively) and a subtle mutation in the region of homology is used to target the locus of interest, generating a structure consisting of a linear duplication of the region of homology interrupted by the vector sequence, which is neo^r. Depending on the result of any branch migration event, the mutation may be in either the left duplication, the right duplication, both duplications or neither duplication of homology. The clone is then subjected to backselection (negative selection; in this case GANC) and intra-chromosomal recombination events between the duplications of homology either regenerate the wild-type structure or insert the mutation, depending on the site of crossover.

Figure 1-8


neo
HSVtk
Homologous Recombination
Target Locus
Positive Selection: G418
HSVtk
neo
G418r GANCs
Negative Selection: GANC
neo
HSVtk
Homologous Recombination
Targeted Mutant: G418s GANCr

neo cassette introduced into the region interacting with the Locus Control Region at the expense of endogenous promoters driving globin gene transcription (Shehee *et al.*, 1993).

The plug and socket system developed at North Carolina also leaves the selectable marker in the genome and so the same caveat must apply here as well. However, this system is useful where many mutations are to be produced, since the same "socket" can be used to generate many different mutants by the use of different "plug" constructs. The positive selection at the second step is also more effective than negative selection strategies, such as are used in hit and run or tag and exchange.

Recombinase enzyme systems rely on the use of either the Cre/*lox-P* (derived from bacteriophage P1) or FLP recombinase (derived from yeast) systems to remove the selectable marker sequence. Recombinase-mediated systems require the placement of the necessary *cis*-acting sequences either side of the sequence to be deleted, and leave a single copy of this sequence in the genome (about 50 bp) after excision by the specific enzyme, but are very efficient.

Targeted recombination events are rare, of the order of 10^{-5} to 10^{-6} per transfected cell (see above), but as only 1 in 10^3 cells integrates the transfected DNA it is possible to enrich for targeted clones by selecting for transfected cells with a selectable marker. Thus the coelectroporation strategy also introduces a selectable marker into the genome, but relies on the fact that the mouse genome is sufficiently large (2000 centimorgan) that the chance of the introduced mutation and the *neo* selectable marker becoming closely linked is only about 0.05%. Separation of the selectable marker from the introduced point mutation can therefore be achieved by meiosis. However, with this strategy a laborious PCR screen is required and a mouse must be

generated before the allele of interest can be studied in the absence of the selectable marker.

The only methods therefore which definitively effect the precise modification of the genome and introduce no other changes which might complicate analysis of the phenotype are microinjection of sequences, tag and exchange and hit and run. The former however has only been achieved on one occasion, as yet without germline transmission of the produced allele, and therefore its generality as a strategy must remain in doubt.

The hit and run strategy was the first of the latter two to be shown to be practicable (Hasty *et al.*, 1991c; Valencius & Smithies, 1991b), and has been employed on several occasions since (Rudolph *et al.*, 1993a, 1993b; Serwe & Sablitzky, 1993; Bautista & Shulman, 1993; Ernst *et al.*, 1994; Gorry *et al.*, 1994). This system requires the construction of only one vector, making it preferable in situations where only one mutation is to be introduced at the locus. As shown in the diagram, reversion may be either to wild-type or the desired mutant. Reversion occurs at widely different frequencies depending on the gene studied (Hasty *et al.*, 1991c), as is seen for initial targeting events (Camerini-Otero & Kucherlapati, 1990). "Hit and run" has also been used to generate conventional insertion-mutant type structures where replacement vectors have proved ineffective (Rudolph *et al.*, 1993a).

Tag and exchange, or double replacement, relies on a second targeting event with another construct to introduce the planned mutation (figure 1-7). This method is likely to be useful where multiple independent modifications of the locus are required, since the same "tagged" gene may be used for any number of exchanges. Furthermore, the tag itself may have utility as a knockout allele against which to compare subsequent, more subtle, mutations. In such cases however, it should be noted that the use of the HSV*tk* gene as the negative selectable marker is

inappropriate because testicular expression from a cryptic promoter is incompatible with germline transmission of the mutant allele (Al-Shawi *et al.*, 1991). Loss of the second selectable marker can occur in ways other than by HR, and a high background of non-targeted cells has been observed after electroporation of the second vector (Bradley, 1993), but this should be less of a problem when using *hprt* selection rather than, for example, HSV*tk* selection (Stacey *et al.*, 1994) because positive selection for gene function may be maintained until introduction of the second construct.

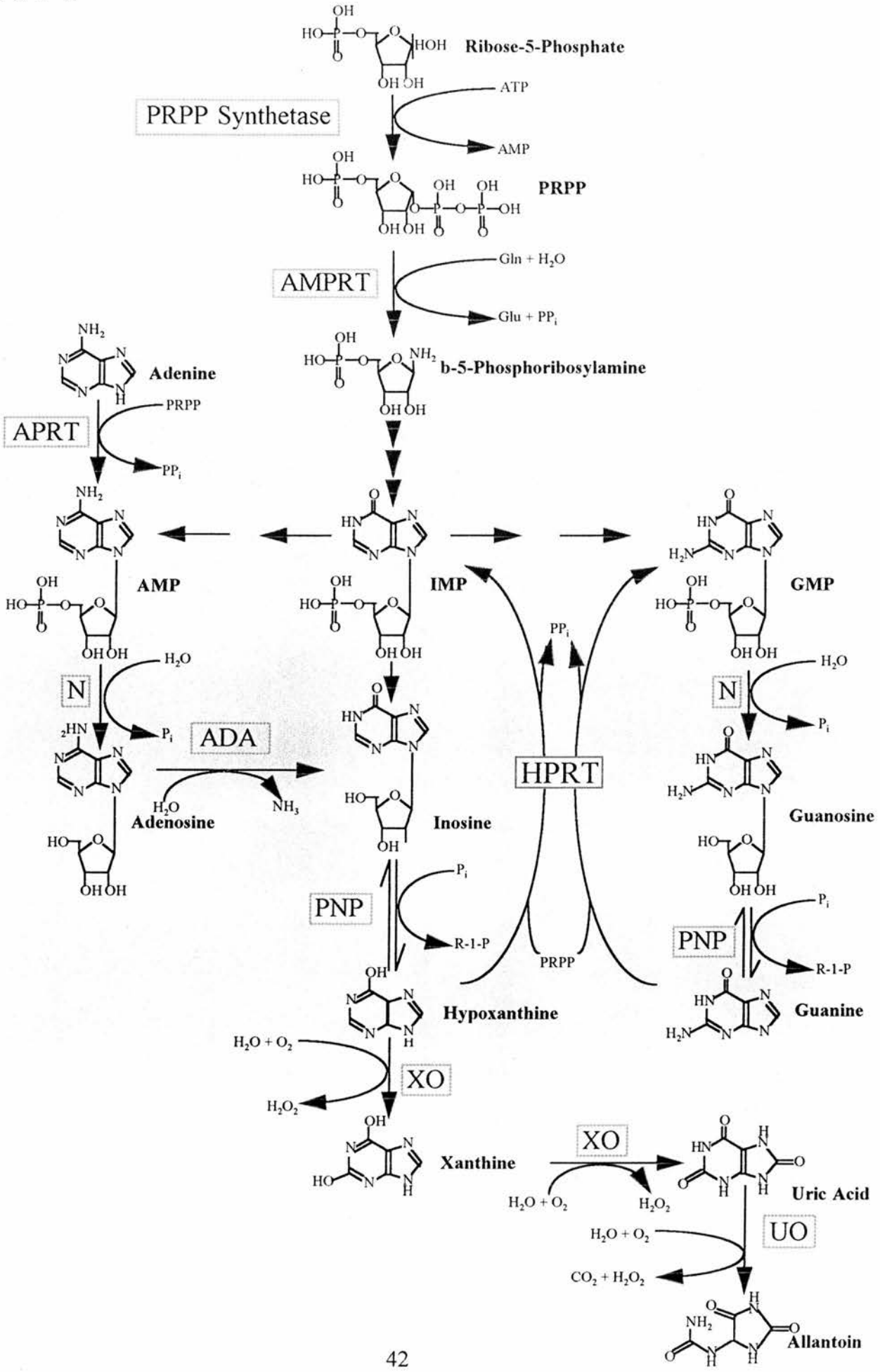
1.4 The *Hprt* System (Reviewed by Stout & Caskey, 1985)

The hypoxanthine-guanine phosphoribosyltransferase (*hprt*) gene has proved to be a useful system in studies aimed at developing gene targeting strategies or studying the mechanism of HR in mammalian cells. Two features make *hprt* suitable for such experiments. *Hprt* is X-linked, and therefore a single copy of the gene may still be observed at the level of the phenotype, because the gene is hemizygous in males and in females only one copy of the gene will be expressed because of X-chromosome inactivation. As remarked in section 1.1.5, most ES cells are male and therefore can be expected to behave identically with respect to *hprt*-dependent selection. *Hprt* function may be selected either for or against in cultured cells by chemical selection, and thus a simple phenotypic change can be used to follow the genotype of the gene. *Hprt* is a housekeeping gene, expressed at a low level in most tissues. It is a part of the purine metabolism pathway (figure 1-9; see, for example, Voet & Voet, 1990), serving to salvage hypoxanthine and guanine by transfer of a phosphoribosyl group from 5'-phosphoribosyl-1-pyrophosphate (PRPP) to produce inosine monophosphate (IMP) or guanosine monophosphate (GMP), respectively. Cells can also synthesise these mononucleotides *de novo*, but this pathway is blocked by the drug aminopterin and the cell can thus be made dependent upon its salvage pathway and will die unless functional *hprt* can provide purines. This is the basis of the HAT (hypoxanthine-

Figure 1-9: Purine Metabolism

The figure shows a schematic outline of the purine salvage pathways in mammals. Abbreviations: ADA, adenosine deaminase; AMP, adenosine monophosphate; AMPRT, amidophosphoribosyltransferase; ATP, adenosine triphosphate; GMP, guanosine monophosphate; HPRT, hypoxanthine-guanine phosphoribosyltransferase; IMP, inosine monophosphate; N, nucleotidase; P_i, inorganic phosphate; PNP, purine nucleoside phosphorylase; PP_i, pyrophosphate; PRPP, 5'-phosphoribosyl-1-pyrophosphate; R-1-P, ribose-1-phosphate; UO, urate oxidase (this step does not occur in humans); XO, xanthine oxidase

Figure 1-9


Ribose-5-Phosphate
PRPP Synthetase
ATP
AMP
PRPP
AMPRT
Gln + H2O
Glu + PPi
b-5-Phosphoribosylamine
Adenine
APRT
PRPP
PPi
AMP
IMP
GMP
N
H2O
Pi
ADA
Adenosine
NH3
Inosine
HPRT
PPi
PRPP
Guanosine
PNP
Pi
R-1-P
Hypoxanthine
Guanine
H2O + O2
H2O2
XO
Xanthine
Uric Acid
UO
CO2 + H2O2
Allantoin

aminopterin-thymidine) selection (Szybalski & Szybalska, 1962): hypoxanthine is provided as a substrate for *hprt*-catalysed purine conversion because aminopterin blocks the *de novo* pathway; thymidine is also included in the selection as aminopterin, being an antifolate drug, also inhibits the thymidylate synthase reaction.

Hprt is also capable of phosphoribosylating the toxic guanine analogue 2-hydroxy-6-mercaptopurine (6-thioguanine, 6-TG) and this is fatal for the cell (Stutts & Brockman, 1963). It is thus possible to select either for function or for lack of function in *hprt* very simply.

The structure of the murine *hprt* gene is shown in figure 2-1, together with the structure of the mutant $hprt^{bm\text{-}3}$ allele in the ES cell line E14TG2a (Hooper *et al.*, 1987), which carries a deletion encompassing the promoter sequences and exons 1 and 2 (Thompson *et al.*, 1989).

1.5 The Kirsten-*Ras* Proto-Oncogene.

1.5.1 General Introduction: The *Ras* Genes

The *ras* genes comprise a gene family and are ubiquitous amongst eucaryotic organisms, having been identified in mammals, birds, insects, molluscs, nematodes, fungi and plants (reviewed by Barbacid, 1987; Valencia *et al.*, 1991). They are part of a greater superfamily of genes which all encode small guanine nucleotide-binding proteins of 20-25kd in size, which as well as *ras* also include the families of the *rac, rap, rab, ral* and *rho* genes.

Mammals encode 3 *ras* genes, designated K-*ras*-2, H-*ras*-1 and N-*ras*, in the mouse present on chromosomes 6, 7 and 3, respectively (O' Brien, 1984). They encode highly-related proteins generically known as p21 (Shih *et al.*, 1979). In humans and

rats, pseudogenes of H-*ras* (H-*ras*-2) and K-*ras* (K-*ras*-1) have also been characterised.

The Kirsten-*ras* proto-oncogene (hereafter, K-*ras*) was one of the first genes discovered to have an association with malignancy, being described in the context of the Kirsten rat sarcoma virus as the transforming element of the virus (Kirsten & Mayer, 1967). K-*ras* was subsequently shown to be a mutant homologue of a cellular gene transduced by the retrovirus (Ellis *et al.*, 1981). H-*ras* was identified similarly (Harvey, 1964; DeFeo *et al.*, 1981). N-*ras* was identified as a transforming gene present in a neuroblastoma cell line homologous to H-*ras* and K-*ras* (Shimizu *et al.*, 1983a).

The biology of *ras* has provoked intense study in the last 15 years because of the association of the *ras* genes with human malignancies (reviewed by Bos, 1989); approximately 30% of cancers, including almost 100% of pancreatic carcinomas, 50% of colon and thyroid cancers and 30% of lung malignancies and myeloid leukaemias, have mutant *ras* alleles.

The effects of *ras* may be observed in bioassays, where *ras* genes or proteins are added to a system to elicit a *ras*-dependent response. This may be cell-proliferation dependent, such as focal transformation of established rodent fibroblasts, DNA synthesis, tumour formation or transformation of a primary cell line in cooperation with another oncogene such as *myc*, or morphological such as the induction of differentiation in an appropriate cell type (reviewed by Lowy and Willumsen, 1993). The wide variety of bioassays available for *ras* is a reflection of its involvement in diverse effects such as proliferation and differentiation in a multitude of biological systems.

1.5.2 Structure of the *Ras* Genes

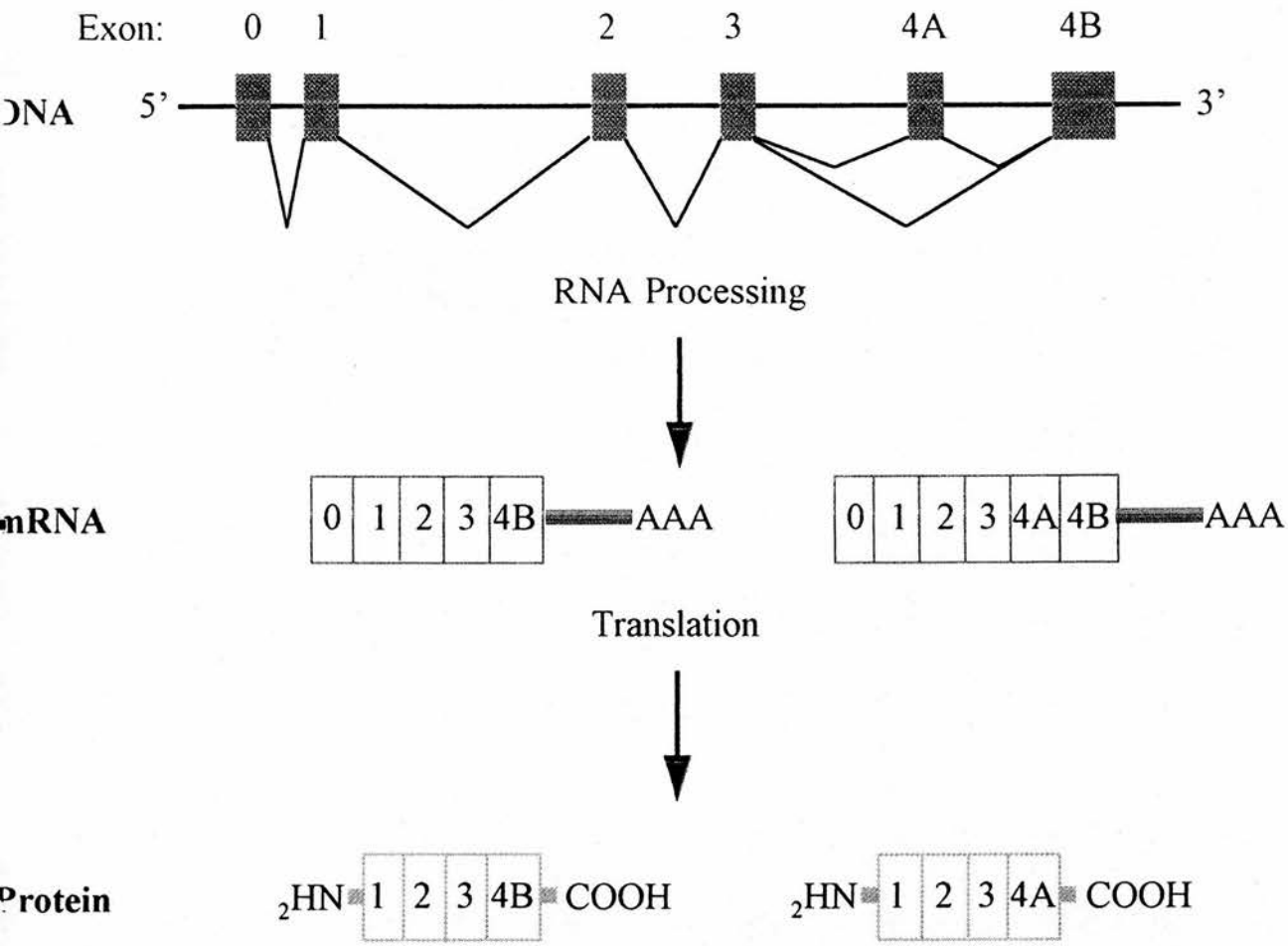
In this and the following sections, it should be noted that many characteristics of K-*ras* also apply to other *ras* genes as well, and therefore where appropriate *ras* or $p21^{ras}$ will be used to denote general findings.

The three mammalian *ras* genes have a similar structure, consisting of a 5' noncoding exon (exon 0) and four coding exons. The introns however are very heterogeneous, so that the genes vary in size: N-*ras* and H-*ras* span approximately 7kb and 3 kb respectively, whereas K-*ras* is much larger, covering 35-40kb. K-*ras* also differs from the other *ras* genes in that it codes for alternative fourth coding exons designated 4A and 4B (McGrath *et al.*, 1983; Shimizu *et al.*, 1983b; Capon *et al.*, 1983), making a total of 6 exons in all coding for two isomorphic proteins which differ in their C-terminal regions. A diagram of the K-*ras* gene and its transcription and translation products is shown in figure 1-10. N-*ras*, H-*ras* and K-ras^A all encode proteins of 189 amino acids; exon 4B has one less codon and therefore K-ras^B is 188 amino acids long. Between *ras* genes the splicing junctions of the exons correspond exactly, suggesting that the proteins are derived from a common ancestral protein. They are highly conserved both between species and between each other (Barbacid, 1987; Lowy & Willumsen, 1993). K-ras^A and K-ras^B proteins are identical from the N-terminus to residue 164 (i.e. to approximately 1/3 way through exon 4) except for differences at residues 151 and 153, but are then extremely heterogeneous in the remaining 24/25 amino acids leading to the C-terminus. However, between species exons 4A and 4B are highly conserved, suggesting that the heterogeneous regions of the proteins are associated with specificity of function. This structure of a highly-conserved N-terminal region and a divergent C-terminal region is usual in *ras* genes (Barbacid, 1987; Valencia *et al.*, 1991; Lowy & Willumsen, 1993), but within the heterogeneous region some conservation is observed. A cysteine residue is always

Figure 1-10: The Kirsten-*ras* Gene

K-*ras* is a gene comprising 6 exons, including a 5' untranslated exon (exon 0) and five translated exons, designated 1, 2, 3, 4A and 4B, spanning approximately 40kb of genomic sequence (not to scale). Alternative splicing possibilities may generate two mRNA species, depending on whether exon 4A is included or not. Because there is a nonsense codon at the end of exon 4A, the two protein isoforms generated either have the amino acid sequence corresponding to exons 1-2-3-4A or 1-2-3-4B, differing just at their C-termini.

Figure 1-10


Exon:
0
1
2
3
4A
4B
DNA
5'
3'
RNA Processing
mRNA
0 1 2 3 4B AAA
0 1 2 3 4A 4B AAA
Translation
Protein
2HN 1 2 3 4B COOH
2HN 1 2 3 4A COOH

present at codon 186 (185 in K-*ras*^{B}) and is followed by two aliphatic amino acids and a residue which is usually either serine or methionine. The C-terminal four residues are therefore collectively referred to as the CAAX motif, and their strong conservation between genes and species reflects their importance for *ras* function (see section 1.2.3.2).

The promoter of K-*ras*, like that of the other two *ras* genes, is located in the region of the untranslated exon 0 (Jordano & Perucho, 1986; Hoffman *et al.*, 1987; Yamamoto & Perucho, 1988). It lacks a TATA and CAAT box, but contains multiple copies of the GC box sequences GGGCGG/CCGCCC both 5', 3' and internal to exon 0. Transcriptional activity appears to be bi-directional from these sites (Hoffman *et al.*, 1987). Such sequences are recognised by the Sp1 transcription factor (reviewed by Kadonaga *et al.*, 1986) and are characteristic of "housekeeping" genes. Sequences at the 5' end of the promoter region have also been shown to have enhancer-like activity (Jordano & Perucho, 1988) and to bind nuclear factors other than Sp1 (Jordano & Perucho, 1988; Hoffman *et al.*, 1990).

1.5.3 Biological Activities of *Ras*: Proliferation, Differentiation and Neoplasia

The *ras* genes and proteins have been implicated in a diverse collection of cellular phenomena, with observations depending on the system under study. *Ras* mutations are observed in many cancers (Bos, 1989), and the *ras* status of the cell affects the way it behaves *in vitro* and *in vivo*.

Microinjection of oncogenic p21^{ras} protein induces proliferation of resting cells (Feramisco *et al.*, 1984; Stacey & Kung, 1984), and the introduction of an activating point mutation into the endogenous H-*ras* gene of immortal untransformed fibroblasts predisposed them to transformation (Finney & Bishop, 1993). Transgenic

mice carrying activated *ras* alleles are predisposed to tumours of the tissues in which the promoter driving expression of the transgene are active (Quaife *et al.*, 1987; Sinn *et al.*, 1987). Transformation may also be induced by overexpression of the c-*onc* form of *ras*, suggesting that the transformation of cells by *ras* is an inappropriate manifestation of a normal cellular process (Chang *et al.*, 1982). Conversely, disruption of activated K-*ras* by HR in colon carcinoma cell lines resulted in a less malignant phenotype as defined by morphology, the ability to grow in soft agar, and growth rates both *in vitro* and as tumours in nude mice (Shirasawa *et al.*, 1993).

Activation of *ras* is observed in response to external mitogenic signals such as epidermal growth factor (Kamata & Feramisco, 1984), and the action of the *ras* proto-oncogene-mediated signalling pathway(s) is essential for the proliferation of NIH3T3 cells in response to serum-stimulation (Mulcahy *et al.*, 1985; Cai *et al.*, 1990). *Ras* proteins also induce processes such as membrane ruffling and pinocytosis in addition to proliferation in resting cells (Bar-Sagi & Feramisco, 1986). In yeast, a deficiency of both the RAS genes was observed to be incompatible with viability (Tatchell *et al.*, 1984). Thus the *ras* signalling pathway plays a vital role in the cellular activation, proliferation and function of normal cells.

In addition to its role in cellular proliferation, which upon deregulation may create a transformed phenotype, *ras* has also been shown to play a role both in promoting and inhibiting differentiation. Pheochromocytoma line PC12 cells differentiate in response to nerve growth factor, observed morphologically by the generation of neurites, but microinjection of inactivating antibodies to $p21^{ras}$ prevents this response (Hagag *et al.*, 1986), as does the transfection of a dominant-negative H-*ras* mutant (Szerberenyi *et al.*, 1990). By contrast, myogenic differentiation is inhibited by the expression of activated *ras* proteins in immortal myoblasts (Olson *et al.*, 1987; Gossett *et al.*, 1988) but not in EC cells (Rudnicki *et al.*, 1989). This difference in the differentiation

response of cells to $p21^{ras}$ indicates that factors other than *ras* status, such as interactions with other oncogenes and the exact lineage of the cell, are also important in its decision to commit to a certain pathway of differentiation.

Ras is now known to be directly involved in several specific eucaryotic developmental processes, including vulval induction in *C. elegans* (Han & Sternberg, 1990; Beitel *et al.*, 1990), eye development of *Drosophila* (Simon *et al.*, 1991), and the progression of preimplantation-stage embryos in the mouse (Ahmad & Naz, 1993; Yamouchi *et al.*, 1994)

1.5.4 Biochemistry of *Ras*

1.5.4.1 The GTP-GDP Cycle

The *ras* genes encode small monomeric molecules which migrate as a 21kd protein. $p21^{ras}$ is a guanine-nucleotide binding protein, specifically GTP and GDP, (Scolnick *et al.*, 1979; Shih *et al.*, 1980; Tamanoi *et al.*, 1984; Temeles *et al.*, 1985) and incorporates a GTPase activity (Gibbs *et al.*, 1984; McGrath *et al.*, 1984; Sweet *et al.*, 1984; Manne *et al.*, 1985; Temeles *et al.*, 1985). This GTPase activity effectively limits the half-life of the GTP-bound form of the molecule, and is attenuated in oncogenic mutants of *ras* such as the val^{12} mutant. The GTP•p21 form of the protein has been shown to be the physiologically "active" form of the protein. *In vitro* GTP•p21 but not GDP•p21 causes activation of *Saccharomyces cerevisiae* adenylyl cyclase (Field *et al.*, 1987; 1988). Microinjection into various cell types of $p21^{ras}$ complexed to non-hydrolysable analogues of GTP but not GDP was found to induce effects similar to those observed in bioassays of activated oncogenic *ras* proteins (Trahey & McCormick, 1987; Satoh *et al.*, 1987).

The $p21^{ras}$ proteins bind guanine nucleotides with very high affinity ($Kd \approx 10^{-11}$ M) (Feuerstein *et al.*, 1987a; 1987b) and thus the spontaneous rate of dissociation of the

nucleotide-p21 complex is very low, $\approx 10^{-5}$ moles s^{-1} mol^{-1} complex (John *et al.*, 1990). This binding is much tighter than is required to saturate the $p21^{ras}$ with nucleotide with the concentrations of GTP ($>10^{-4}$ M) and GDP ($>10^{-5}$ M) found in the cytoplasm, so that $p21^{ras}$ effectively cycles slowly between its GTP- and GDP-bound forms. Because the cellular concentration of GTP is higher than that of GDP dissociation of the bound nucleotide results in its rapid replacement from the cytosol by GTP, and thus dissociation of the nucleotide•p21 complex "activates" the molecule (see above). Certain point mutations change the exchange rate and are activating because they are associated with an unusually high level of GTP•p21 (reviewed in Lowy & Willumsen, 1993). The intrinsic GTPase activity of $p21^{ras}$ is also very low, $k_{cat.GTP}$ of the order of 0.02 min^{-1} (Neal *et al.*, 1988). Therefore in the cell the $p21^{ras}$ molecule effectively binds GTP, where it is in the active form, and subsequently hydrolyses the GTP to GDP and P_i with a half-life of 1-5 hours, in the process inactivating itself. Activated mutant *ras* molecules have a GTP•p21 half-life of 3-9 times longer (Sweet *et al.*, 1984; Gibbs *et al.*, 1984; Lacal *et al.*, 1986; John *et al.*, 1988). The GTP•p21 form of the complex is thought to have several conformations, one of which is an efficient catalyst for hydrolysis of GTP (reviewed in Bourne *et al.*, 1991). This explains its intrinsically slow rate of hydrolysis, although other possibilities such as participation of GAP residues directly in the hydrolysis mechanism have been proposed (Bollag & McCormick, 1991).

Such low intrinsic activities both for the rate of exchange (activation) of *ras* and the rate of hydrolysis (inactivation), but with the rate of hydrolysis exceeding that of exchange, mean that in most cells, *ras* is predominantly in the GDP•p21 state (Gibbs *et al.*, 1990; Satoh *et al.*, 1990a; Downward *et al.*, 1990). Activation of the cell, for example activation of mouse fibroblasts by epidermal growth factor (EGF) or platelet derived growth factor (PDGF) elevates GTP•p21 levels some 2-4-fold (Gibbs *et al.*, 1990; Satoh *et al.*, 1990a, 1990b), or in human T-cells GTP•p21 increases from 5%

total complex to 80% total complex upon stimulation (Downward *et al.,* 1990). To activate the cell by increasing the fraction of *ras* in the GTP•p21 state requires a 100-fold drop in the ratio of exchange rate:GTPase rate. Regulation of this ratio, and hence of the biological activity of $p21^{ras}$, is believed to be achieved through effector proteins which either accelerate the rate of GTP hydrolysis (GTPase activating proteins, GAPs) or nucleotide exchange (guanine nucleotide exchange factors, GEFs), thus moving the equilibrium of GTP•p21: GDP•p21 (reviewed by Bollag & McCormick, 1991; Boguski & McCormick, 1993; Feig, 1994). In mammals the GAPs include $p120^{GAP}$ (Trahey & McCormick, 1987) and neurofibromin (the product of the NF1 neurofibromatosis tumour suppressor gene: Xu *et al,* 1990); the GEFs include mSos1, mSos2 and hSos1, (murine and human homologues of the *Drosophila Son of Sevenless* gene: Bowtell *et al.*, 1992; Chardin *et al.*, 1993), Ras-GRF/$cdc25^{Mm}$ (a homologue of the yeast cdc25 gene: Martegani *et al.*, 1992; Shou *et al.*, 1992; Wei *et al.*, 1992) and *Vav* (Gulbins *et al.*, 1993). The exchange factor SmgGDS (Small GTP-binding protein Guanine nucleotide Dissociation Stimulator: Yamamoto *et al.,* 1990) appears to be specific for K-ras^{B} amongst the *ras* proteins, although it does interact with related proteins such as *Rac*1, *Rho*A and *Rap*1A. These results taken together have lead to the proposal of a model for $p21^{ras}$ action as is shown in figure 1-11.

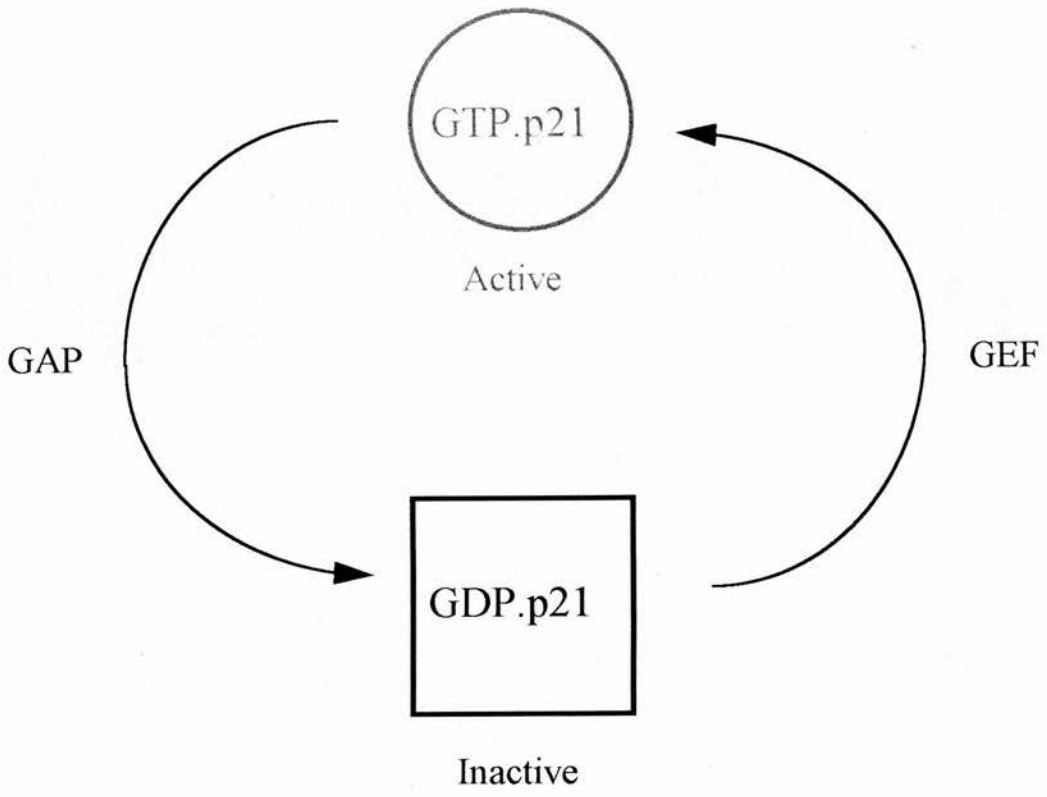
1.5.4.2 Subcellular Localisation: C-Terminal Modification

Ras proteins are synthesised on ribosomes in the cytosol (Shih *et al.*, 1982; Ulsh *et al.*, 1984; Fujiyama & Tamanoi, 1986) but are localised at the inner surface of the plasma membrane (Willingham *et al.*, 1980; Willumsen *et al.*, 1984a; Fujiyama & Tamanoi, 1986). This localisation is essential for the normal activity of *ras* proteins and mutants which remain in the cytoplasm are incapable of transforming cells, even if they carry oncogenic mutations as well (Willumsen *et al.,* 1984a, 1984b), or of

Figure 1-11: The *ras* GTP-GDP Cycle

Regulation of $p21^{ras}$ by guanine nucleotide exchange factors (GEFs), which activate $p21^{ras}$ by catalysing exchange of bound GDP for cytosolic GTP, and GTPase activating proteins (GAPs), which inactivate $p21^{ras}$ by accelerating hydrolysis of bound GTP to GDP. Unbound $p21^{ras}$ is not shown because the binding constant of the molecule for guanine nucleotide is so high that these species are very short-lived and thus presumed to have little biological significance.

Figure 1-11



activating the MAP kinase cascade in fibroblasts (see below; Leevers & Marshall, 1992). Membrane localisation is mediated by several post-translational modifications of the protein which increase the hydrophobic nature of the molecule rather than the presence of hydrophobic sequences within the primary structure (reviewed by Newman & Magee, 1993; Glomset & Farnsworth, 1994). Attachment to the membrane requires the prenylation of $p21^{ras}$; a farnesyl (C15) moiety is added to the sulphydryl group of $cys^{186/185}$ (the C of the CAAX motif) by a thioester link (Hancock *et al.*, 1989, 1991). In the case of H-*ras*, N-*ras* and K-ras^{A} the cysteine residues slightly N-terminal to cys^{186} are also palmitoylated, but K-ras^{B}, which lacks the necessary cysteine residues, has a polylysine motif present in the heterogeneous C-terminal region of the protein which is involved in membrane targeting (Hancock *et al.*, 1990; Jackson *et al.*, 1994). The basic residues of the polylysine region may enhance the membrane targeting of the molecule by permitting charged interactions with negatively charged head groups of acidic membrane phosphoglycerides (Glomset & Farnsworth., 1994), and this domain has recently been shown to be required by K-ras^{B} in malignant transformation (Jackson *et al.*, 1994). Following the isoprenylation reaction, the terminal three amino acids of the CAAX motif are removed by proteolysis and the C-terminal carboxyl group is methylesterified, and these steps are required for fully efficient membrane localisation (Hancock *et al.*, 1991).

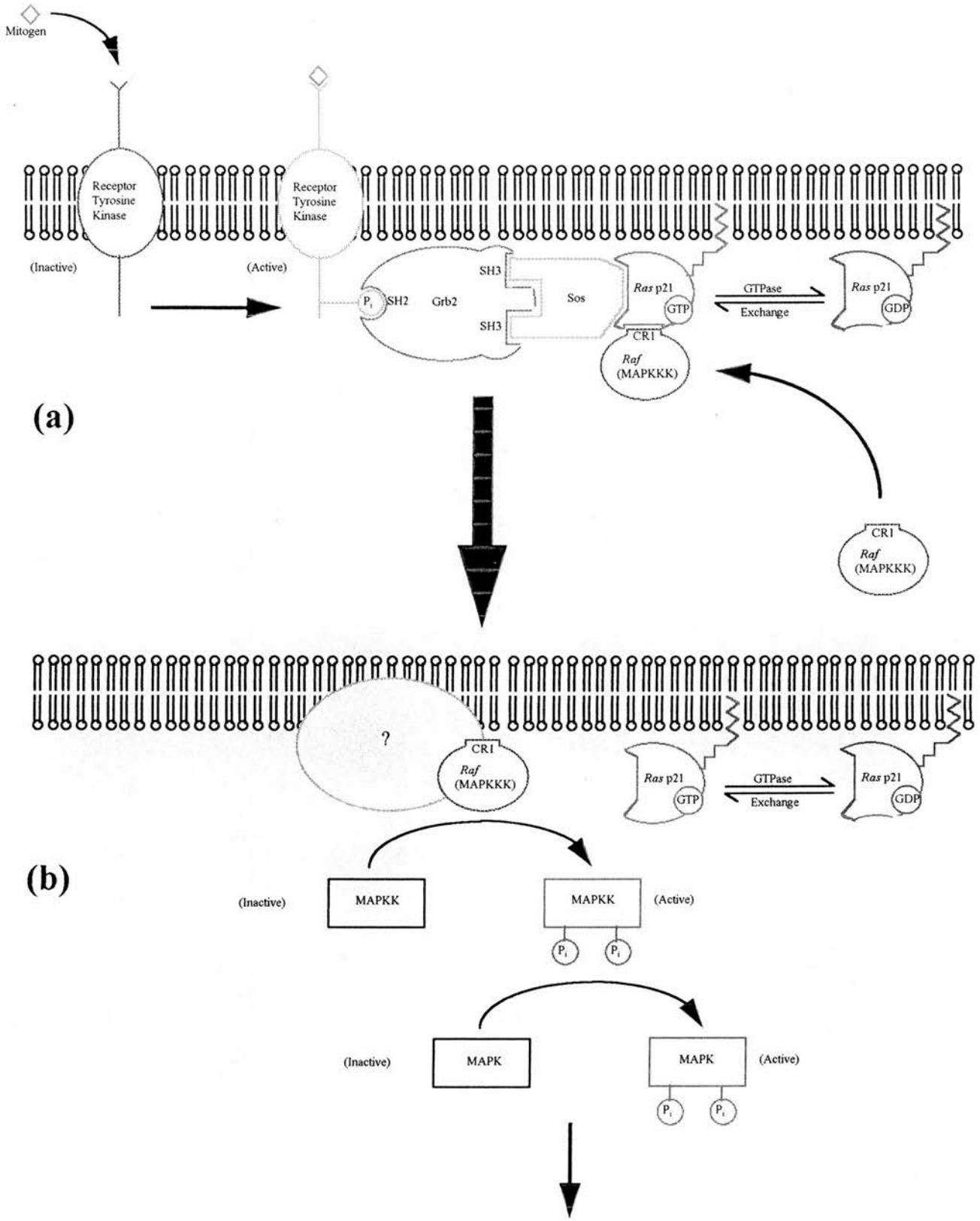
1.5.4.3 How $p21^{ras}$ Activates the Cell

Results obtained in the last two years have now produced a very clear picture of the way in which *ras* is able to activate the cell (figure 1-12; for reviews see McCormick, 1993; Feig, 1993; Moodie & Wolfman, 1994; Hall, 1994; Avruch *et al*, 1994). Sos has recently been shown to be an activator of $p21^{ras}$ in response to cellular activation, for example by EGF activating the EGF receptor tyrosine kinase. Interestingly, *Sos*

Figure 1-12: *Ras* Activation of the Cell

A mitogen, such as EGF, binds to its receptor tyrosine kinase (RTK), which is activated and autophosphorylates. The phosphorylated cytoplasmic domain of the RTK may then bind with the SH2 domain of an adapter protein such as Grb2. The SH3 domains of Grb2 interact with Sos, recruiting it to the cell membrane. There is therefore a high effective local concentration of Sos at the plasma membrane. Sos is a GEF for $p21^{ras}$, and thus *ras* is activated by exchange of GDP for GTP. *ras* then recruits *raf* to the membrane, interacting through its CR1 region. Once *raf* is membrane-localised, it may interact with other proteins (?) and activate the MAP kinase cascade, leading to upregulation of various cell effectors such as transcription factors.

Figure 1-12



- **Transcription Factors:** :***Elk*-1, c-*jun*, c-*fos*, c-*myc*, ATF-2**
- **Other Ser/Thr Kinases** :**Rsk, MAPKAP-K2**
- **Inflammatory Mediators** :**$cPLA_2$**
- **Cytoskeleton**

requires prenylation of *ras* to efficiently promote guanine nucleotide exchange (Porfiri *et al.*, 1994). Guanine nucleotide exchange is promoted through adapter proteins such as Grb2 (the mammalian homologue of Sem-5 of *C. elegans* and drk of *Drosophila*), so called because it has both SH2 and SH3 "Src homology domains". The SH2 and SH3 domains provide a link between the activated (autophosphorylated) receptor and *Sos*, respectively. Once converted to the GTP• p21 state by *Sos* in response to the extracellular mitogenic signal, *ras* may initiate a kinase cascade. *Ras* recruits another member of the GTPase superfamily to the complex, namely *raf*. The localisation of *raf* at the plasma membrane allows activation by *raf* of the MAPK (mitogen activated protein kinase) cascade. This cascade has pleiotropic affects on the cell, influencing cytoplasmic organisation, cytoplasmic phospholipase A_2, other serine/threonine kinases such as *Rsk* and MAPKAP-K2 and transcription factors such as c-*fos*, c-*myc*, *c-jun*, ATF-2 and Elk-1.

1.5.5 Expression of *Ras*

Several workers have investigated the expression of *ras* genes in both normal and dysplastic tissue, at the RNA and protein levels. In tumours the expression of *ras* genes is elevated with respect to control tissues in about 50% of cases when RNA transcripts are assayed, and similar results have been obtained with immunoblot assay of p21 *ras* proteins (reviewed by Barbacid, 1987). Mutant alleles of K-*ras* have been observed to be preferentially expressed over normal alleles in about two thirds of colorectal carcinomas examined (Kotsinas *et al.*, 1993), and this may provide a selective growth advantage to the tumour.

In normal proliferating liver, *ras* was observed to be upregulated several-fold (Goyette *et al.*, 1983). Systematic studies of *ras* expression in the normal tissues of several mammalian species have also been undertaken, to clarify the role of the genes

in development and in normal cell function. The *ras* genes appear to be expressed ubiquitously, but with variations in levels between different genes and cell types. In the mouse, the H-*ras* gene is expressed ubiquitously at approximately constant levels in both the embryo and extra-embryonic tissues, with higher levels observed in bone, kidney, brain and skin post-natally (Muller *et al.*, 1982; Slamon & Cline, 1984; Leon *et al.*, 1987). K-*ras* is also expressed ubiquitously at approximately constant levels until about day 16 of development, when levels decrease to about 25-50% relative to those observed in midgestation embryos (Muller *et al.*, 1983). In the adult K-*ras* is most prevalently expressed in the gut, lung and thymus, and was less represented in tissues such as skin, liver, skeletal muscle, ovary and spleen (Leon *et al.*, 1987). K-*ras* transcript and protein levels are modulated in some tissues with age (Tanaka *et al.*, 1986; Leon *et al.*, 1987). It is of note that H-*ras* appeared to be most highly represented in the skin because H-*ras*, and not K-*ras* or N-*ras*, is a target for oncogenic mutations in the skin of mice with chemically induced tumours (Balmain *et al.*, 1984).

N-*ras* was expressed at highest levels in thymus and testis, and lowest in liver and kidney, but again appears to be represented to some extent in all tissues (Leon *et al.*, 1987).

Consistent with the observation that *ras* genes are ubiquitously expressed, they have also been observed in very early-stage embryos and in cells of the germline. In male germ cells the 3 *ras* genes show distinct levels of expression: K-*ras* mRNA being elevated in pachytene cells and N-*ras* in early round spermatids. In contrast, H-*ras* expression is constant during spermatogenesis (Sorrentino *et al.*, 1988). In preimplantation-stage mouse embryos all three *ras* genes are expressed (Pal *et al.*, 1993). Modulation of expression of H-*ras* and K-*ras* was not observed during differentiation of F9 EC cells (Lockett & Sleigh, 1987).

During rat development, analysis with a polyspecific anti-p21 *ras* antibody showed very wide protein expression of p21 *ras*, but in earlier stages of development staining was weaker in some extraembryonic tissues (Brewer & Brown, 1992). Despite the association of *ras* with proliferation of tissues, examination of tissues with anti-p21^{ras} monoclonal antibodies has frequently revealed the highest levels of expression in terminally differentiated tissues, such as adult (and embryonic) brain and epithelial cells of the endocrine glands, or non-dividing tissue such as heart muscle (Spandidos & Dimitrov, 1985; Tanaka *et al.*, 1986; Furth *et al.*, 1987; Chesa *et al.*, 1987).

CHAPTER 2-

THE FIDELITY OF GENE TARGETING

The Fidelity of Gene Targeting

2.1 Introduction

Much of this thesis (chapters 4 and 5) concerns itself with hit and run gene targeting. Because this technique was designed for the purpose of introducing small mutations into the genome, a high degree of precision is required to produce the desired results, arguably not the case for knockout-type experiments. As described in the introduction, section 1.3.3.3, relatively little is known about the fidelity of gene targeting events. Some early results suggested that the recombination process may be prone to errors (Thomas & Capecchi, 1986; Doetschman *et al.*, 1988; Thompson *et al.*, 1989; Brinster *et al.*, 1989). Although intra-chromosomal recombination is known to be very accurate (Stachelek & Liskay, 1988), it is not possible to extrapolate this assumption to gene targeting because some of the enzymatic systems which accomplish this HR event are likely to be different, perhaps including some of those involved in the extra-chromosomal recombination mechanism. Because intra-chromosomal recombination seems to proceed with fidelity, it is likely that undesired mutations co-introduced in a hit and run experiment would be introduced with the initial insertion event ("hit" stage).

The only precise study of the fidelity of gene targeting, by chemical mismatch analysis, suggested that gene targeting is in fact quite accurate (Zheng *et al.*, 1991). Zheng and co-workers used a knockout of the *hprt* locus in mouse and human cells as their test system, and then analysed independent G418- and 6-TG-resistant clones for the presence of mutations suggesting errors in the targeting process. Their analysis relied on the mutation being relatively close to the site of integration (within approximately 700bp of the insertion site), and was designed primarily to detect small mutations within this region.

The experiments described in this chapter were intended to further the understanding of the fidelity of insertion events in the context of a hit and run experiment.

2.2 Experimental Design

The *hprt*$^{b\text{-}m3}$ mutation (Thompson *et al* 1989), present in the E14TG2a ES cell line (Hooper *et al.*, 1987), was chosen as a convenient system in which to study errors in recombination for reasons already outlined in section 1.4. To this end, a targeting event was designed, such that accurate recombination would restore function in the *hprt*$^{b\text{-}m3}$ gene. Many errors in recombination would therefore be observable at the level of the phenotype. The structure of the *hprt*$^{b\text{-}m3}$ locus is shown in figure 2-1, where it may be seen that the mutation rendering the E14TG2a line resistant to 6-TG is a deletion encompassing the promoter region and exons 1 and 2. There is thus no significant possibility of background reversion to wild-type, as there might be with a point mutation. The experimental design is shown in figure 2-2. A vector incorporating sufficient upstream control sequences to drive the repaired *hprt* gene (Melton *et al.*, 1986) plus a fragment spanning exons 1 and 2 is therefore capable of repairing the *hprt*$^{b\text{-}m3}$ gene (Doetschman *et al.*, 1987; Thompson *et al.*, 1989). A *neo* gene was also included in the construct, and was used to initially select the clones on the basis of their resistance to G418 rather than to HAT medium. Clones were then studied both by Southern analysis and HAT selection. This allowed the comparison of genotype with phenotype; a lack of correlation between the results of the Southern analysis and the ability of the cell to survive or otherwise in HAT medium might signify a clone which has failed to recombine correctly. The level of *hprt* enzyme activity was also assessed in HAT-resistant (HAT^r) clones, in case small mutations introduced during the insertion event had deleterious effects on the protein but did not completely ablate gene function, resulting in an *hprt* allele with reduced activity compared to wild-type.

Figure 2-1: The Murine *hprt* Locus

The structures of the wild-type (HAT^r/6-TG^s) *hprt* locus and the mutant (HAT^s/6-TG^r) $hprt^{b\text{-}m3}$ locus, present in the E14TG2a murine ES cell line. Exons are marked as grey blocks, with transcribed but untranslated regions hatched. Intronic sequence is shown as a black line. The sizes of different restriction fragments observable with exon-specific probes are shown in kb. Selected restriction enzyme sites marked: B, *Bam*H I; E, *Eco*R I; X, *Xho* I.

Figure 2-1

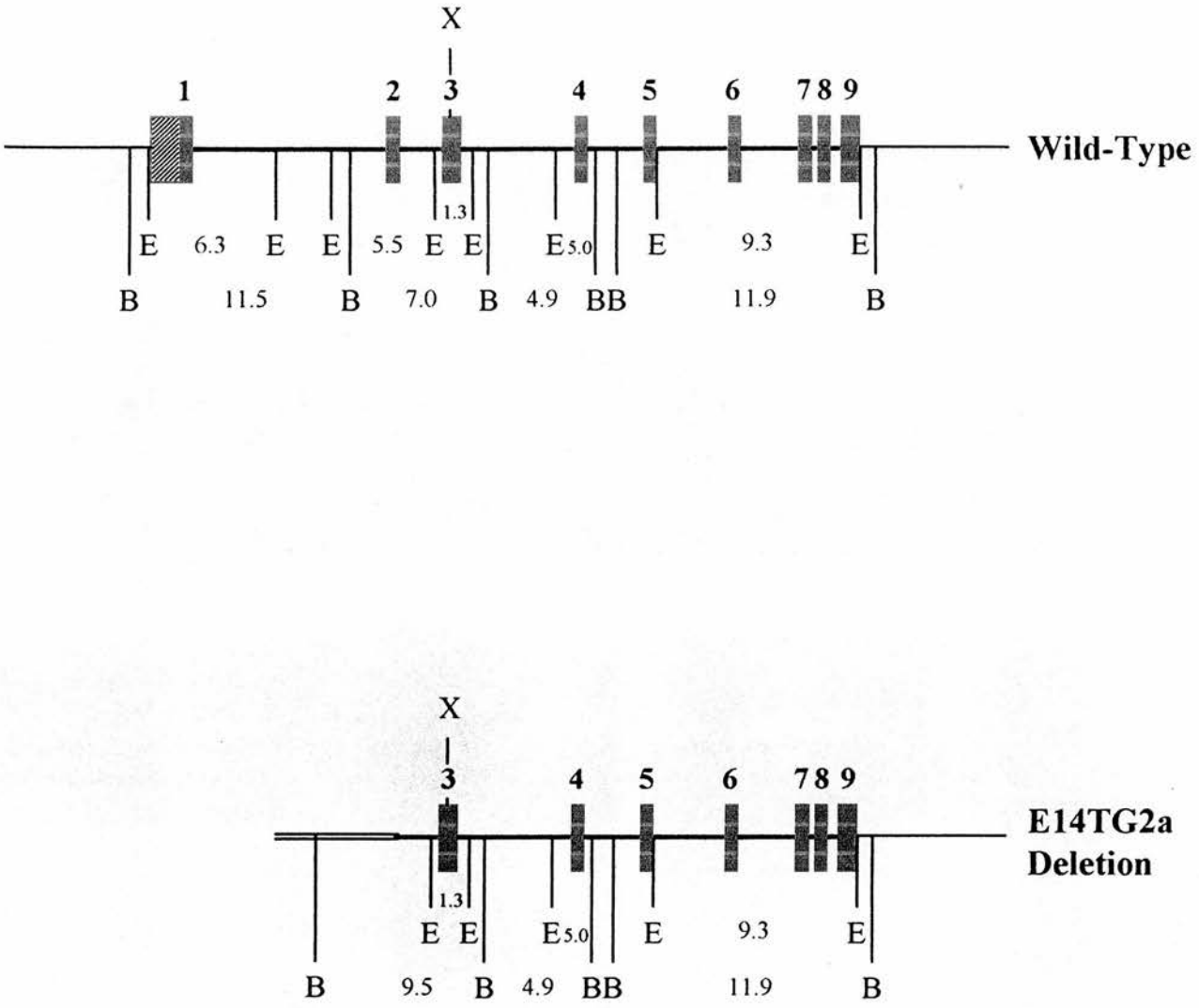

X
1
2
3
4
5
6
7 8 9
Wild-Type
1.3
E
6.3
E
E
5.5
E
E
E5.0
E
9.3
E
B
11.5
B
7.0
B
4.9
BB
11.9
B
X
3
4
5
6
7 8 9
E14TG2a
Deletion
1.3
E
E
E5.0
E
9.3
E
B
9.5
B
4.9
BB
11.9
B

Figure 2-2: A Strategy for Examining the Fidelity of Gene Targeting

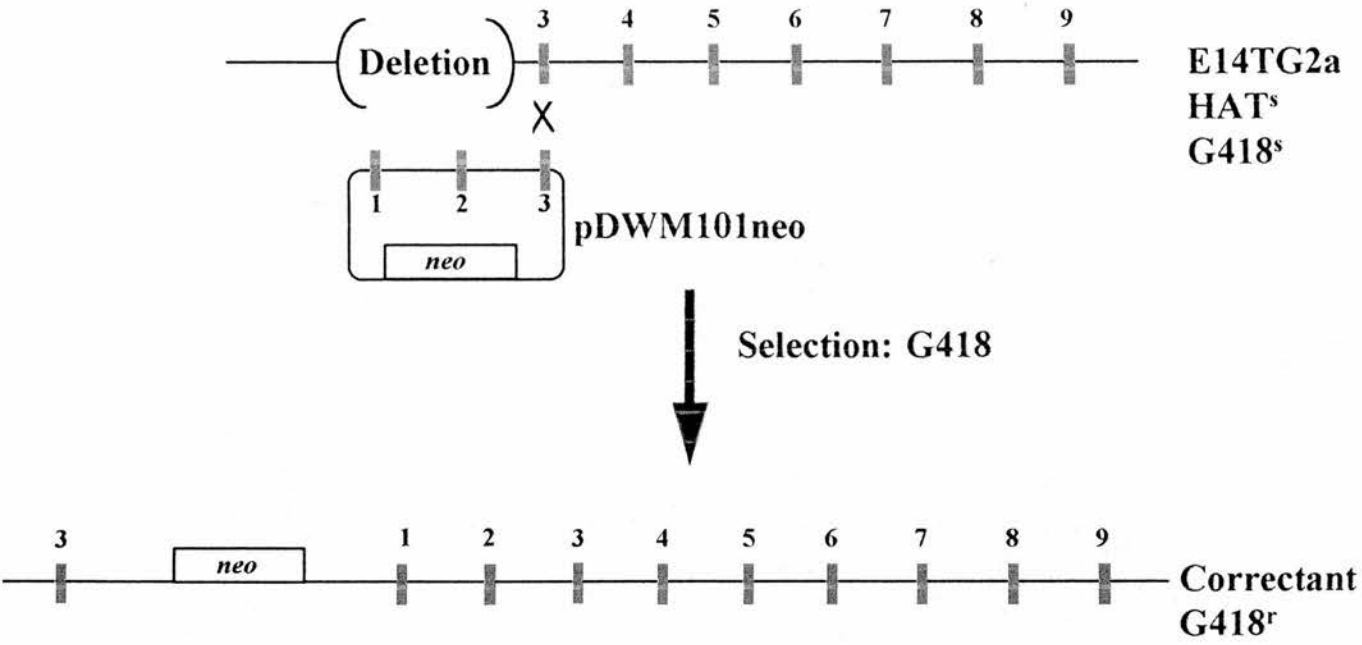
The *hprt*$^{b\text{-}m3}$ locus is targeted with a correction vector designed to convert phenotype from HAT^{s}/6-TG^{r} to HAT^{r}/6-TG^{s}. Clones are isolated by virtue of their resistance to G418, rather than HAT, and therefore selection is not dependent upon restoration of function. The correction vector is based upon pDWM101 (Thompson *et al.*, 1989) and is shown in figure 2-3. Vector pDWM101 contains 9kb of the murine *hprt* gene, between 2.3kb and 4.2kb of which is homologous to the mutant *hprt*$^{b\text{-}m3}$ allele (depending upon the exact but unknown location of the 3' endpoint of the deletion encompassing exons 1 and 2 in E14TG2a). The vector has the same organisation as the wild-type gene except that intron 1 is reduced from 10.8kb to 4.1kb. The homologous sequence is derived from a mouse myeloma library (Melton *et al.*, 1984) and as such is considered not isogenic with the 129/ola sequence of the target locus.

In experiment HPRT/I, a single plate of electroporated cells was selected in HAT medium, to generate targeted clones selected at the level of the phenotype as positive controls for those generated by selection in G418; These clones were designated HI/H*x*.

After isolation, clones are then screened in duplicate, at the level of the genotype by Southern analysis and at the level of phenotype by HAT selection. Clones which appear targeted by Southern analysis but which are HAT^{s} may indicate errors in the HR process.

The *hprt* enzyme activity of a cohort of targeted clones was also analysed to check for anomalous levels of activity as a consequence of point mutations which do not completely ablate gene function.

Figure 2-2



ANALYSIS OF CLONES: (1) Southern: Genotype

(2) HAT Selection: Phenotype

Are there any clones of the correctant structure by genotype that are HAT^s, and that may thus be failed recombinants?

HPRT Enzyme Activity analysis of HAT^r clones to ascertain the likelihood of small mutations failing to completely ablate gene function

Such an analysis may detect mutations not obvious from the study of Zheng *et al.* These include point mutations, small rearrangements and deletions occurring at some distance from the site of integration of the vector. There is some evidence that such mutations occur during insertion events (Doetschman *et al.*, 1987; Thompson *et al.*, 1989). Such errors may be difficult to detect by Southern analysis, and unless they occurred close to the linearisation site would not have been detectable by mismatch analysis of PCR products from around this site. Unlike the Zheng study, the clones described in this chapter were not pre-selected as being of the predicted targeted structure before analysis.

2.3 Targeting Vector pDWM101neo

The targeting vector pDWM101neo was constructed as a variant of pDWM101 (Thompson *et al.*, 1989), provided as a gift by Dr David W. Melton of the Department of Molecular Biology, Edinburgh. A *neo* gene under the control of the PGK promoter and PGK polyadenylation sequence was excised from the vector pSPGKneo (obtained from Hein te Riele of Amsterdam; shown in figure 7-1 [a]) as a 1.5kb *Bgl* II fragment. This fragment was cloned into the *Sal* I site of pDWM101 by blunt-ended ligation to create pDWM101neo. pDWM101neo is shown in figure 2-3, together with the predicted structure of the correctant *hprt* allele. The vector was linearised at the *Xho* I site in exon 3.

2.4 Results

In two separate electroporations, vector pDWM101neo was transfected into E14TG2a ES cells as described, and clones were selected in 300μg/ml G418 and expanded to approximately $1x10^6$ cells (a confluent single well of a 24-well tissue culture plate). They were then split into two, with half being used for Southern analysis and the remainder tested for the ability to survive in HAT medium. The *hprt*

Figure 2-3: *hprt* Correction Vector pDWM101neo

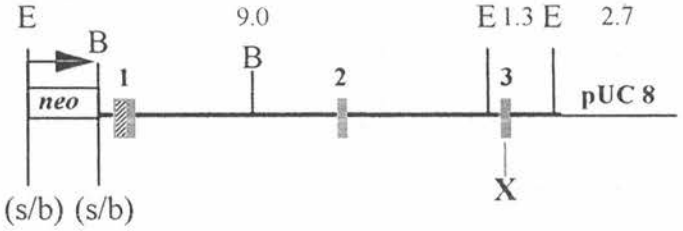
a) Vector pDWM101neo, derived from vector pDWM101 (Thompson *et al.*, 1989). The vector includes exons 1-3 inclusive and 650 bp upstream sequence, shown as grey or hatched blocks respectively, which is sufficient to drive corrected *hprt* gene expression (Melton *et al.*, 1986). Intronic sequence is shown as a thick black line; bacterial sequence as a thin black line. A PGK-*neo* cassette was cloned into the vector, orientation as shown. The vector is linearised at the unique *Xho* I site in exon 3. Selected restriction enzyme sites are shown: B, *Bam*H I; E, *Eco*R I; X, *Xho* I; s and b, *Sal* I and *Bgl* II sites lost upon blunt-ended cloning-in of the *neo* construct.

b) Restriction enzyme digestion of vector pDWM101neo. Lanes: 1, Uncut (*Sal* I); 2, *Eco*R I; 3, DNA molecular size markers.

c) Predicted structure of the corrected $hprt^{b\text{-}m3}$ allele. *Bam*H I restriction enzyme sites are shown, with fragment sizes shown in kb.

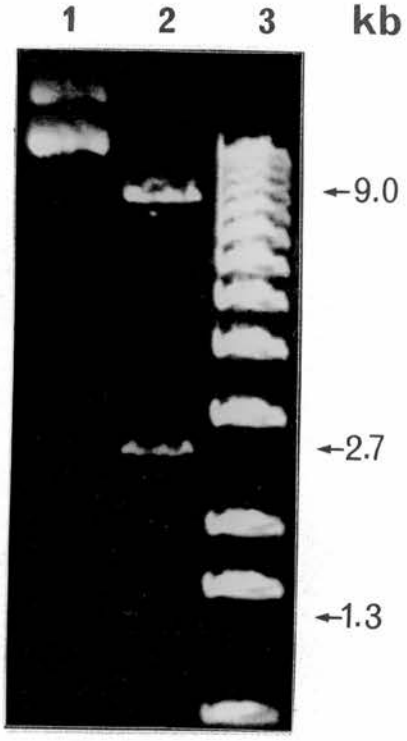
Figure 2-3

(a)

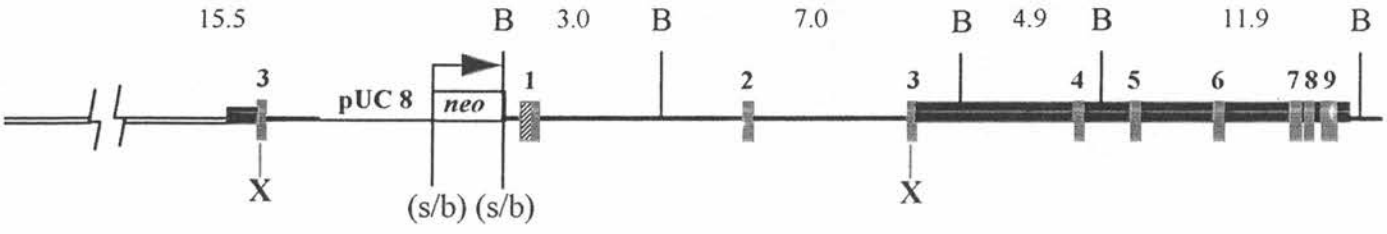


pDWM101neo

(b)



(c)



Correctant

cDNA plasmid pHPT5 (obtained from Dr David W. Melton; Konecki *et al.*, 1982) was used as a source of probes for Southern analysis. Either a 170 bp *Xho* I/*Hinc* II fragment largely comprising exon 3 or a 1.3kb fragment comprising the entire *hprt* cDNA was used as a probe. Table 2-1 shows the results of these experiments. Figure 2-4 shows the results of Southern analysis of some HAT-resistant clones and some $G418^r$, HAT^s clones, using exon 3 of the *hprt* gene as a probe.

2.4.1 Accuracy of HAT Selection Data

From table 2-1 it may be seen that the frequencies of targeting measured by phenotype (i.e. HAT resistance) are higher than those suggested in the Southern analysis. This was more noticeable in the second experiment. Especially in experiment HPRT/II, it was observed that clonal growth rates in HAT varied widely. This was unexpected, and suggested that many clones may be mixed rather than clonal, and that the variable rate of growth in HAT medium was related to the initial proportion of HAT^s contaminating cells. To test this hypothesis, HAT^r clones obtained in both experiments HPRT/I and HPRT/II were re-tested for their ability to survive in HAT medium. The results of this experiment are shown in table 2-2. Clearly, some clones were originally incorrectly assigned as HAT^r. Two tested "clones" were seen to be capable of surviving in both HAT and 6-TG. One of these (HI/122) was a targeted clone from the G418 selection of experiment HPRT/I and the other (HI/H5) was a clone originally selected in HAT medium. To differentiate between the possibilities that either these two clones might be failed recombinants which maintained a sufficient level of *hprt* activity to survive, albeit poorly, in HAT as well as 6-TG, or were mixed clones, they were split into two and selected in both selective media for 10-14 days. Southern analysis and HPRT enzyme activity analysis (see below) were then performed on the selected cell populations, and it was concluded that these were mixed clones. Figure 2-5 shows Southern analysis of HAT^r and $6\text{-}TG^r$ subclones from

Figure 2-4: Southern Analysis of Clones Obtained with Vector pDWM101neo

The figure shows the results of Southern analysis on E14TG2a clones incorporating pDWM101neo. *Bam*H I digest, probed with exon 3 of the murine *hprt* gene. The 9.5kb E14TG2a band, and the 15.5kb and 7.0kb bands diagnostic for the correctant structure are shown.

Lanes 1-5: Targeted clones carrying a corrected *hprt* gene, and $G418^r/HAT^r$; lanes 6-16: Untargeted clones carrying a random integration of vector pDWM101neo, and $G418^r/HAT^s$. Some clones, such as those in lanes 6 and 14, show an extra band of an unpredicted size. These clones are interpreted as ones where recircularisation of the vector has probably occurred prior to vector integration at a random locus. The vector copy of exon 3 (the linearisation site) is thus protected from exonucleolytic degradation at the ends of the vector prior to recombination and will therefore give a signal upon probing a restriction digestion with an exon 3 probe.

Table 2-1: Clones Obtained with Vector pDWM101neo

The table summarises the genotypic (Southern) and phenotypic (HAT selection) results for the clones obtained with pDWM101neo.

Figure 2-4 Southern Analysis of Clones Targeted and Untargeted with Vector pDWM101neo

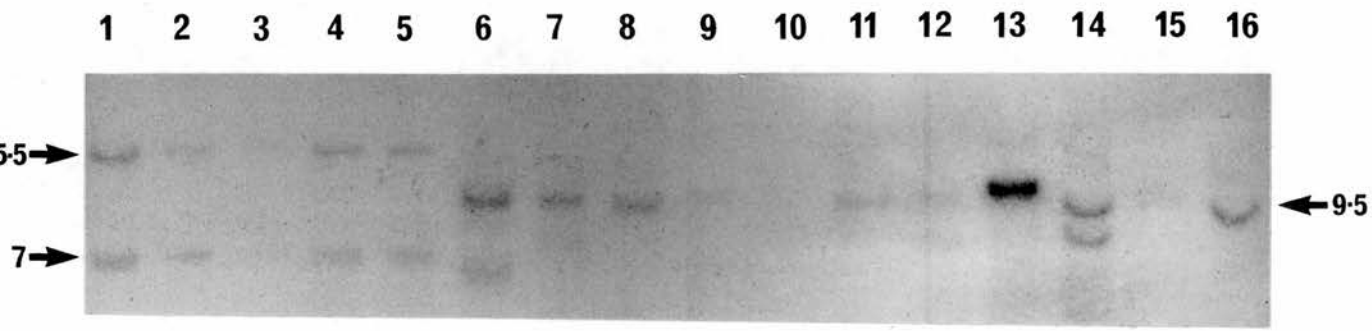


Table 2-1. Clones Obtained with Vector pDWM101neo

Experiment	# G418^r (1)	# G418^r + HAT^r (2)	# Screened by Southern	# Not Targeted by Southern	# Targeted by Southern
HPRT/I	167	5 3%	96[3]	93	3
HPRT/II	284	60 21%	75[4]	73	2

Notes:

1) Clones selected in 300μg/ml G418.

2) The number of clones originally selected in G418 initially scored as positive for the ability to survive in HAT medium.

3) Includes those clones assessed as G418^r/HAT^r in column 3; 2 clones, HI/36 and HI/56, appeared to be untargeted when studied by Southern analysis.

4) Includes 18 of those clones assessed as G418^r/HAT^r; 16 clones appeared to be untargeted when studied by Southern analysis.

Table 2-2: Analysis of Targeted Clones

Unexpected results described in the text (section 2.4.1) where it was observed that some clones initially assigned as HAT^r appeared to be untargeted when studied by Southern analysis suggested that many clones from this experiment may in fact be mixed. The results of additional selection experiments carried out on these "nominally targeted" clones to test this hypothesis are summarised in the table. The second column shows whether the clone was derived from an initial selection in HAT medium or in G418. The next column shows the number of clones tested. Not all clones were HAT^r upon re-selection. For clones from experiment HPRT/I, it was found that clone HI/56 was in fact HAT^s/6-TG^r (i.e. mutant phenotype), and that clones HI/H5 (initial selection HAT) and HI/122 (initial selection G418) appeared to be HAT^r/6-TG^r. These latter two clones were then accordingly re-tested at the level of the genotype as shown in figure 2-5.

Figure 2-5: Mixed HAT^r/HAT^s Clones

The figure shows the results of Southern analysis on subclones obtained from clones HI/H5 and HI/122 by re-submitting the clones to either HAT or 6-TG selection for a period of 10-14 days. *Bam*H I digest, using the murine *hprt* cDNA as a probe. It may be seen that subclones obtained by 6-TG selection have the 9.5kb band (present in E14TG2a) diagnostic of the deletion spanning the promoter and exons 1 and 2. These are therefore clones carrying a random integration of vector pDWM101neo and are $hprt^-$. The subclones subjected to HAT selection for the same time lack this band, but show the 7.0kb band diagnostic of a targeted correction and are therefore $hprt^+$.

Because a full-length cDNA clone was used as a probe in this experiment, other bands are also visible. The 15.5kb correctant band and the murine *hprt* pseudogene are visible as a single large intense band at the top of the blot, above the 9.5kb $hprt^{b\text{-}m3}$ deletion band. The small band common to all lanes below the 7.0kb correctant band is the 4.9kb band encompassing exon 4 (see figures 2-1 and 2-3).

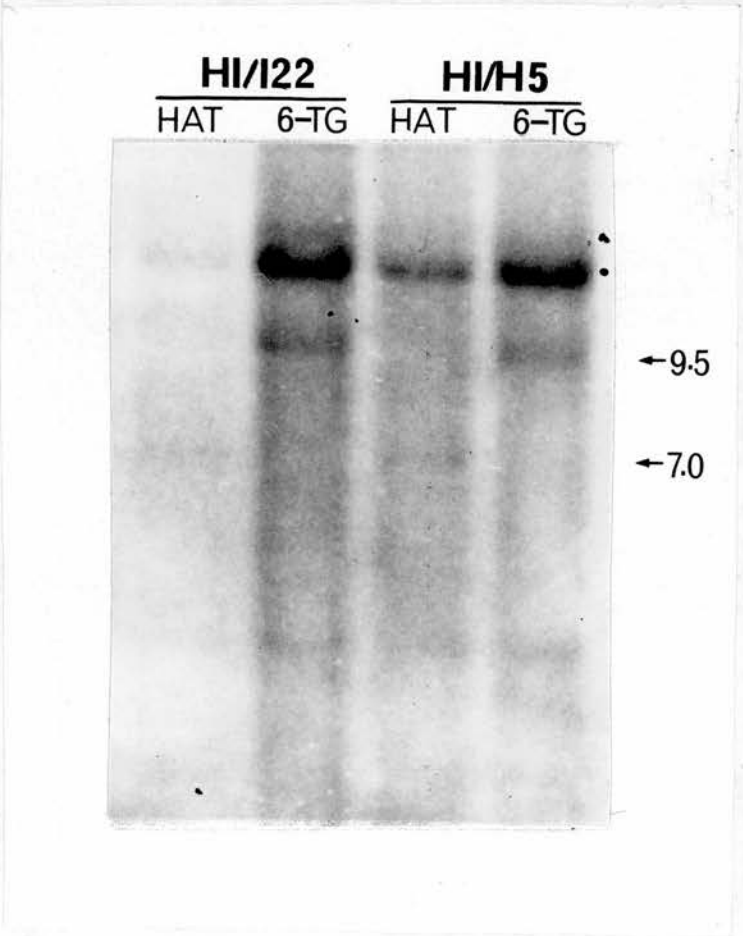
Table 2-2 Phenotypic Analysis of Nominally Targeted Clones

Experiment	Initial Selection of Clones	# HAT^r Clones Tested	# HAT^r on Re-Selection	# Tested for 6-TG Resistance	# 6-TGr on Re-Selection
HPRT/I	G418	4	2	2	2 (1)
	HAT	1	1	1	1
HPRT /II	G418	26	7 27%	ND	ND

Note:

(1) These two clones comprise one found to be HAT^r, but which grew slowly in HAT (HI/122), and one HAT^s (HI/56).

Figure 2-5 Southern Analysis of HAT^r and 6-TG^r Subpopulations of Clones HI/H5 and HI/122.



HI/H5 and HI/122, where it is clear that each is in fact a mixture of two clones, only one of which is targeted, and a second round of HAT selection was required to completely purify the clone. Table 2-3 and figure 2-6 (see below) also include the *hprt* enzyme activity data for both HAT and 6-TG-selected subpopulations of HI/122 and HI/H5, where it may be seen that they have significantly different levels of *hprt* activity depending on which subpopulation has been selected.

2.4.2 Enzyme Activity Analysis of Clones Targeted with pDWM101neo

No clones which appeared to be targeted by Southern analysis were found to be HAT^s. This suggests that failed recombinants are relatively rare, but it is likely that in some cases small mutations in a gene would not completely ablate function, but result in a reduced level of activity. A clone with a residual level of *hprt* activity may still be capable of surviving in HAT medium. To examine this possibility a cohort of the HAT^r clones from both experiments were analysed for HPRT enzyme activity.

In addition to the $G418^r/HAT^r$ clones tested, the *hprt* activity of E14 and E14TG2a was also measured as positive and negative controls. To examine the possibility that targeting and selection procedures may alter *hprt* activity relative to wild-type as an artefact of culture, some E14 clones which have integrated an irrelevant construct at random loci, and some $G418^r/HAT^s$ clones from experiment HPRT/I, were also analysed.

The results of this experiment are shown in table 2-3 and figure 2-6.

Table 2-3. HPRT Enzyme Activity of Targeted and Untargeted Cells

Notes:

1) An *hprt* enzyme assay was always run with E14 (*hprt*$^+$) and E14TG2a (*hprt*$^-$) cells present as positive and negative controls, respectively. The *hprt* enzyme activity was calculated by multiple linear regression for all the clones present in a particular assay based on a zero value and three different protein concentrations (1, 5 and 10μg), corrected for background counts with an EDTA-poisoned blank. Calculations were carried out with the "Minitab" version 9 statistical package or the Jandel Scientific "Sigmastat" statistical package.

2) Standard deviation for the activity coefficient calculated from the regression is shown, and is given as an error bar in figure 2-6.

3) P value gives the significance of the recorded activity coefficient. Only HAT^r clones have a significant level of enzyme activity ($P \leq 0.05$).

4) Clones HI/H2, HI/H4 and HI/H5 were initially selected in HAT medium as controls for clones isolated by virtue of their resistance to G418 in experiment HPRT/I. HI/H5 had appeared to be both HAT^r and 6-TG^r (see table 2-2, line 2). Clones HI/8, HI/9, HI/36, HI/46, and HI/150 are HAT^s, untargeted negative controls. Clone HI/56 originally appeared to be targeted but was subsequently shown not to be (table 2-2, row 1, final column). Clone HI/122 appeared to be both HAT^r and 6-TG^r, and was shown to be a mixed clone like HI/H5 (table 2-2 and figure 2-5). Both these clones appear in the table twice, defined by the selection to which each pure subclone is resistant. Clones HI/122 (HAT^r) and HI/149 are targeted clones obtained from experiment HPRT/I by G418 selection (table 2-1, row 1). Clones HII/252, HII/349, HII/361 and HII/464 are HAT^r clones obtained by G418 selection in experiment HPRT/II (table 2-1, row 2), and shown to be HAT^r (table 2-2, row 3).

Table 2-3. HPRT Enzyme Activity of Targeted and Untargeted Cells

Type	Clone	HAT^r	HPRT Activity (cpm/μg Protein)	Std. Dev.	P Value
WT	E14	✓	1402	124	<0.001
WT	E14	✓	1160	73	<0.001
WT	E14	✓	1328	55	<0.001
WT	E14	✓	498	95	0.002
WT	P66 (E14)	✓	946	135	<0.001
WT	W38 (E14)	✓	1069	120	<0.001
WT	W44 (E14)	✓	1249	165	<0.001
WT	W55 (E14)	✓	964	180	<0.001
WT	W65 (E14)	✓	1311	184	<0.001
WT	W76 (E14)	✓	1045	150	<0.001
WT	W85 (E14)	✓	773	163	<0.001
Mutant	E14TG2a	×	-201	169	0.24
Mutant	E14TG2a	×	86	73	0.25
Mutant	E14TG2a	×	-64	52	0.25
Mutant	E14TG2a	×	-26	95	0.79
HAT^r	HI/H2	✓	669	153	<0.001
HAT^r	HI/H4	✓	734	157	<0.001
HAT^r	HI/H5 (HAT^r)	✓	233	73	0.005
$G418^r/HAT^s$	HI/8	×	-214	153	0.17
$G418^r/HAT^s$	HI/9	×	-131	149	0.39
$G418^r/HAT^s$	HI/36	×	-201	153	0.20
$G418^r/HAT^s$	HI/46	×	-169	157	0.29
$G418^r/HAT^s$	HI/56	×	-54	73	0.47
$G418^r/HAT^s$	HI/122 (6-TG^r)	×	115	73	0.13
$G418^r/HAT^s$	HI/150	×	-136	155	0.39
$G418^r/HAT^s$	HI/H5 (6-TG^r)	×	60	73	0.42
$G418^r/HAT^r$	HI/122 (HAT^r)	✓	387	73	<0.001
$G418^r/HAT^r$	HI/149	✓	804	142	<0.001
$G418^r/HAT^r$	HI/149	✓	700	73	<0.001
$G418^r/HAT^r$	HII/252	✓	326	52	<0.001
$G418^r/HAT^r$	HII/349	✓	299	95	0.020
$G418^r/HAT^r$	HII/361	✓	364	52	<0.001
$G418^r/HAT^r$	HII/464	✓	1193	5	<0.001

Figure 2-6 HPRT Enzyme Activities of Targeted ES Cell Clones

The figure shows the *hprt* enzyme activity for each clone tested (see table 2-3 for values).

Multiple points in a single column indicate that more than one value has been obtained for a particular type of clone.

Symbols:

1) Inverted solid triangles. E14 cells (4 values).

2) Upright solid triangles. E14 cells which have integrated an irrelevant vector at a random locus (7 values). Reading from left to right, clones: P66, W38, W44, W55, W65, W76, W85.

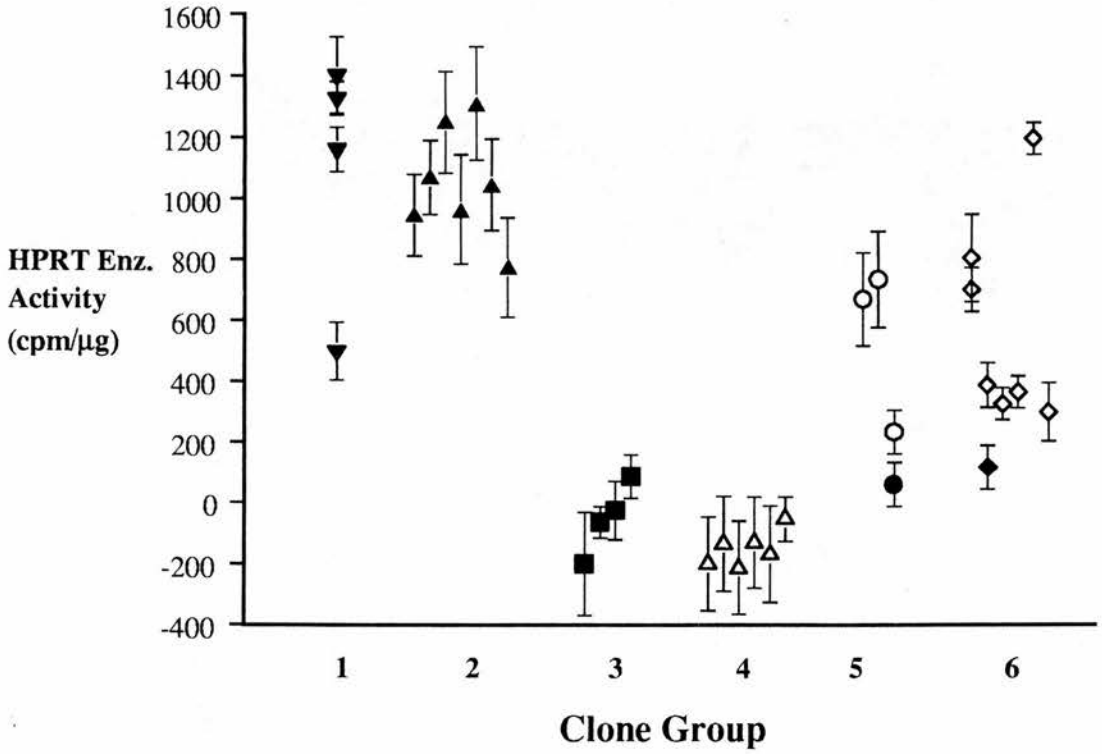
3) Solid squares. E14TG2a cells (4 values).

4) Open upright triangles. E14TG2a cells transfected with pDWM101neo but not targeted at the *hprt* locus (6 values). reading from left to right: HI/36, HI/150, HI/8, HI/9, HI/46, HI/56.

5) Open circles. Correctant E14TG2a clones targeted at the *hprt* locus with pDWM101neo and selected for that event in HAT medium (3 values). Reading from left to right, clones: HI/H2, HI/H4, HI/H5. The value with a filled circle in the same column as HI/H5 is the subclone selected from HI/H5 for the ability to survive in 6-TG. It may be seen that this clone has the mutant genotype and is thus a separate, untargeted clone, by referring to figure 2-5.

6) Open diamonds. Correctant E14TG2a clones targeted at the *hprt* locus with pDWM101neo and selected for that event in G418 (7 values). Reading from left to right, clones: HI/149 (duplicate values were obtained for this clone), HI/122, HII/252, HII/361, HII/464, HII/349. The value with a filled diamond in the same column as HI/122 is the subclone selected from HI/122 for the ability to survive in 6-TG. It may be seen that this clone has the mutant genotype and is thus a separate, untargeted clone, by referring to figure 2-5.

Figure 2-6


1600
1400
1200
1000
800
600
400
200
0
-200
-400
HPRT Enz. Activity (cpm/µg)
1
2
3
4
5
6
Clone Group

2.5 Discussion

2.5.1 Failed Recombinants Appear to be Rare

In experiments HPRT/I and HPRT/II a total of 451 $G418^r$ clones were isolated and analysed by phenotype. A little over a third of these (171; 38%) were screened by Southern analysis using either exon 3 or the entire cDNA of the *hprt* gene as a probe, and no evidence was found for failed recombinants; that is, there were no clones which had a defined targeting event when analysed by Southern hybridisation which were HAT^s. This result places an upper bound on the frequency of such events, with a 95% confidence limit of 2% of stably-transfected cells, or 46% of targeted events, although here (targeted clones analysed by Southern hybridisation) the sample size is small (5 clones) and the frequency of failed recombinants may be much lower.

2.5.2 Mixed Clones Were Obtained at High Frequency

A surprising result from these experiments was that there was a sizeable proportion of clones which were HAT^r yet when analysed by Southern appeared to be untargeted. This was, if anything, the exact opposite of what was expected from the experiment, since even if HR insertion events proceed with perfect fidelity the number of HAT^r clones may only equal the number of clones shown to be targeted at the level of the genotype, but never exceed it. This therefore suggested that some "clones" were in fact mixtures, and that depending on the nature, or lack, of the selection applied to a subculture of cells, different populations of cells outgrew the others to become the majority. A small proportion of HAT^r cells present may not be noticeable on Southern hybridisation but following subculture and selection in HAT medium, may overgrow the untargeted *hprt*$^-$ cells.

Therefore groups of clones originally assigned as HAT^r were re-tested for their ability to survive in HAT medium. These results suggested that a sizeable proportion (about

50-75%) of "clones" so produced were in fact mixtures, with a very few HAT^r cells contaminating a majority of $G418^r/HAT^s$ cells. In the case of two clones which appeared to be resistant to both HAT and 6-TG, it was formally shown by repeated Southern analysis after subselection that the clones were mixed.

This problem may have arisen for several reasons. For example, it is very difficult when plating out ES cells to ensure perfect cloning because ES cells tend to form clumps of a very few cells when they first adhere to the substrate. At higher plating densities, clones can grow into each other, so that what may appear to be a single clone is in fact composed of more than one clone. The creation of mixed clones from otherwise pure populations of cells is also possible during either feeding of plates during selection or picking clones, because cell aggregates are released from the plate surface into the medium by mechanical disruption, allowing the movement of cells to other sites on the plate or their contamination of wells containing other clones, during the picking process.

2.5.3 HPRT Activities of Targeted Clones Are Lower than those of Wild-Type.

It may be that a mutation introduced into the locus by the recombination event affects the function of the protein but does not completely destroy it, and therefore the relative *hprt* enzyme activity of a group of the targeted (HAT^r) clones was analysed biochemically and compared to that of wild-type and mutant cells. The individual results for each clone tested are shown in table 2-3 and figure 2-6. The clones were also grouped according to the starting cell type (wild-type E14 or mutant E14TG2a), whether they had been subjected to a targeting procedure or not, and of those transfected with vector pDWM101neo, whether they were untargeted or were identified as targeted after initial selection in G418 or HAT medium. Between 3 and 8 values were obtained for each cell type, and the mean *hprt* enzyme activity for each

is shown in table 2-4 and figure 2-7. Using the Mann-Whitney non-parametric two-sample rank test it was shown that there is no significant difference in *hprt* enzyme activity between unmodified E14 cells and E14 cells carrying a *neo*-containing construct at an irrelevant locus. There was also no significant difference in *hprt* enzyme activity between unmodified E14TG2a cells and E14TG2a cells transfected with pDWM101neo but which have not undergone a recombination event at the *hprt* locus. However, the difference in *hprt* enzyme activity between wild-type and mutant cells is significant, as expected. It was therefore concluded that any observed differences in enzyme activity could not be attributed to culture or selection artefacts. There was also no significant difference in *hprt* enzyme activity between E14TG2a cells targeted with pDWM101neo whether the cells were selected initially in HAT medium or in G418. This suggested that the *hprt* gene is no more likely to be mutated if the correction event is selected at the level of phenotype or screened at the level of genotype, and was therefore taken as implying that the recombination event generally proceeds with fidelity. However, it is obvious from figure 2-7 that there is a striking difference in *hprt* activity between wild-type cells and targeted correctants, and this difference is significant. The standard deviations of these groups appear comparable, suggesting similar distributions around different mean enzyme activities. Does this suggest that most of the correctants have undergone errors in recombination, with perhaps clone HII/464 being the only faithful event out of the 9 clones analysed (see figure 2-6)? This is a possible explanation, but the Dixon parameter for the *hprt* enzyme activity of this clone in comparison with the other clones targeted with pDWM101neo is 0.493, providing no evidence that this clone is a significant outlier from the group ($P>0.05$; Dixon, 1953). Furthermore, it is now known that in ES cells and mice carrying an *hprt* allele corrected with the parent vector (pDWM101) of the construct used in the present study *hprt* transcription occurs at a reduced level (in most tissues, approximately 30% of normal) compared

Table 2-4 and Figure 2-7: Mean HPRT Enzyme Activities

This Table and Figure together show the mean *hprt* enzyme activity for each type of clone described in the present study. These are:

1) Wild-type E14 cells (*hprt*$^{+}$; HAT^{r}).

2) E14/G418^{r} clones are E14 (and thus HAT^{r}) clones which have been subjected to G418 selection and cloning. They were included in the analysis to monitor for any change in *hprt* activity caused by long-term culture and selection.

3) E14TG2a cells (*hprt*$^{bm\text{-}3}$; HAT^{s}).

4) E14TG2a/G418^{r}/HAT^{s} clones are a cohort of clones from experiment HPRT/I which were shown to be untargeted both by Southern analysis and HAT selection.

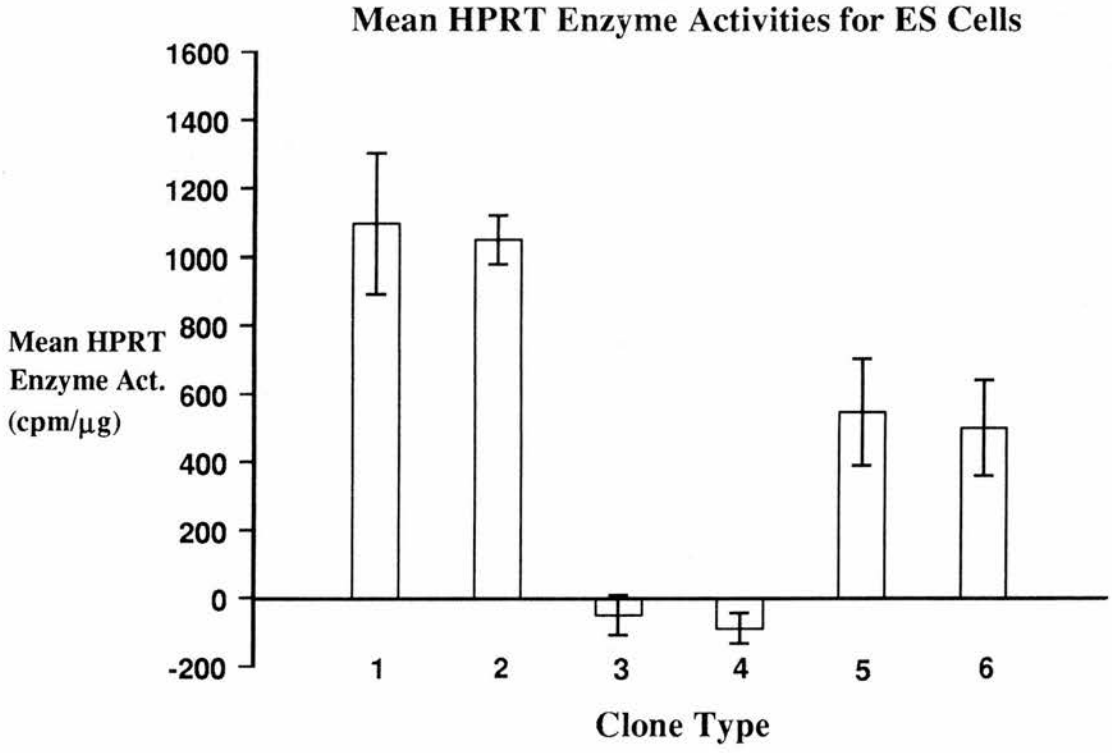
5) E14TG2a/HAT^{r} clones were targeted with vector pDWM101neo and selected in HAT medium as positive controls for the predicted targeted repair of the *hprt* gene.

6) E14TG2a/G418^{r}/HAT^{r} clones were targeted with vector pDWM101neo and selected in G418. Analysis by Southern and/or HAT selection showed they were targeted at the *hprt* locus.

Table 2-4 Mean HPRT Enzyme Activities

N^{0}	Clone Type	n	Mean Activity	Std. Dev.	Std. Error
1)	**E14**	4	1097	412	206
2)	**E14 /G418^{r}**	7	1051	184	70
3)	**E14TG2a**	4	-51	118	59
4)	**E14TG2a/G418^{r}/HAT^{s}**	8	-90	123	44
5)	**E14TG2a/HAT^{r}**	3	545	272	157
6)	**E14TG2a/G418^{r}/HAT^{r}**	6	499	346	141

Figure 2-7



to that of the wild-type allele (Melton, 1990). This was attributed to suppression of transcription caused by the presence of bacterial plasmid sequences immediately upstream of the *hprt* promoter. This is most likely to be the cause of the reduction in *hprt* activity seen in these clones.

2.6 Conclusion

There was no incontrovertible evidence obtained in this study for a high frequency of unfaithful HR events occurring in insertion events. Although most targeted correctant events analysed had a variably-reduced level of *hprt* activity compared with wild-type ES cells, it is unlikely that such changes were caused by the introduction of deleterious mutations during the integration event. No clones analysed at the level of Southern blot showed anomalous patterns and the reduction in *hprt* enzyme activity was also observed to an identical extent in clones which were directly selected for targeted correction of the *hprt* allele in HAT as well as those subjected to selection in G418, (and which were thus not dependent upon the restoration of gene function for their survival). The introduction of the *neo* gene may have contributed to the effect on expression caused by the bacterial plasmid sequences, resulting in a wide variability of expression. This objection perhaps represents a fault in the experimental design (which was not obvious initially) as pDWM101 was originally thought to fully restore *hprt* activity (Thompson *et al.*, 1989). The reduction in expression observed makes the separation of those clones which have a reduced level of transcription due to interfering sequences from those with a point mutation difficult. Further work to clarify this issue is required. I suggest the use of RT-PCR of the *hprt* mRNA from targeted clones followed by single-stranded conformational polymorphism (SSCP) analysis as a possible strategy. Digestion of the cDNA into 3 or 4 approximately equal-sized fragments prior to separation and SSCP analysis could be expected to detect 80-90% of mutations introduced into the coding sequence of the gene

(Hayashi, 1992). Subject to these reservations however, the conclusion from this work is that, consistent with previous results (Zheng *et al.*, 1991), gene targeting by insertion does appear to proceed with a level of accuracy compatible with the accomplishment of experiments requiring a high degree of fidelity in the HR event, such as hit and run.

CHAPTER 3-

INTRODUCTION OF A DELETION BY HIT AND RUN

Introduction of a Deletion by Hit and Run

3.1 Introduction

The advantage of the Hit and Run technique is its fine control of the genetic changes created; that is the ability to introduce a mutation into the mammalian genome with great precision, leaving no other changes which might feasibly complicate analysis of the phenotype produced. It is therefore likely to find most application in the creation of subtle mutations, such as single base-pair substitutions, with the objective for example of producing strains of mice which provide models of human congenital diseases by carrying similar alleles in homologous genes. Precise analysis of this process is difficult as such mutations are often difficult to detect because the analysis must be locus and sequence-specific. To permit the relatively simple study of a hit and run experiment through each step, it was therefore decided that the mutation should be sufficiently large as to be observable by Southern blot at each step. Frequencies of all the observed events would therefore be easily calculable, without the need either to study a selectable gene such as *hprt*, or to analyse by sequencing large numbers of clones.

This experiment used the vector pKiD0PNT described below to attempt the deletion of a 2.3kb region encompassing exon 0 and the associated promoter sequence of the murine K-*ras* gene. In addition to the characterisation of the hit and run process, it was expected that a K-*ras* knockout would be produced, which would be a valuable resource in the study of K -*ras* involvement in development and neoplasia.

3.2 Targeting Vector pKiD0PNT

Targeting vector pKiD0PNT was constructed by Dr D. James Williamson of our group, designed to introduce a deletion into the K-*ras* gene of the mouse. The predicted deletion would span approximately 2.3 kb of sequence 5' to the first

translated exon of the K-*ras* gene, completely encompassing the untranslated exon 0 and the promoter region. It was constructed from the vector pPNT (Tybulewicz *et al.*, 1991; see figure 7-1 [b]) by incorporating a 4 kb *Sal* I fragment from vector PBKT-142 (Guerrero *et al.*, 1984) from which a 2.3 kb *Xba* I fragment containing exon 0 and the surrounding sequence had been removed. The vector is shown in figure 3-1, and figure 3-2 shows the predicted changes in the genome made by this vector.

3.3 Results

3.3.1 Insertion Step

Vector pKiD0PNT was electroporated into E14 ES cells as described, and 158 $G418^r$ clones were successfully screened using probe 3 (figure 3-2) as an external probe located within the region of homology. Probe 3 comprises the *Xba* I fragment of the deletion and is therefore an external probe because it is not homologous to any vector sequences. This slightly unconventional strategy was adopted because problems of specificity were initially experienced with the probes designed for this experiment, probes 4' (5' external) and 5 (internal). One clone, designated P66, gave a band-shift to a larger size as expected for a hit event, and was accordingly screened further to verify its structure. Use of probe 4' initially suggested that P66 was a targeted clone of the structure indicated in figure 3-2 (see figure 3-3 [a]).

3.3.2 Back Selection of Clone P66: High Cell Density

Clone P66 was accordingly expanded to 10^8 cells and back-selected in ganciclovir with the objective of producing clones carrying the 2.3 kb deletion in the K-*ras* gene. The density of cells was $5x10^6$ cells per 100mm plate. 36 clones were obtained, designated P66/1-P66/36 inclusive.

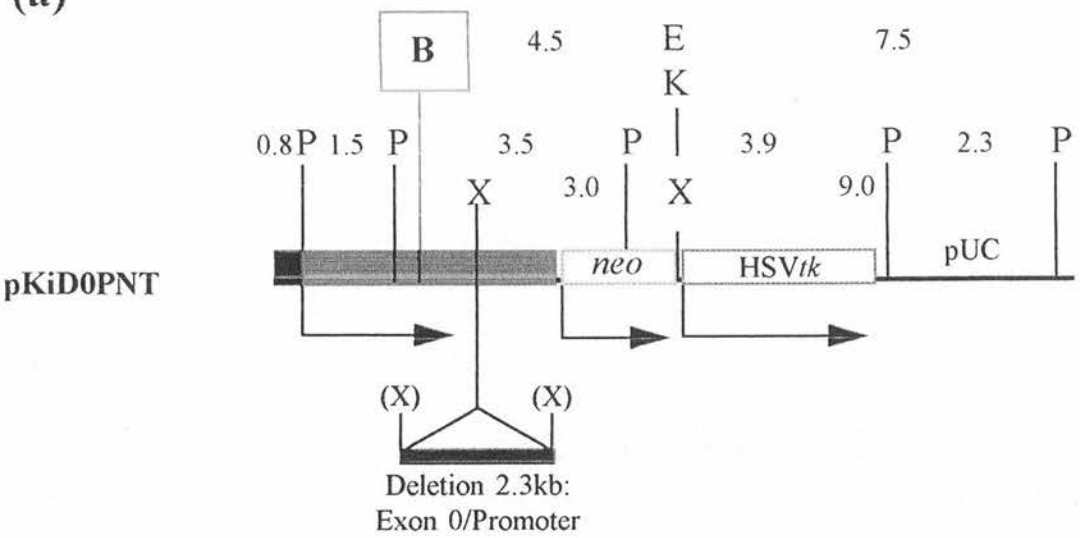
Figure 3-1: Targeting Vector pKiD0PNT

a) The structure of vector pKiD0PNT. The vector carries 3.5kb non-isogenic homology (source: hybrid of mouse strains AKR-RF/J) to the 5' upstream sequences of murine K-*ras* (grey block). A short region of bacterial plasmid derived from the construct PBKT-142 (Guerrero *et al.*, 1984) is also present (black block). The vector carries a 2.3kb deletion including the promoter sequences and exon 0, encompassed by two *Xba* I sites (X). The vector is derived from pPNT (see figure 7-1 [b]) and as such includes a *neo* and an HSV*tk* gene (orientation as shown) and a bacterial plasmid (pUC 18, shown as a thick black line). Relevant restriction sites are shown: B, *Bam*H I, which is the linearisation site of the vector; X, *Xba* I; P, *Pvu* II; K, *Kpn* I; E, *Eco*R I. The sizes in kb of fragments from selected restriction digestions of the construct are shown.

b) Restriction enzyme digestion of vector pKiD0PNT. Lanes: 1, *Bam*H I; 2, *Bam*H I & *Eco*R I; 3, *Eco*R I; 4, *Pvu* II; 5, *Xba* I; 6, DNA molecular size markers. N.B.: the 0.8kb *Pvu* II fragment is not visible in lane 4 because it is too small.

Figure 3-1

(a)


B
4.5
E
K
7.5
0.8P 1.5 P
3.5
P
3.9
P
2.3
P
X
3.0
X
9.0
pKiD0PNT
neo
HSVtk
pUC
(X)
(X)
Deletion 2.3kb:
Exon 0/Promoter

(b)

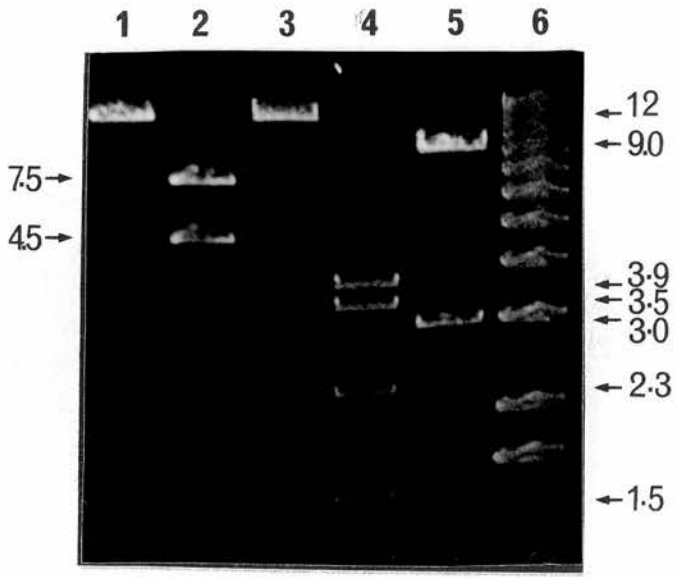

1
2
3
4
5
6
7.5
4.5
12
9.0
3.9
3.5
3.0
2.3
1.5

Figure 3-2: Targeting Structures for Vector pKiD0PNT

The figure shows, to scale, the structure of the wild-type murine K-*ras* locus in the exon 0-exon 1 region, and the predicted insertion and deletion structures introduced at the locus with vector pKiD0PNT. Relevant probes are shown: 4' and KR8XK, 5' external; probe 5, internal; probe 3, external located internal to the homology. Relevant restriction sites are also shown: B, *Bam*H I; E, *Eco*R I; X, *Xba* I; K, *Kpn* I; H, *Hin*d III.

The expected band sizes (in kb) for each of the possible structures with internal and external probes on an *Eco*R I digest are shown in the table below:

Probe	Wild-Type	"Hit" Insertion	"Hit & Run" Deletion
Internal (Probe 5)	8.7+1.5	9.6(5')+11(3')+1.5	7.8
5' External (Probe 4')	8.7	9.6	7.8

Figure 3-2

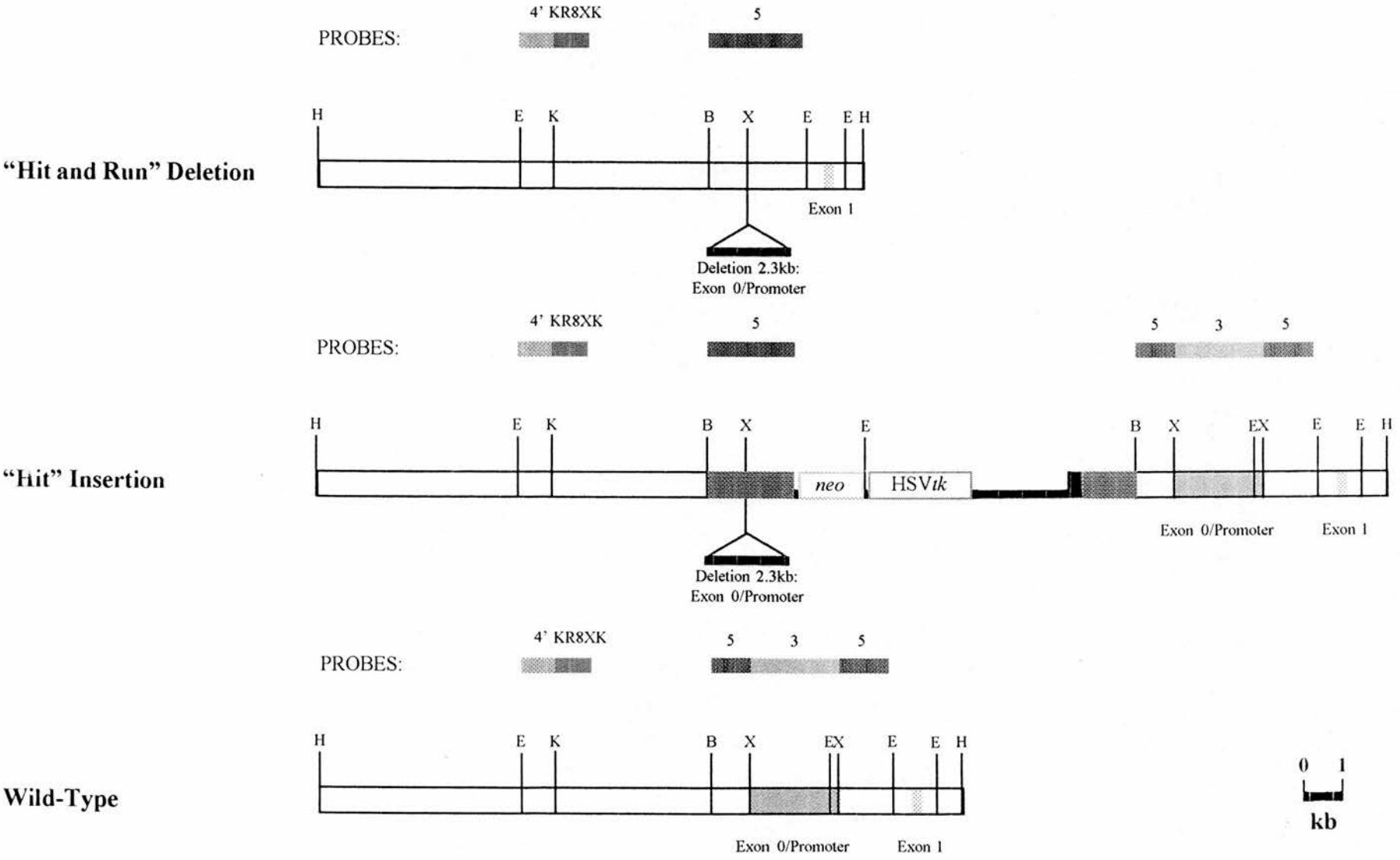

PROBES:
4' KR8XK
5
"Hit and Run" Deletion
H
E
K
B
X
E
E
H
Exon 1
Deletion 2.3kb:
Exon 0/Promoter
PROBES:
4' KR8XK
5
5
3
5
"Hit" Insertion
H
E
K
B
X
E
neo
HSVtk
B
X
EX
E
E
H
Exon 0/Promoter
Exon 1
Deletion 2.3kb:
Exon 0/Promoter
PROBES:
4' KR8XK
5
3
5
Wild-Type
H
E
K
B
X
EX
E
E
H
Exon 0/Promoter
Exon 1
0
1
kb

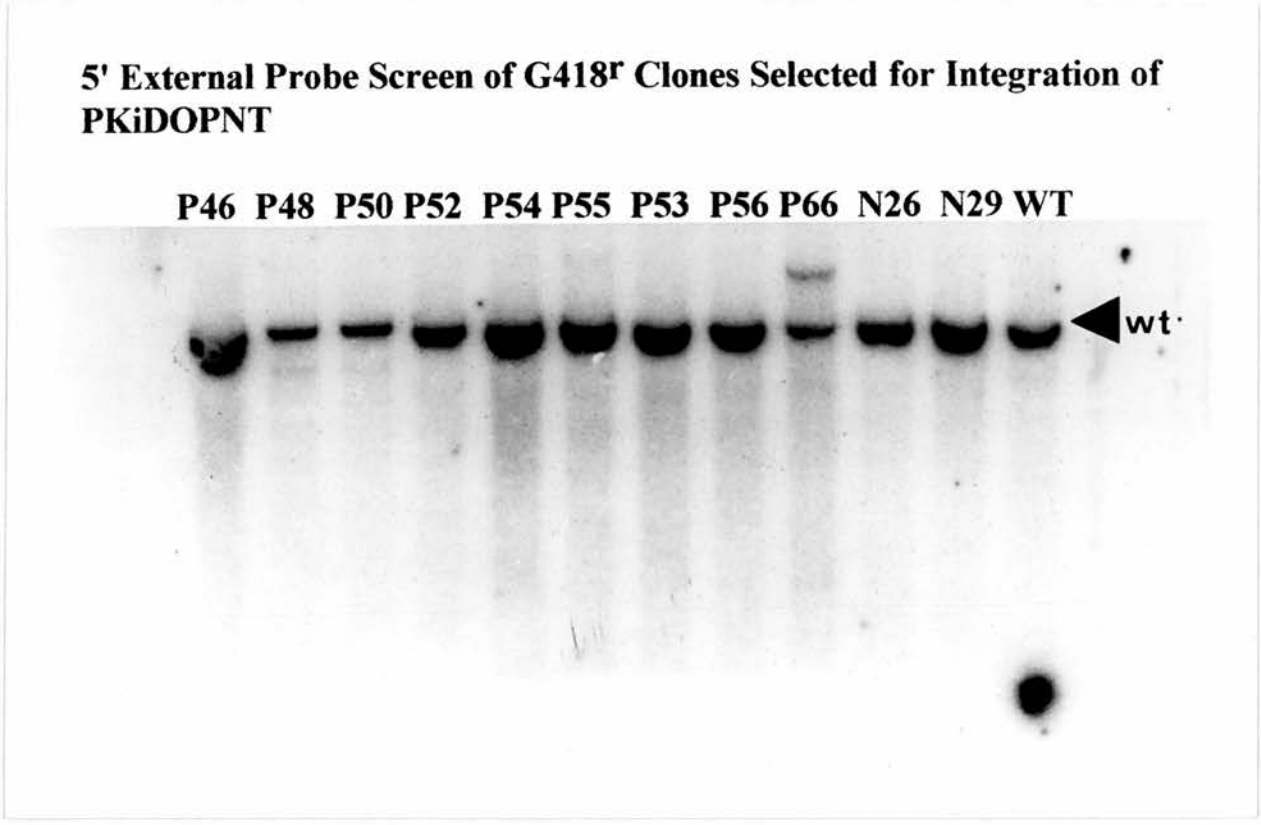
Figure 3-3: Vector pKiD0PNT Integrates at the K-*ras* Locus

a) Use of 5' external probe 4' to check an *Eco*R I digest of DNA prepared from *neo*$^+$ clones obtained with vector pKiD0PNT and initially screened with probe 3. The 8.7kb band due to the wild-type allele and present in all clones is indicated.

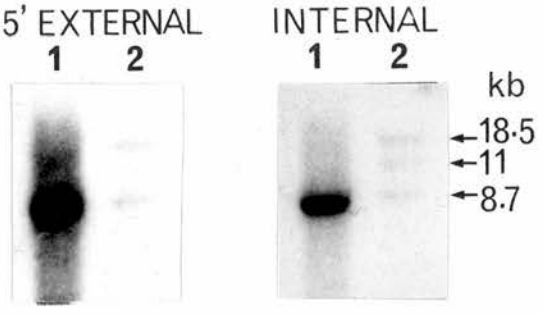
b) The same blot of 129/Ola wild-type (lane 1), and clone P66 (lane 2) DNAs digested with restriction endonuclease *Eco*R I, probed with a 5' external probe (KR8XK; left-hand blot) and internal probe 5 (right-hand blot).

Figure 3-3

(a)


5' External Probe Screen of G418r Clones Selected for Integration of PKiDOPNT
P46 P48 P50 P52 P54 P55 P53 P56 P66 N26 N29 WT
wt

(b)


5' EXTERNAL
1 2
INTERNAL
1 2
kb
18·5
11
8·7

These clones were analysed by Southern Blot and PCR, using the following analyses:

1) Internal Probes (P5, *neo*, HSV*tk*)

2) 5' External Probes (P4' & KR8XK)

3) *neo* PCR

Sample data obtained with internal probe 5 are shown in figure 3-4. The clones appeared to fall into several broad classes (table 3-1).

3.3.3 Back Selection of Clone P66: Low Cell Density

Small numbers of clones were obtained in the selections described in sections 3.3.2, 3.3.5.2 and 4.4, and this was probably due in major part to metabolic cooperation deleting many clones by "kiss of death" which would otherwise have survived. This hypothesis was tested as described in section 4.5.1, and it was concluded that for ganciclovir selection 10^5 cells/100mm plate is the optimum number to plate. P66 was therefore plated out at this cell density, and the backselection was repeated. 90 clones were screened using internal probe 5 and 5' external probe KR8XK. The result of this selection is shown in table 3-2 with sample data from the Southern analysis shown in figure 3-5.

3.3.4 Clone P66 has a Modified K-*Ras* Allele, but with an Unpredicted Structure

Figure 3-3 (b) shows the results of Southern analysis of HM1 (strain 129) wild type and P66 K-*ras*-modified ES cell genomic DNA with internal probe 5 and 5' external probe KR8XK, performed on the same blot. Although P66 was originally thought to be a correctly-targeted clone, the data in the figure, and the results of the back selection at a low cell density (described in the previous section) suggest that this is

Figure 3-4: Backselection of Clone P66 (High Cell Density)

The figure shows sample data obtained with internal probe 5 on an *Eco*R I digest of some clones obtained by backselection of clone P66. The 8.7kb wild-type band is indicated. The clone in lane 18 was originally thought to be a hit and run revertant carrying a deletion.

Table 3-1

All the data obtained for the 36 clones obtained by backselection of clone P66 under high cell density with all the probes shown are summarised.

Figure 3-4

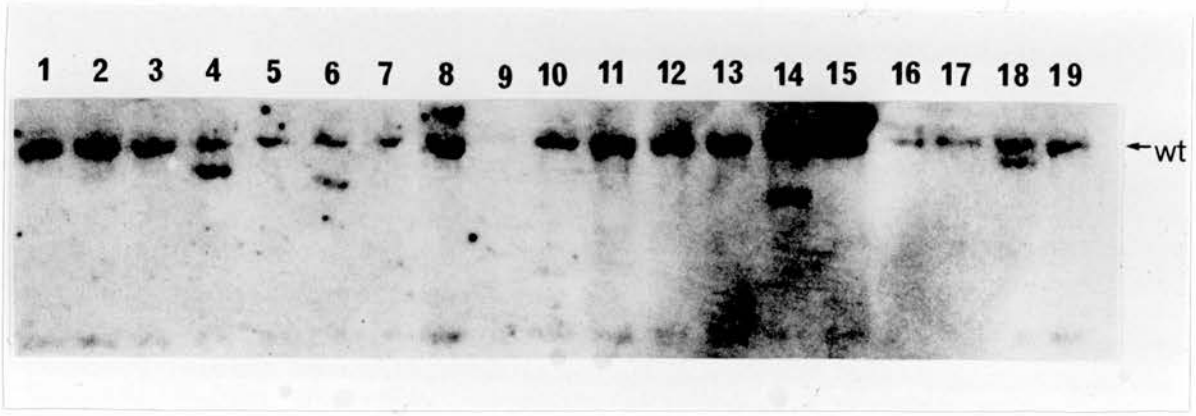


Table 3-1.

High-Density Backselection of K-*ras* Clone P66

	Internal Probe	External Probe	*neo* Status	Number
Class 1	WT	WT	-	14
Class 2	WT	WT	+	5
Class 3	WT+1 band	WT	+	9
Class 4	WT+1 band	WT	-	1
Not fully assigned; WT based on a subset of the probes described				2
Not fully assigned; *neo*$^+$				5

Note: "WT" indicates the band pattern expected for a clone which is wild-type at the K-*ras* locus is observed; i.e. an 8.7kb band is detected with internal probe 5 and 5' external probe 4' on an *Eco*R I digest.

Figure 3-5: Backselection of Clone P66 (Low Cell Density)

The figure shows sample data obtained with internal probe 5 on an *Eco*R I digest of some clones obtained by backselection of clone P66. The 8.7kb wild-type band and large band (~18kb) due to the modified allele of clone P66 are both indicated. Lane 1 is wild-type 129/Ola strain DNA, and lane 2 is clone P66. Lanes 3-13 inclusive are revertant clones of P66 obtained after selection in ganciclovir. Two patterns are observed for the revertant clones; either a pattern identical to that of the wild-type allele, or a pattern lacking the middle, 3', band, but retaining a wild-type and a large, 5', band.

Table 3-2

All the data obtained for the clones obtained by backselection of clone P66 under low cell density with both internal (probe 5) and 5' external (KR8XK) probes are summarised.

Figure 3-5

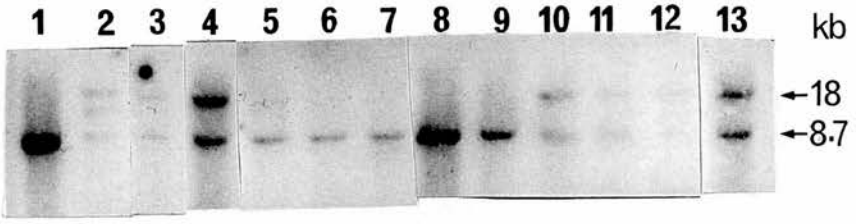


Table 3-2

Low-Density Backselection of K-*ras* Clone P66

	Internal Probe	External Probe	Number
Class 1	WT	WT	55
Class 2	WT+18kb band	WT+18kb band	35

Note: "WT" indicates the band pattern expected for a clone which is wild-type at the K-*ras* locus is observed; i.e. an 8.7kb band is detected with internal probe 5 and 5' external probe 4' on an *Eco*R I digest.

not the case. As can be seen from the Southern analysis, the band pattern differs from that expected. The smallest band, of approximately 8.7 kb, is due to the wild type allele on one chromosome. For a correct targeting event of the kind shown in figure 3-2 two larger bands are expected with internal probe analysis. These additional bands are 9.6kb (5' fragment) and 11kb (3' fragment) in size. The largest fragment observed is approximately 18kb. Because this fragment is also observed with the external probe, it is concluded that the K-*ras* gene is disrupted, with the larger fragment at the 5' end of the modified locus. Because the large fragment was also observed in many clones obtained upon back selection of P66, it was concluded that this fragment does not incorporate a functional HSV*tk* gene. This discrepancy is not due to a simple error in genomic mapping because were the 5' *Eco*R I site located further away from the locus, making the targeted fragment larger, the wild-type fragment would also be correspondingly large.

As shown in figure 3-1, there is an *Eco*R I site in vector pKiD0PNT, immediately 5' to the HSV*tk* gene, and it is this site which should limit the size of the detected 5' *Eco*R I fragment of the targeted locus to 9.6kb. The fragment observed is some 8kb larger than expected, indicating that an *Eco*R I site is missing. On back selection even an unfaithful excision event would usually be expected to remove this site because of its proximity to the HSV*tk* gene, and thus change the observed sizes of both the 3' and 5' fragments by merging them into one fragment. It was therefore concluded that either the 3' *Eco*R I site of the 18kb 5' fragment is not the 5' *Eco*R I site of the 3' fragment, or that the commonly occurring excision event restores a fragment of the same size (18kb).

3.3.5 Fast-Track Hit and Run with pKiD0PNT

3.3.5.1 Rationale

The targeting frequency of K-*ras* with vector pKiD0PNT is low: less than 1% of integrations result in a homologous recombination event. Furthermore, the only event observed was not the one anticipated in the experiment design. This may be due to a feature of the experimental design, or of the K-*ras* locus in the mouse-for example the location of the gene may attenuate expression, thus selecting for competing events such as the one suggested for P66 by selecting for events which integrate more than one *neo* gene at the insertion stage (see discussion section).

In an attempt to circumvent this, and also to investigate the efficiency of the hit and run process, the experiment described in section 3.3.5.2 was performed. The assumption upon which this experiment rests is that the number of clones subjected to backselection is much increased; to approximately 4000 in this case. Thus the chance of carrying out the backselection upon a rarely-occurring correctly targeted clone is increased some 27-fold. The individual number of cells representing each neo^+ clone is smaller, but still numbers approximately $5x10^4$ cells for each clone. If it is assumed that the majority of clones are random integrants, the corresponding majority of cells in the selection will lack the requisite duplication of homology to undergo a homologous recombination event and will die in ganciclovir selection. Some such cells can be expected to form clones by undergoing a point mutation or deletion in the HSV*tk* gene. A mutation frequency of between 1 in 10^6-10^8 may be expected (data on *hprt,* see, for example, Valencius and Smithies, 1991b) resulting in about 20 $ganciclovir^r$ clones derived from random events (mutation rate of 1 in 10^7 on $2x10^8$ cells) expected in this experiment. Reversion frequencies of from 1 in 10^3 to 1 in 10^6 have been recorded for backselection of characterised "hit" clones (Valencius & Smithies, 1991b, Hasty *et al.*, 1991c). This suggests a range of approximately 2-

2000 revertant clones would be obtained from such an experiment, assuming an initial targeting frequency of 1 in 150 $G418^r$ colonies. It was therefore expected that the Fast-Track protocol would produce relatively few clones at the end of the experiment, but that those obtained would be enriched for ones which had reverted back to wild type or the desired mutant from a targeted hit event over those derived from random integration events which had managed to survive by loss of the HSV*tk* gene by mutation.

3.3.5.2 Results of the Fast-Track Hit and Run Experiment

HM1 cells were used for this targeting experiment because of their reputedly highly efficient targeting and germline transmission frequencies (Magin *et al.*, 1992). 10^8 ES cells were electroporated with pKiD0PNT as described, and subjected to selection in G418 for 14 days. Approximately 4000 clones were obtained. All these clones were pooled and expanded to $2x10^8$ cells. These cells were then also subjected to back selection as described in section 7.4.2. 58 clones were obtained, designated HKiD01-HKiD058 inclusive. These clones were studied by Southern analysis as shown in figure 3-6 and summarised in table 3-3.

3.3.6 Chimaera Production

Initial results based on internal probe data (probe 5) suggested that 4 clones, P66/32, HKiD0/10, HKiD0/52 and HKiD0/55 had undergone the desired "run" event to produce a clone carrying the intended deletion. These clones were injected into blastocysts and chimaeras were made. Germline transmission was not observed. After 1 year, the 5 chimaeras were killed, subjected to necropsy and the tissue samples taken were examined histologically by a pathologist (Dr David J. Harrison, Edinburgh). Three of the mice had tumours. The first mouse had a teratoma.

Figure 3-6: Fast-Track Hit and Run with Vector pKiD0PNT

The figure shows sample data obtained with internal probe 5 on an *Eco*R I digest of some clones obtained by a fast-track hit and run protocol as described in the text. The 8.7kb wild-type band is indicated. Clones in lanes 10 and 13 were originally thought to be potential hit and run deletion revertants. Lane 17 includes a wild-type control.

Table 3-3

All the results obtained with the probes shown for the clones derived by the fast-track hit and run protocol described in the text are summarised.

Figure 3-6

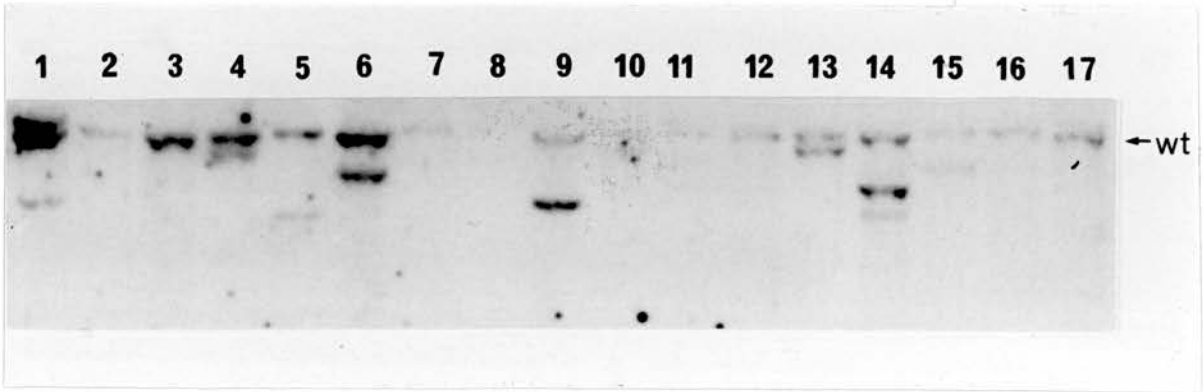


Table 3-3

Clones Obtained by Fast-Track Hit and Run with pKiD0PNT

	Internal Probe	External Probe	*neo* Status	Number
Class 1	WT	WT	-	16
Class 2	WT	WT	+	2
Class 3	WT+1 band	WT	+	9
Class 4	WT+1 band	WT	-	3
Class 5	WT+>1 band	WT	-	1
Class 6	WT	WT+18kb	-	1
Class 7	WT+>1 band	WT+18kb	-	1
Class 8	WT+1band	WT+18kb	-	1
Not fully assigned; WT based on a subset of the probes described				18
Not fully assigned; *neo*$^+$				6

Note: "WT" indicates the band pattern expected for a clone which is wild-type at the K-*ras* locus is observed; i.e. an 8.7kb band is detected with internal probe 5 and 5' external probe 4' on an *Eco*R I digest.

Teratomas have a high incidence in strain 129 (Stevens, 1983; Hardy *et al.*, 1990) and it is therefore unlikely that this tumour was a direct result of the mutation at the K-*ras* locus. A second animal, (a female), had a malignant liver tumour. However, a PCR (Cui *et al.*, 1993; kindly performed by Dr Jane Armstrong) to decide the sex of the tumour tissue showed it to lack a Y chromosome. This result does not formally exclude the possibility that the liver tumour was derived from a karyotypically abnormal cell (e.g. an XO segregant of an ES cell-derived XY cell). However, because the mouse was a female and the levels of chimaerism observed, as estimated by coat colour mosaicism, were uniformly low (<10%), it was most likely that the tumour was in fact derived from host blastocyst tissue rather than the XY (male) E14 ES cell line. The third animal examined had a malignant leukaemia. Glucose phosphate isomerase isozyme analysis was performed on part of the tissue, and showed a slight elevation in *GPI-1s*^{a}. The host strain of mouse is *GPI-1s*^{b}, whereas strain 129 is *GPI-1s*^{a}. However a representation of *GPI-1s*^{a} of approximately 20% minimum is necessary to conclude that a tissue is composed to a significant extent of ES cell-derived cells because of the relatively high background of GPI isozyme analysis, and the slight elevation observed in the leukaemia sample was not sufficient to confirm that the tumour was ES cell-derived. From these results it was concluded that none of the chimaeras' disease could be unequivocally attributed to the introduced mutation and, because of the small sample size of the chimaeras, no further analysis was carried out on these animals.

3.4 Discussion

3.4.1 The Structure of Clone P66

Only one clone characterised at the stage of insertion of vector pKiD0PNT into the ES cell genome appeared to have integrated the vector at the K-*ras* locus. This

clone was designated P66. As considered in section 3.3.4, it appears however that this clone does not have the structure predicted (shown in figure 3-2).

3.4.1.1 Common Competing Events Seen in Targeting Experiments Do Not Account for the Results Obtained with Clone P66.

The following sections consider possible structures for the unpredicted K-*ras* insertion clone P66, and some mechanisms which might have produced such structures.

3.4.1.1.1 Recircularisation of the Vector

Should the linearised vector recircularise prior to integration into the genome by homologous recombination, it may undergo either an insertion event at an unplanned point in the locus or integrate by replacement. Here however, the insertion would result in either the designed structure shown in figure 3-2 or a reverse, discussed as possibility (2) in the next section. If the vector integrates by replacement, it would produce a deletion at the designed site but this would never be detected because the resulting clone would be *neo*$^-$ and therefore fail to survive G418 selection. Recircularisation of the vector prior to integration by HR in an unplanned fashion therefore does not account for the structure of P66.

3.4.1.1.2 Branch Migration

Upon generation of a cross-over structure at the insertion site of the vector, branch migration of the Holliday Junction along the homology can occur. This may result in the changing of the structure obtained from that predicted by conversion of a mutation to wild-type or vice-versa when the heteroduplex is repaired (Hasty & Bradley, 1993). Four possibilities may be expected in the case of an insertion vector carrying a mutation in an arm of homology. First, the predicted structure, where the

mutation is present in one of the linear duplications of homology (in this case, the 5' one), derived from the vector, and the other duplication is wild-type, being derived from the endogenous locus. This structure has been ruled out above (see section 3.3.4). Second, a reverse of the predicted structure, where the mutated duplication and the wild-type one are swapped relative to each other due to branch migration and heteroduplex repair. The third possibility is a locus where both of the duplicated regions of homology are wild-type. Fourth, a locus where both of the duplicated regions of homology contain the mutation. These possible structures are shown schematically with the predicted *Eco*R I fragments detected with internal and external probes in figure 3-7.

None of the above listed events can account for the structure of P66, because none of these possibilities would result in a structure with such a large 5' fragment. Possibilities (2) and (3), where there would be no 2.3kb deletion at the 5' duplication of homology, would appear completely wild-type when tested with the external probe. Possibility (2) would only produce one larger band (approximately 10kb) detectable by the internal probe, plus a small band of 3.6kb. Possibility (4) would not have been identified by the initial screen, which relied on the presence of the 2.3kb fragment in the targeted locus, and in any case would give the mutated band size (9.6kb) for the 5' fragment with the external and internal probes which was not observed, and a 3' fragment of only approximately 10kb when analysed with the internal probe.

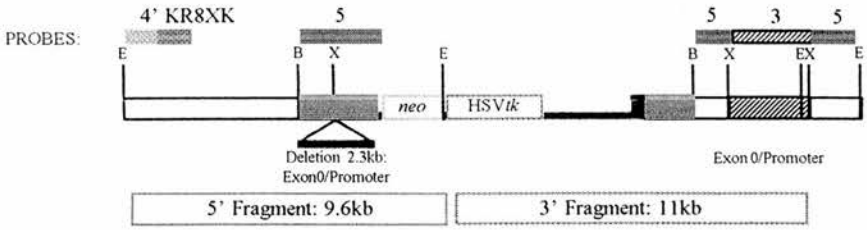
Furthermore, there is no reason to suggest that a backselection carried out on any of these possible structures would result in an unpredicted event in preference to a conventional homologous recombination which deletes the vector and one of the regions of duplicated homology.

Figure 3-7: Structures Produced Upon Branch Migration by pKiD0PNT

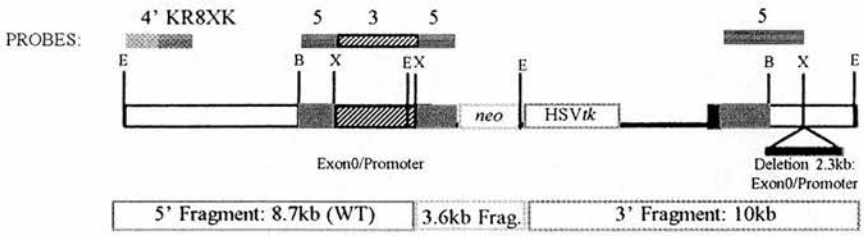
The figure shows, to scale, the alternative possible structures produced at the K-*ras* locus by vector pKiD0PNT, should branch migration and heteroduplex repair occur during homologous recombination. Below each structure the predicted band sizes which would be observed with a 5' external probe or internal probe are shown.

Figure 3-7

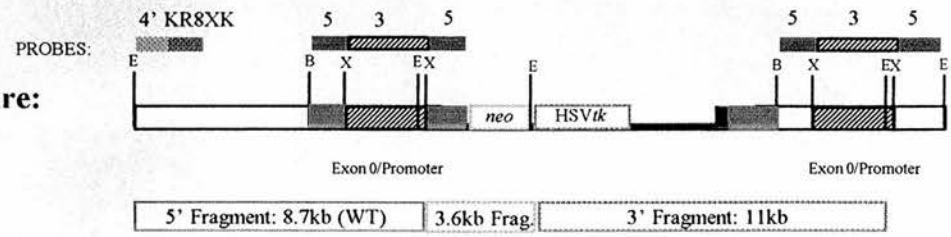
-edicted Structure:
ion on 5' Arm,
-1 3' Arm



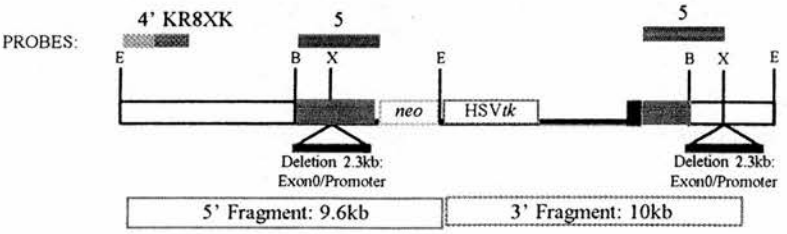
-verse Structure:
on on 3' Arm,
1 5' Arm



-uble Wild Type" Structure:
both 5'& 3' Arms



-uble Mutant" Structure:
n on both 5'& 3' Arms



3.4.1.1.3 Concatenation of the Vector.

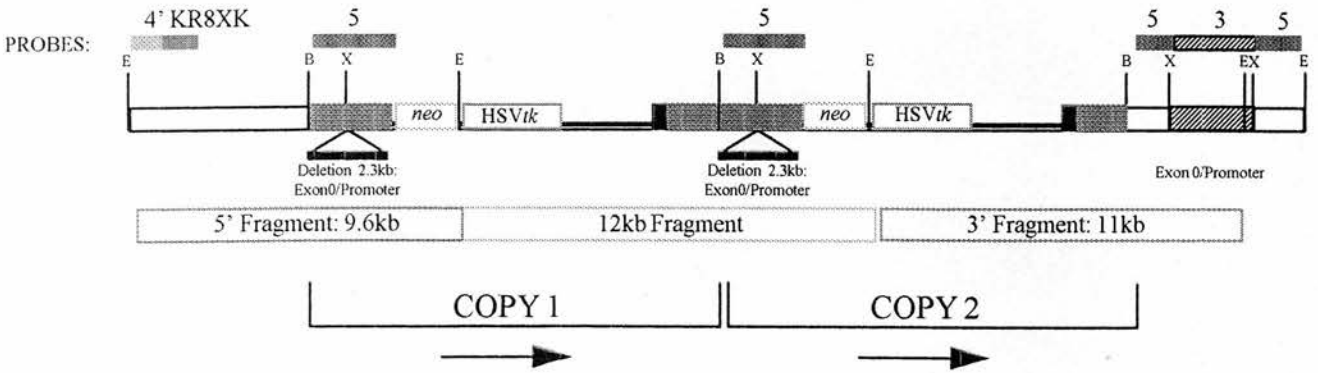
Vector pKiD0PNT is linearised with the restriction endonuclease *Bam*H I, and as such has cohesive ends when electroporated into ES cells. Thus multiple tandem copies of the vector could conceivably become integrated into the genome by concatenation into a linear array by virtue of their sticky ends prior to insertion. However, were this the case one would expect to see a minimum of 4 bands when studying the locus by *Eco*R I digestion and internal probe analysis. These would be the bands expected for the targeting event as predicted, and extra band(s) representing additional copies of the vector in the genome. Integrated copies of the vector arranged "head-tail" would give an extra 12kb band with the internal probe representing one or more copies of the complete vector. This band would be expected to become progressively more intense depending on the number of copies of the vector integrated. A 9kb fragment would be observed for an event linking two long arms of homology. The internal probe would fail to hybridise with certain *Eco*R I fragments of concatamers joined short arm to short arm, because it has no homology to this arm of the vector. The possible results of such an event are shown in figure 3-8. One would expect that the backselected clones obtained from a clone carrying multiple copies of the vector would be comparable with those obtained from a clone with a single copy, because survival in ganciclovir demands that all of the integrated copies of the HSV*tk* gene be removed in the excision event. My observations were not consistent with these predictions, and therefore it is concluded that integration of multiple copies of the vector in a tandem array into the K-*ras* locus do not account for the structure of clone P66.

Concatenation of the vector followed by its integration into the genome by replacement instead of insertion also produces a structure similar to the predicted

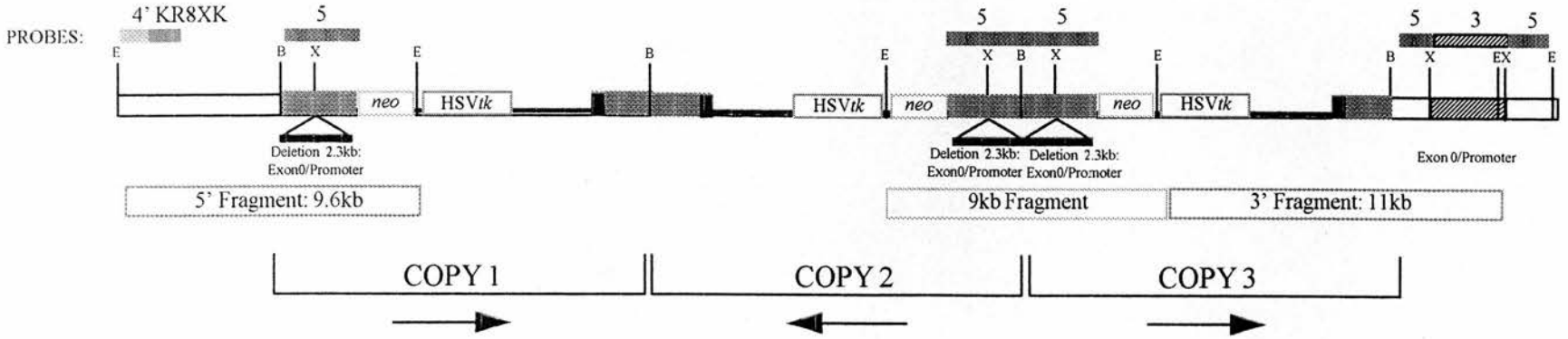
Figure 3-8: Structures Produced Upon Concatenation of pKiD0PNT

The figure shows, to scale, the two alternative structures produced at the K-*ras* locus by vector pKiD0PNT should concatenation occur. The arrow indicates the direction of the short arm (3' end) of the homology. Below the structure the bands observed by Southern analysis with 5' external probes or internal probes on *Eco*R I digest are shown. Structure 1 shows the bands obtained with a vector with concatenates "head-tail", and structure two show the more complicated possibilities observed if "head-head" or "tail-tail" concatenation events occur.

(1) "Head-Tail" Linear Concatenation



(2) "Tail-Tail/Head-Head" Linear Concatenation



event, or possibility (4) considered in section 3.4.1.1.2, which as already described is incompatible with the observations.

3.4.1.1.4 Intra-Vector Recombination

As shown in figure 7-1 (b), targeting cassette pPNT also has internal linear homologous duplications of sequence, namely the PGK promoter and poly-A sequences which drive the two selectable markers. Each duplication is approximately 400bp long. Recombination can thus feasibly occur between these regions, resulting in the deletion of either the *neo* gene (by recombination between the PGK promoters) or the HSV*tk* gene (by recombination between the PGK poly-A sequences). Whilst this mechanism is attractive in that it results in deletion of the *Eco*R I site, and thus may result in a larger fragment than expected when analysed with the external probe, the resulting clone would be either G418^s or ganciclovir^r. This is not the case: Clone P66 was isolated after G418 selection, and is therefore *neo*$^+$. The results of backselection of this clone show that it is ganciclovir^r (and is therefore also HSV*tk*$^+$).

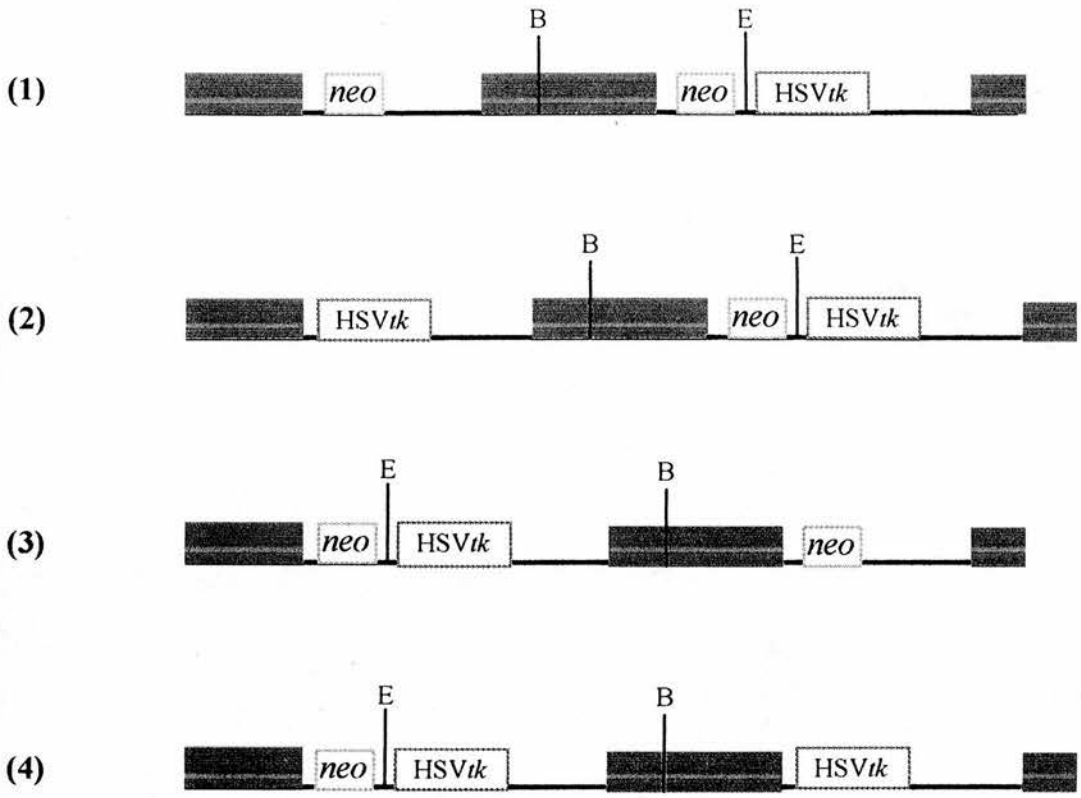
3.4.1.2 Variations on a Theme

Intra-vector recombination is the only mechanism so far suggested which would result in the loss of a restriction site and thus increase the size of the fragments observed with the internal and external probes used to analyse this locus. Concatenation of the vector prior to such an event could generate several possibilities, shown schematically in figure 3-9. The final structure generated depends on whether a *neo* or HSV*tk* gene is deleted by recombination between either the PGK promoter or PGK poly-A sequences respectively, and whether such a recombination occurs in the 5' or 3' copy of the vector.

Figure 3-9: Structures Produced By Concatenation/ECR Events With pKiD0PNT

Schematic structures obtained by a combined concatenation and ECR event upon pKiD0PNT, leading to a double structure which has lost one of the selectable markers. The *Bam*H I site (where concatenation is assumed to occur) and diagnostic *Eco*R I site are shown. The structures are described further in the text, but briefly, it is seen that there are four possibilities depending on whether intra-vector recombination occurs in the 5' (structures 1 and 2) or 3' (structures 3 and 4) copy of the vector, and whether the *neo* (structures 2 and 4) or the HSV*tk* (structures 1 and 3) gene is lost upon recombination.

Figure 3-9


(1)
B
E
neo
neo
HSVtk
(2)
B
E
HSVtk
neo
HSVtk
(3)
E
B
neo
HSVtk
neo
(4)
E
B
neo
HSVtk
HSVtk

Structures of the type shown in parts 3 and 4 of figure 3-9 are not relevant to the present discussion. This is because the deletion including the diagnostic *Eco*R I site is at the 3' end of the construct, so that the 5' *Eco*R I fragment of the structure detectable by the 5' external and internal probes is of the same size and structure as that predicted and shown in figure 3-2. Structure 2 has an HSV*tk* gene in the 5' arm, which appears unlikely because many clones obtained from back selection retain the 5' large fragment, yet survive ganciclovir selection. Inactivation of the remaining HSV*tk* gene by point mutation is unlikely in such cases because there are two copies of the gene, and thus the chance of both copies being inactivated to permit survival in ganciclovir is correspondingly smaller.

Therefore if such an event occurred in the system presently under discussion, the integrated vector is most likely to be that of structure 1 in figure 3-9. Figure 3-10 shows the structure for clone P66, based on vector pKiD0PNT undergoing such a concatenation/recombination event prior to insertion. The structure is consistent with the Southern analysis data from both the internal and external probes, in that the 5' fragment is the correct size, 18kb, making it the largest fragment to be observed at the locus. An *Eco*R I site has been lost, resulting in only two (larger than wild type) bands observable in the targeted locus with the internal probe, and one with the external probe and probe 3. As suggested from backselection data, the large 5' fragment has no HSV*tk* gene.

3.4.2 Low Cell Density Back Selection of Clone P66

This back selection is discussed first, because the results obtained are less heterogeneous than those obtained from the high density back selection of clone P66 and the "Fast-Track Hit and Run" protocol described.

Figure 3-10: Structure of Clone P66

Clone P66 is postulated to have been generated from a pair of copies of vector pKiD0PNT which have undergone a concatenation/ECR event prior to insertion into the K-*ras* locus. The figure shows the suggested structure of the modified locus, where it may be seen, consistent with experimental data, that:

1) By 5' external probe a large, 18kb, fragment is detectable.

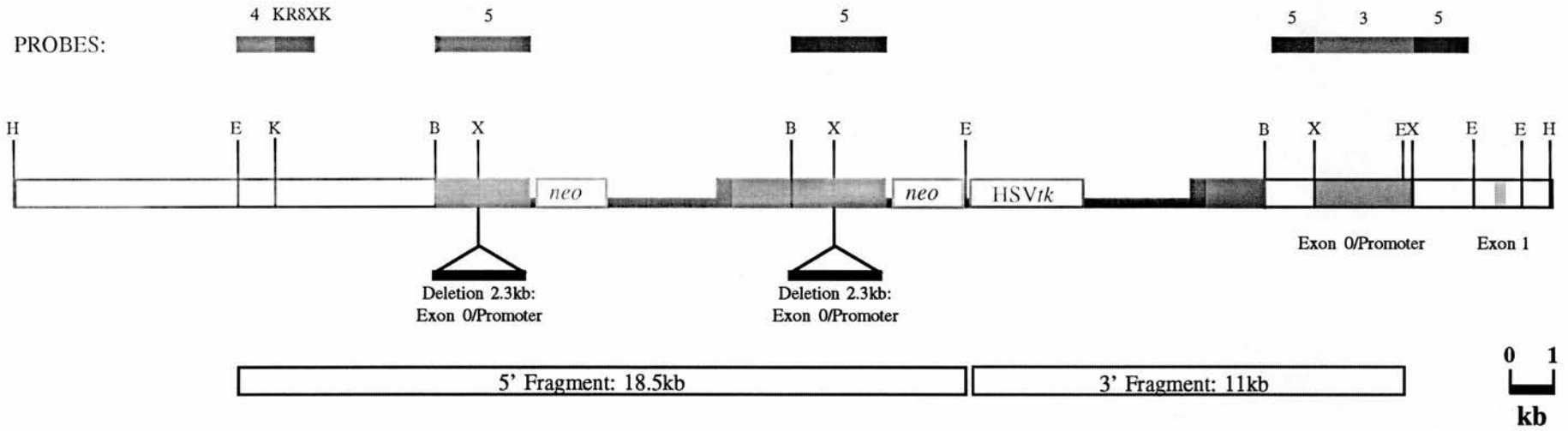
2) This large fragment is also detectable with the internal probe.

3) A fragment larger than wild-type (8.7kb) and consistent with a targeted modification of the gene, is detectable by the internal probe and the external probe located within the region of homology at the 3' end of the locus.

4) The large 5' fragment has no copy of the HSV*tk* gene, and therefore recombination events leaving the fragment in place will produce clones which are $GANC^r/G418^r$, as observed.

5) The structure has a triplication of homology, and therefore multiple reversion possibilities exist.

Proposed Structure of K-*ras* Clone P66



Of the clones for which a result was obtained, 55 (61%) appeared to be wild type, and 35 (39%) retained the 18kb band at the 5' end of the P66 targeted locus. As may be seen from consideration of figure 3-10, there are more possibilities for intra-chromosomal recombination within the P66 locus than in the simpler structure originally predicted. This is because the K-*ras* locus is thought to contain a triplication of the regions of homology, and also contains duplications of homology due to the sequence of pUC and *neo* from the targeting cassette. Some of the possibilities are irrelevant, as they do not result in loss of the HSV*tk* gene, for example between the two *neo* genes. There are four groups of possible combinations for recombination. The first possibility is a resolution to wild-type, and is a result of recombination between the outer two lengths of homology recombining 5' to the intended deletion. Second, a hit and run deletion, as desired, can be produced by recombination between the outer two lengths of homology 3' to the *Xba* I site which forms the 3' boundary of the deletion. The third possibility is the production of a modified locus with a single large *Eco*R I fragment, observable by both internal and 5' external probes. This is 17-18kb in size, and as such is likely to be unresolvable from the 18kb fragment of P66 by agarose gel electrophoresis. There are two possible variants of this structure, depending on the site of cross-over, but both produce a clone with a similar Southern analysis pattern because there is an *Eco*R I site at the extreme 3' end of the *Xba* I fragment comprising the deleted region. Both of these possible structures are generated by recombination between the middle and 3' homology repeats, or between the duplications of the pUC plasmid sequence. These two additional possible recombinant structures are shown in figure 3-11. The fourth and final possible recombination event which could occur and result in a $GANC^r$ clone is the deletion of the remaining HSV*tk* gene by recombination of its poly-A sequence with that of the adjacent *neo* gene.

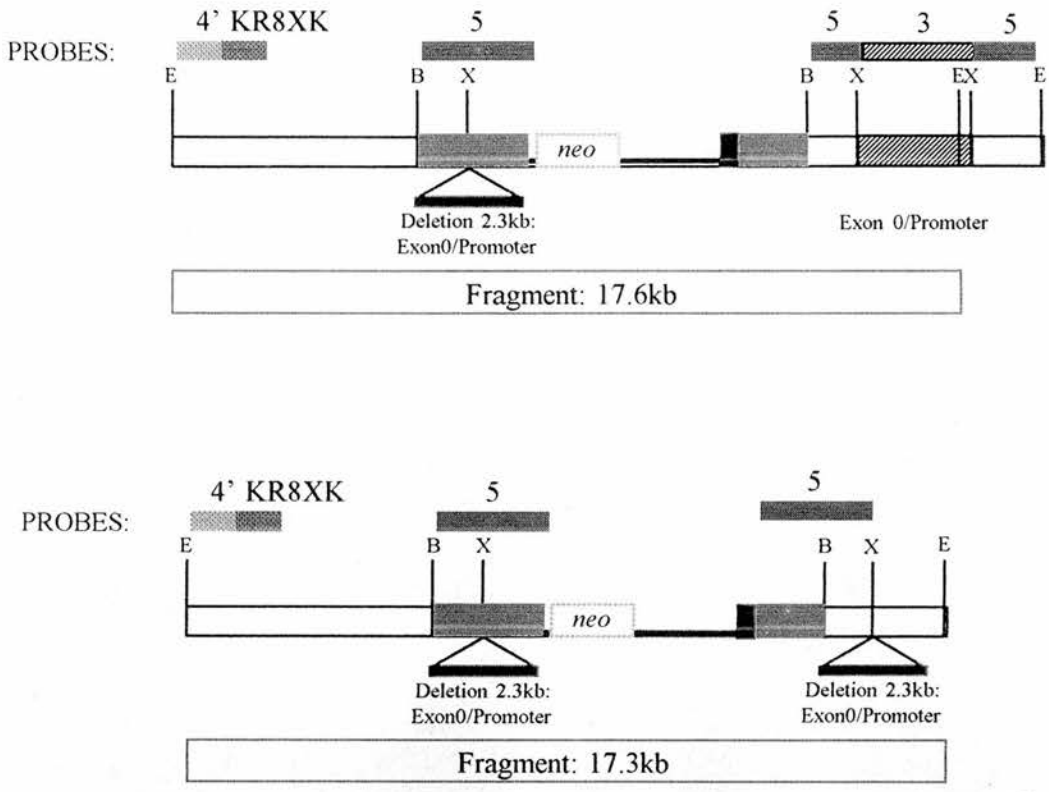
Figure 3-11: Alternate Reversion Structures from Clone P66

There are two additional possible reversion events from the structure of clone P66 shown in figure 3-10, which may be generated by recombination between the 3' and middle triplications of homology, or between the duplication of plasmid sequences. These structures are shown opposite. It may be seen that both structures lack an HSV*tk* gene and are therefore GANC^r, and that each generates a large *Eco*R I fragment detectable with both the internal and 5' external probes and indistinguishable in size from the 5' fragment of clone P66. No other fragment, other than the wild-type 8.7kb fragment expected from the unmodified allele on the other chromosome, is detectable with the internal probe, i.e. the approximately 10kb 3' fragment seen in clone P66 with the internal probe is lost.

Table 3-4: Reversion of Clone P66

The table shows the possible reversion structures obtained from a back selection experiment upon clone P66, and the approximate frequencies expected for each structure should the probability depend exclusively upon the length of homology involved in the recombination event generating that structure.

'igure 3-11



■ble 3-4
-ssible Reversion Events from Backselection of Clone P66

Event	Homology (kb)	% Expctd.	% Obsvd.
Resolution to Wild Type	2.4	21.8	61
Hit and Run Deletion	1.1	10.0	0
neo$^+$ Structure with large *Eco*RI Fragment	7.1 (6kb + 1.1kb)	64.5	39
Deletion of HSV*tk* gene	0.4	3.6	0

Thus it is concluded that on backselection of P66, events of type 1 and 3 are observed, but not events of type 2 or 4. Table 3-4 gives the approximate length of homology available in the P66 structure over which a recombination event would generate each type of reversion event delineated above. The penultimate column shows the expected frequencies for each type of event if it is assumed that each has a probability proportional to the length of homology involved in producing the required crossover. The final column shows the results actually obtained by experiment, described in section 3.3.3.

The data do not closely fit with the predicted events. The most obvious possible reason is because the suggested structure of P66, although it is the only one which has been found to be consistent with the available data, has not been formally shown to be the true structure. Thus the results obtained from backselections on this clone could not agree with predictions because of inaccuracies in the supposed starting structure. However, it is almost certainly naive to predict targeting frequencies based simply on lengths of homology, as it is well-known that many other factors play a part, for example whether the regions of homology recombining are isogenic (Te Riele *et al.,* 1992; Van Deursen & Wieringa, 1992; Deng & Capecchi, 1992). Special features of the primary sequence may also be significant, as is implied by the observation that targeting frequencies have been found to vary by three orders of magnitude between different genes (Camerini-Otero & Kucherlapati, 1990). Although 400bp is not below the limit at which intrachromosomal recombination has been shown to occur (Waldman & Liskay, 1988), such a short length of homology is known to be less efficient at promoting recombination (Liskay *et al.,* 1987). Thus a frequency sufficiently small as to be undetectable in this analysis for event (4), that is recombination between duplicated PGK polyadenylation sequences to delete the HSV*tk* gene, does not seem surprising. However, reversion to wild-type is favoured

over event (3), despite the length of homology available for such a cross-over being approximately only 1/3 the size. Event (3) may be selected against because it leaves a duplication of homology still present in the genome which is also capable of recombination, and the conditions used have strongly selected for clones in which homologous recombination events have occurred at this locus. Therefore there may also be selection for an event which spans the entire locus rather than just a part of it, since the machinery accomplishing the event is present at the locus and has a viable substrate on which to work.

Why are no hit and run deletion events observed? A measurable percentage, albeit not high, would be expected from the structure as drawn, as well as from the designed structure or indeed many other structures which could be produced by integration of this vector. This is difficult to answer. As previously remarked, branch migration of the crossover is a common event during recombination. Should this occur in this case, with the Holliday Junction moving through the site of deletion, a large (2.3kb) region of single-stranded DNA is produced and in some cases this would be repaired using information from the other strand. This would reduce the frequency of backselected clones carrying a deletion, and also has the effect of enriching for the reversion to wild-type event (as observed).

3.4.3 High Cell Density Back Selection of Clone P66

Results obtained from this population of clones more closely resembles those obtained from the Fast-Track protocol than the clones obtained by backselection of Clone P66 at a low cell density. In both of these experiments the cells were selected at a sufficiently high density that metabolic cooperation could occur. Because a widely differing population of revertant clones was obtained by backselection of clone P66 depending on the plating density, it was concluded that metabolic co-

operation during the selection significantly changes the nature of the selection (and hence the population of surviving clones).

Clones characterised as showing a wild type pattern constituted the largest single group of clones in this population, as might be expected from the discussion of the other backselection on this clone described in section 3.4.2 above (39%-44%).

3.4.3.1 Random Deletion

Random deletion (unfaithful excision) is a strong candidate as a mechanism for removal of the HSV*tk* gene under conditions of metabolic stress caused by backselection at a high cell density. The *Eco*R I site is immediately adjacent to the HSV*tk* gene and as such would be expected to be lost in the majority of such deletions, resulting in an *Eco*R I fragment of any size between 2-3kb and 30kb. As there are two *neo* genes present in the locus, it is quite feasible that the clone remain *neo*$^+$ after such an event. However one would not expect clones which have undergone such deletions to appear wild type when analysed with the external probe as a rule. This is theoretically possible, in that the deletion may span from 5' of the site of hybridisation of the external probe to the HSV*tk* gene, but this is unlikely from a statistical point of view to be the most common event; for it to be so would require the HSV*tk* gene to favour deletion in the 5' direction and deletion to frequently initiate within the gene or plasmid sequence so that a band of unpredicted size is seen with the internal probe. The result obtained for one clone, comprising a class 4-type event described in section 3.3.2, is consistent with this mechanism.

3.4.3.2 Re-Integration Events

Pickup events, where a target sequence recombines with a vector and then re-integrates into the genome elsewhere have been recorded in the past (Adair *et al.*, 1989), and extrachromosomal events involving the association of homologous and

non-homologous recombination events have recently been described (Sakagami *et al.*, 1994). It appears that in clones of classes 2 and 3 described in section 3.3.2 and table 3-1, (that is wild-type by external probe but still having another band by one or both of the internal probes 5 and *neo*), some vector sequence has re-integrated into the genome at another locus, following excision of a region of sequence incorporating the HSV*tk* gene. This conclusion was reached because the results obtained with the external probe for the vast majority of these clones indicate a wild-type structure, suggesting that a recombination event occured to regenerate the wild-type K-*ras* locus. However, the internal probe indicates that vector sequence is still present. The fragment observed is of a random size. Simple deletions are unlikely to account for all of these clones because the diagnostic *Eco*R I site is immediately adjacent to the HSV*tk* gene and therefore in the majority of cases a band shift would also be observed with the 5' external probe as well. Very large deletions encompassing the entire region containing sequence homologous to the external probe would usually be expected to encompass the regions of homology to probe 5, and in these case the clones could no longer be neo^+ (as is the case with clones of class 2) and yet fail to generate a detectable signal with the internal probe.

3.4.4 Fast Track Hit and Run with pKiD0PNT

3.4.4.1 Screening of clones HKiD01-HKiD058

58 clones were obtained from this selection, as described in section 3.3.5.2. Overall they were similar in character to those obtained from the high density backselection of clone P66, but slightly more heterogeneous (as expected), and with certain significant differences. 16 clones were defined as wild type, and another 18 clones were wild type as far as defined, by external probe(s) and *neo*. Thus whilst some of the 18 incompletely characterised clones were probably derived from random integration events which were neo^-, it appears that the largest single group of clones

obtained from this selection are wild type (27.5%-58.6%). A significant difference between this group of clones from the comparable group produced by backselection of clone P66 is that they are derived from uncharacterised precursor cells: all that is known about the cell from which each clone obtained in this backselection is derived is that it is *neo*$^+$. Therefore it is not known for each wild type clone obtained here whether it is truly wild-type, having integrated the vector at the target locus and then regenerated the wild-type locus by homologous recombination, or whether a random integrant has excised the vector sequence by a deletion mutation so that it generated no signal when analysed with the internal probe.

Most of the remaining clones obtained are wild-type when studied using the external probe, but have one or more extra bands observed using the internal probes (table 3-3: classes 3-6 inclusive). They may or may not be *neo*$^+$. It is possible to account for all of these clones as random integrations of the either vector or some part thereof which have subsequently undergone point mutation or deletion to inactivate the HSV*tk* gene. However, they could also be derived from targeted cells which have undergone deletion/re-integration events similar to those considered in section 3.4.3.2.

3.4.4.2 Fast Track Hit and Run Does Enrich for Events at the Target Locus

No hit and run deletion events were obtained in this experiment, as was the case with both back selections on clone P66. However, three clones produced patterns similar to that seen for some of the clones obtained in the previous backselection carried out on a known clone (clone P66). Clones HKiD054, HKiD056 and HKiD057 (classes 6, 7 and 8, respectively) show a large band with the 5' external probe similar to that seen with P66 and many of its derivative clones. This shows that an insertion event has occurred at the target locus and that the clone is derived from a cell in which

deletion of its functional copy of the HSV*tk* gene has occurred. However consideration of the internal probe results for these clones shows differences between these clones and their conventionally-produced counterparts. First, the internal probe results are all different, showing each to be an independent clone; thus the enrichment for recombination events at the target locus is increased (5.2% compared to 1.7%). The other major difference is that the internal probe pattern does not correspond with that of the external probe: the internal probe results are consistent with a wild-type locus (HKiD054) or show one other smaller band (HKiD057), or more than one other band (HKiD056).

This may indicate a problem with transfer of larger fragments to the filter on the Southern used for the internal probe, which could be the case for HKiD054. However, the observed hybridisation by the internal probe to other clones in this group argues against this, rather that the event occurring in these cells differs significantly from that proposed to have occurred in clone P66. Because they are undetectable by 5' external probe, the subsidiary bands seen are either at the 3' end of the locus, or are elsewhere in the genome, perhaps present as the result of pickup events. They may represent additional integrations of the vector into the genome at random.

The exact nature of the event(s) occurring at the K-*ras* locus is difficult to resolve. It is difficult to conceive of an event which produces a pattern so strikingly similar to that seen on clones known to have undergone recombination at the intended locus, albeit in an unpredicted event, and yet which does not appear to react with the internal probe. This question cannot be answered categorically because the only information reliably pertaining to the K-*ras* locus is that provided by the external probe. Data generated using the internal probe may reflect the occurrence of events elsewhere in the genome. However, the fragment size observed with the 5' external

probe suggests an event similar to that occurring in clone P66, but clearly with further complications: possibly a large deletion of the region generating a large fragment containing no homology duplications. In a similar way to the majority of clones obtained in the high cell density backselection of clone P66, the variably-sized bands visible in the internal probe analysis are possibly due to pickup events, but the possibility also remains that the entire locus has rearranged and created two different fragments visible by internal or external probe.

3.4.5 High Cell Density in Backselection

In the backselections described in sections 3.3.2 and 3.3.5.2 the higher plating density of cells present in the selection introduces other variables into the system. Cells able to survive such selection conditions which eliminate their neighbours are unlikely to be behaving in a similar physiological manner to the wild-type. By virtue of their cellular environment, e.g. the position in the cell cycle, or accumulation of a mutation, they may be more prone to undergoing changes than normal ES cells and thus as well as surviving selection by loss of the HSV*tk* gene other changes permit escape from "kiss of death", and incidentally produce a small but diverse group of clones whose analysis suggests uncommon events. That these clones have arisen from a population that in some way deviates from the normal ES phenotype is also suggested by the fact that of the relatively few chimaeras created from these clones, the level of chimaerism observed was not large (less than 10% by coat colour estimation) and germline transmission of the allele was not achieved. It is therefore likely that the cells surviving the selection were either differentiated or genetically compromised, and the selection conditions enriched for a non-pluripotent subset of cells which have undergone unpredicted events (Sheardown & Hooper, 1992). It is generally regarded that for optimum transfer of the modified allele through the

germline, ES cells should be at their lowest possible passage number and that they should not be exposed to suboptimal growth conditions (Robertson, 1987).

3.4.6 Chimaera Studies

After initial internal blot data were obtained for clone P66/32 and clones HKiD010, HKiD052 and HKiD055, which were initially thought to be targeted hit and run deletions, some chimaeras were made whilst further verification of the structures was sought. Unfortunately these results showed that the clones were not K-*ras* exon 0 deletion mutants and chimaera production was terminated. However, the chimaeras were not killed until after a year, when they were then examined. Few chimaeras were obtained, and those which were obtained were not highly chimaeric and did not pass on their ES cell-derived mutation to their offspring as discussed above.

Although three out of five chimaeras had neoplastic malignancies in evidence upon pathological examination, no significance to the status of the animal's K-*ras* gene could be inferred from these tumours.

3.5 Conclusion

An attempt has been made to introduce a deletion mutation removing the exon 0 region from the K-*ras* gene of the mouse by hit and run gene targeting in ES cells. Although events were observed at the K-*ras* locus itself, these occurred at a low frequency and no event of the type expected was obtained. This may represent an effect of the K-*ras* locus itself on the recombination process, for example being located in a poorly recombinogenic region of the genome, or flanking sequences complicating the selection by downregulating *neo* expression so that there is selection for more than one *neo* gene to become incorporated. An insertion event (clone P66) was observed at the K-*ras* locus, but this was an unpredicted event and backselection of this clone failed to produce a hit and run deletion. The event producing clone P66

appeared to involve a concatenation and extra-chromosomal recombination event prior to the integration of the vector into the K-*ras* locus.

Backselection at a cell density high enough for the ES cells to participate in metabolic cooperation was found to select for a small subpopulation of clones which have undergone multiple events in the genome, thus producing a disparate collection of clones many of which had one or more fragments of the vector incorporated in their genomes. A low density of cells in backselection produces a clearer picture due to the eradication of such complications but was still not able to generate a hit and run mutation.

A "fast track" protocol based on backselection of all *neo*$^{+}$ clones obtained from the initial step with the objective of enriching for those clones which have undergone the rare event of a targeted insertion was also unable to produce the desired mutation, but it was shown that some 5% of the clones obtained by this procedure had experienced a recombination event at the K-*ras* locus, suggesting that it may provide the basis of an improved selective procedure if undesired competing recombination events can be eliminated.

CHAPTER 4-

INTRODUCTION OF POINT MUTATIONS BY HIT AND RUN

Introduction of Point mutations By Hit and Run

4.1 Introduction

The impetus for the development of the hit and run technique is the desire to generate precise, small mutations in the genome which will allow the creation of animal models which are representative of human diseases, or permit studies of aspects of gene function *in vivo* not possible with a knock-out type experiment. To this end, two targeting vectors were used in experiments which aimed to introduce small mutations into the K-*ras* gene. The first, pKiTPNT, aimed to introduce a point mutation into codon 12 of the gene, and thereby introduce an activating gly→asp mutation of the type observed in malignancy. Vector pK4BΔPNT was designed to introduce a frame-shift into exon 4B prior to the polylysine domain and post-translational modification site, with the objective of effectively deleting this splicing variant as a functional molecule. Such experiments, which are designed to introduce smaller mutations, may additionally shed light on the behaviour of pKiD0PNT described in the previous chapter. A much more efficient targeting efficiency might suggest, for example, that the mutation to be introduced by a hit and run experiment affects its efficiency and that the method is not suitable for the creation of large deletions.

4.2 Targeting Vector pKiTPNT

This vector was constructed by Dr D. James Williamson in our group. It is derived from the pPNT targeting cassette (Tybulewicz *et al.*, 1991), described in chapter 7 and shown in figure 7-1, and includes a 2.8kb *Xba* I-*Kpn* I fragment of homology to the murine K-*ras* gene derived from the plasmid PBKT-142 (Guerrero *et al.* 1984). This fragment incorporates the murine exon 1 of K-*ras*, and carries a G→A mutation at position 35 which converts the wild type glycine residue to an aspartic acid

residue. Vector pKiTPNT is shown in figure 4-1, and the structures of the wild-type locus and the designed insertion event are shown in figure 4-2.

4.3 Targeting Vector pK4BΔPNT

4.3.1 Design of pK4BΔPNT

A distinctive feature of K-*ras* is that unlike its close relatives H-*ras* and N-*ras*, it has two splicing variants which differ in their C-termini. The protein is linked to the inner cell membrane by post-translational modifications at the C-terminus (see the Introduction, sections 1.2.2 and 1.2.4), but little is known about the functional differences of the two different K-*ras* p21 proteins. Part of the present study therefore attempts to address this question using two approaches. In chapter 5, an RT-PCR-based study is described which looks at the relative expression of the two alternative fourth coding exons in different tissues during development of the mouse. Secondly, targeting experiments are currently underway to delete one or other of the exons, so that animals and ES cells which are only capable of manufacturing one form of the protein may be studied. Dr D. James Williamson and Dr Charles Patek are carrying out an experiment to force the use of exon 4B by removal of exon 4A by replacement. Vector pK4BΔPNT is designed as a complementary experiment, where the exon 4A splicing variant is the only functional one.

As described in the introduction, section 1.2.4, $p21^{K\text{-}rasB}$ undergoes three post-translational modifications, namely farnesylation of the cysteine residue of the CAAX motif four amino acid residues prior to the C-terminus, followed by proteolytic removal of the terminal three amino acids and carboxymethylation of the new C-terminal substituted cysteine. These modifications are essential for the membrane localisation of $p21^{ras}$, which is essential for function (reviewed in section 1.2.4).

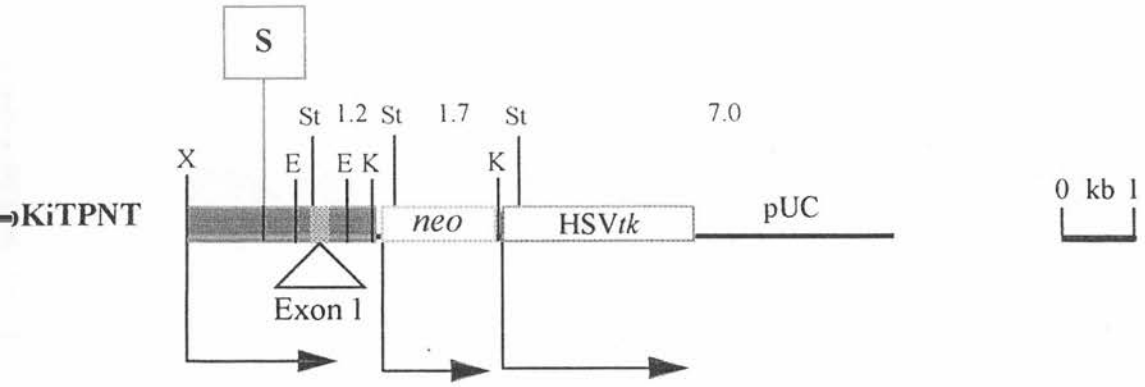
Figure 4-1: Targeting Vector pKiTPNT

a) The structure of vector pKiTPNT. The vector carries 2.8kb of non-isogenic homology (source: hybrid mouse of strains AKR-RF/J) to the region encompassing exon 1 of murine K-*ras* (grey block). The vector carries an activating gly→asp codon 12 activating mutation generated by a G→A transition at position 35 of exon 1. The vector is derived from pPNT (see figure 7-1 [b]) and as such includes a *neo* and an HSV*tk* gene (orientation as shown) and a bacterial plasmid (pUC 18, shown as a thick black line). Relevant restriction sites are shown: S, *Sal* I, which is the linearisation site of the vector; X, *Xba* I; K, *Kpn* I; E, *Eco*R I, St, *Stu* I. The sizes in kb of fragments from selected restriction digestions of the construct are shown.

b) Restriction enzyme digestion of vector pKiTPNT. Lanes:1, *Stu* I; 2, *Sal* I (linearised); 3, DNA molecular size marker.

Figure 4-1

(a)


S
St
1.2
St
1.7
St
7.0
X
E
E K
K
KiTPNT
neo
HSVtk
pUC
Exon 1
0 kb 1

(b)

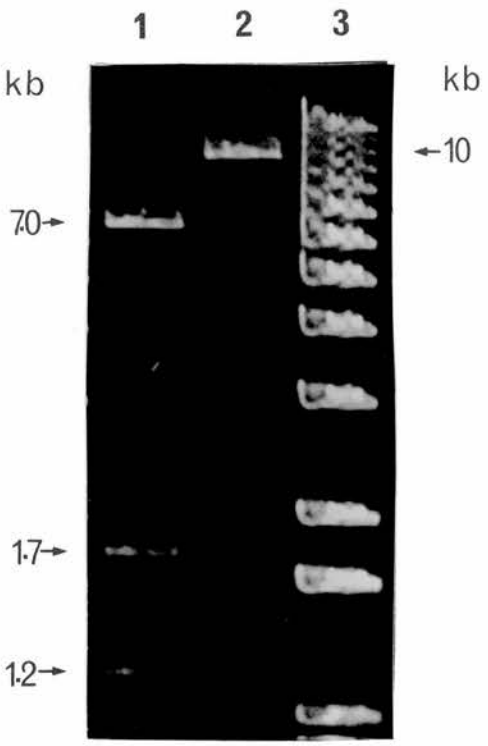

1
2
3
kb
kb
10
7.0
1.7
1.2

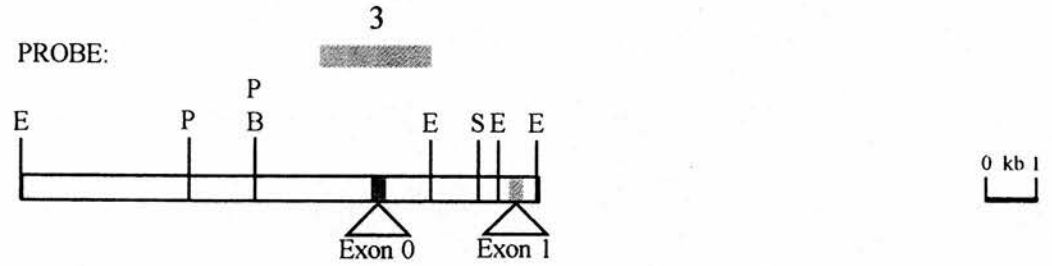
Figure 4-2: Targeting Structures for Vector pKiTPNT

The figure shows, to scale, the structure of the wild-type murine K-*ras* locus in the exon 0-exon 1 region, and the predicted insertion structure introduced at the locus with vector pKiTPNT. Because the introduced mutation is so small, at the scale shown the wild-type and hit and run mutant structures are identical. Probe 3 is shown. Selected restriction sites are also shown: B, *Bam*H I; E, *Eco*R I; S, *Sal* I; K, *Kpn* I; N, *Not* I; P, *Pvu* II.

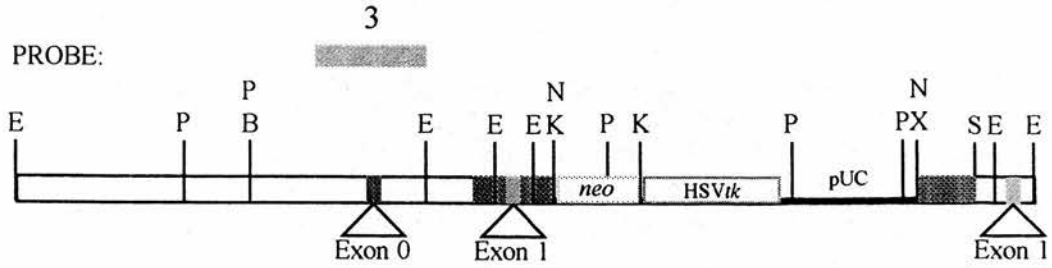
The expected results for Southern analysis of targeted clones obtained with this vector, using probe 3 are shown below:

Restriction Digest	**Wild-Type/"Run"**	**"Hit" Insertion**
Pvu II	9.5kb	6.5kb
*Bam*H I	>12kb	8.5kb

Wild-Type/Hit and Run Locus



Insertion Structure



The structure of the murine K-*ras* gene (George *et al.*, 1985) is similar to that of the human c-K-*ras*2 gene (Shimizu *et al.*, 1983b, McGrath *et al.*, 1983 & Capon *et al.*, 1983), and is shown schematically in the introduction. As may be seen from figure 1-10, deletion of the entire exon 4B by replacement as planned for exon 4A is not an appropriate strategy because immediately 3' to the exon 4B coding sequences are the K-*ras* untranslated sequences, including the polyadenylation signal, and removal of part or all of these sequences would also affect the behaviour of the exon 4A-containing mRNA in addition to that of exon 4B.

Taking these two considerations together, hit and run was therefore selected as the most appropriate strategy for this experiment. A small mutation introducing a premature stop codon into exon 4B prior to the poly-lysine string and the CAAX motif could be expected to functionally ablate the exon 4B variant $p21^{ras}$ protein by preventing correct localisation of the molecule at the cell surface (see introduction). The exon 4A-containing mRNA should be unaffected because no change has been made to any of the untranslated sequences or the size of the mRNA, and therefore the mRNA can be expected to behave identically to the wild-type with respect to stability and polyadenylation.

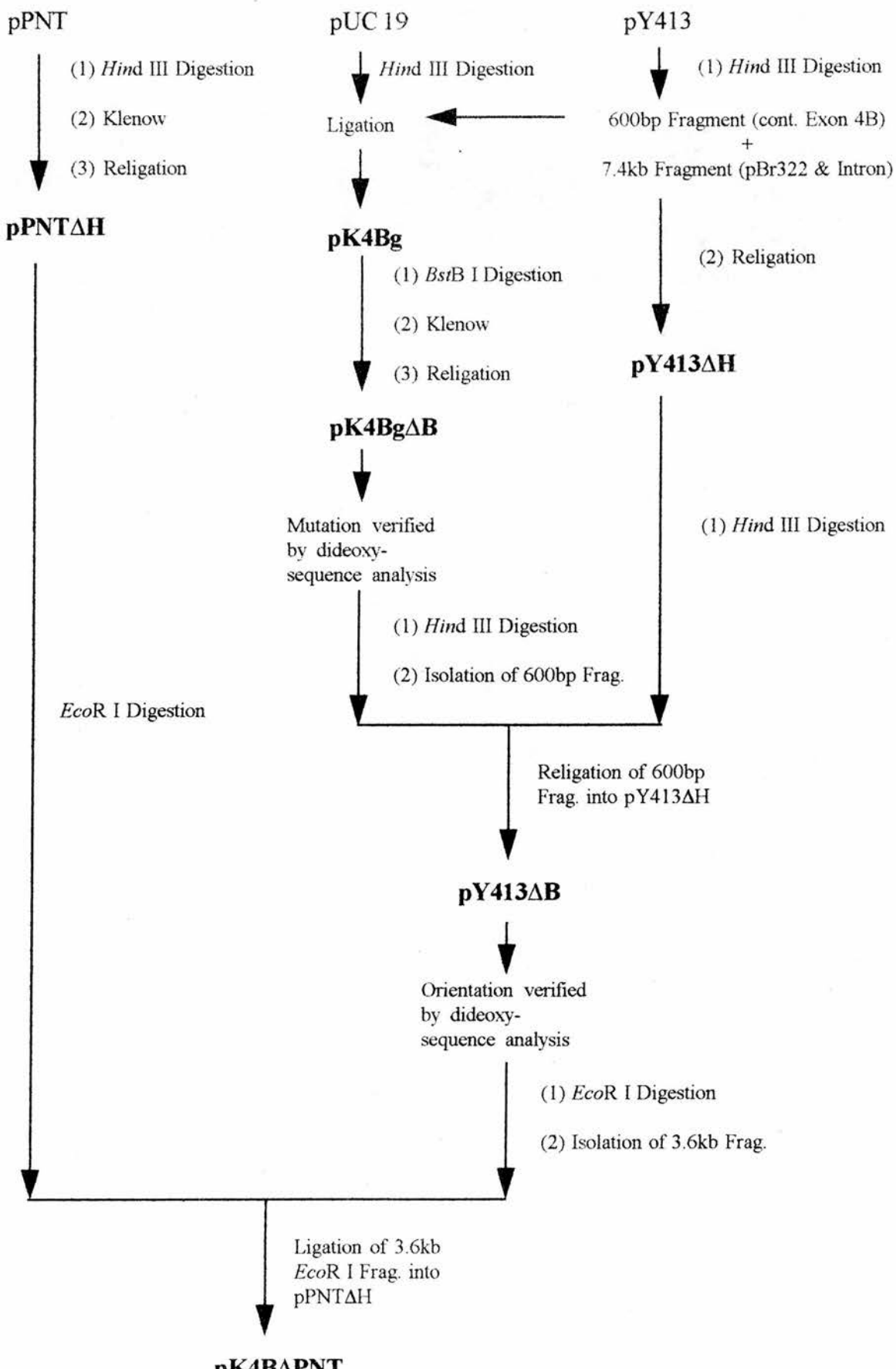
4.3.2 Construction of pK4BΔPNT

Vector pK4BΔPNT was constructed as shown in figure 4-3. It is derived from the pPNT targeting cassette (Tybulewicz *et al.*, 1991), and vector pY413 (George *et al.*, 1985) which carries 3.6kb genomic sequence incorporating the murine K-*ras* exon 4B. The unique *Hin*d III site in pPNT was removed by digestion of the construct with *Hin*d III, filling in of the cohesive ends by Klenow enzyme and then re-ligating the ends. This permitted the use of the unique *Hin*d III site in the homology as a linearisation site prior to electroporation of the construct.

Figure 4-3: Construction of Targeting Vector pK4BΔPNT

A flowchart showing the steps involved in the construction of the targeting construct designed to introduce a missense mutation into exon 4B of murine K-*ras*, pK4BΔ PNT. The vector homology is derived from pY413 (George *et al.*, 1985) and the selectable markers and bacterial plasmid sequences from pPNT. The mutation was introduced at a *Bst*B I site near the start of exon 4B, and a small fragment containing this exon was subcloned out of pY413 for the mutagenesis process because there is a second such site in the homology region. This fragment was then replaced to restore the homology region in vector pY413, and the complete, modified length of homology was then cloned into pPNT lacking a *Hind* III site to generate the targeting construct.

Figure 4-3


pPNT
pUC 19
pY413
(1) Hind III Digestion
Hind III Digestion
(1) Hind III Digestion
(2) Klenow
Ligation
600bp Fragment (cont. Exon 4B)
+
7.4kb Fragment (pBr322 & Intron)
(3) Religation
pPNTΔH
pK4Bg
(2) Religation
(1) BstB I Digestion
(2) Klenow
(3) Religation
pY413ΔH
pK4BgΔB
Mutation verified by dideoxy-sequence analysis
(1) Hind III Digestion
(1) Hind III Digestion
(2) Isolation of 600bp Frag.
EcoR I Digestion
Religation of 600bp Frag. into pY413ΔH
pY413ΔB
Orientation verified by dideoxy-sequence analysis
(1) EcoR I Digestion
(2) Isolation of 3.6kb Frag.
Ligation of 3.6kb EcoR I Frag. into pPNTΔH
pK4BΔPNT

The small *Hind* III fragment of pY413 was removed by digestion with *Hind* III and religation to generate the construct pY413ΔH, and the construct pK4Bg was generated by the subcloning of the small *Hind* III fragment so produced into pUC19. This smaller *Hind* III fragment contains K-*ras* exon 4B and only has one *Bst*B I site which is present in the exonic sequence. pK4Bg was digested at this site with *Bst*B I, treated with Klenow enzyme and then religated, to generate pK4BgΔB which lacks the *Bst*B I site. Dideoxysequencing analysis showed the loss of a T residue in the site, which resulted in a frame-shift giving a premature stop signal as shown in figure 4-4. The small *Hind* III fragment was then cloned back into pY413ΔH at the *Hind* III site to regenerate the genomic fragment contained in pY413, except that a point mutation in exon 4B is now present at the *Bst*B I site. Dideoxysequencing was used to check that the orientation of the fragment was correct. The 3.6kb genomic fragment from pY413ΔB was excised with the enzyme *Eco*R I and cloned into pPNT at the unique *Eco*R I site to produce vector pK4BΔPNT. Vector pK4BΔPNT is shown in figure 4-5, together with the wild-type and insertion locus structures.

4.4 Experiments with pKiTPNT

With the objective of generating an activating point mutation, vector pKiTPNT was introduced into E14 ES cells. 150 clones successfully screened with probe 3 showed no correctly-targeted insertion events, indicating a low targeting frequency with this vector.

Because of this low targeting frequency, at the same time as the comparable experiment with vector pKiD0PNT described in section 3.3.5.2 was carried out, the vector was re-electroporated into ES cells and the approximately 1500 clones so obtained were pooled, expanded to $2x10^8$ cells and backselected in ganciclovir at a

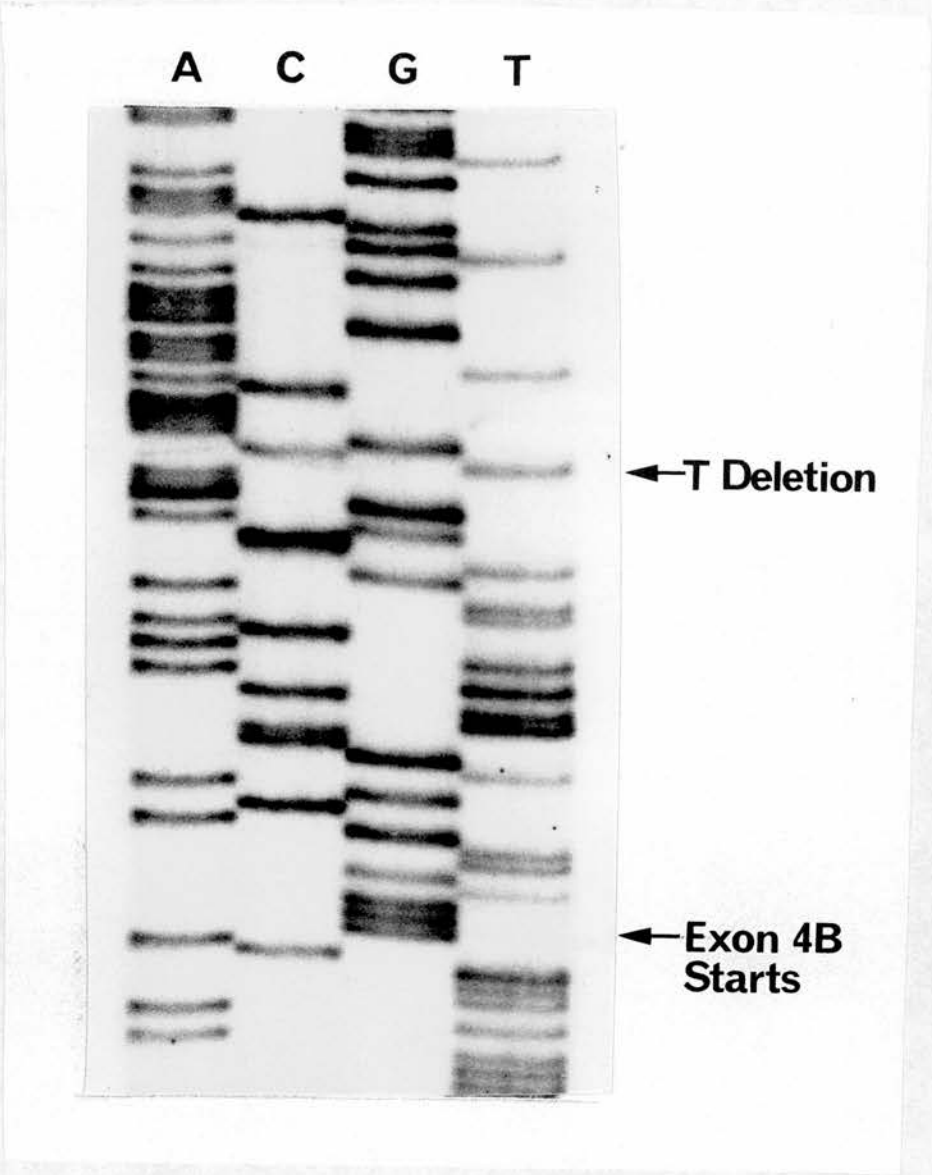
Figure 4-4: Point Mutation to be Introduced into K-*ras* Exon 4B

a) Dideoxysequence analysis of the start of murine K-*ras* exon 4B sequence, as present in vector pK4BΔPNT, showing the deletion of a thymidine at the *Bst*B I site of the wild-type gene.

b) The introduced mutation results in a frame-shift leading to a premature termination of translation of the K-*ras* exon 4B splicing variant prior to the essential post-translational modification domains.

Figure 4-4

(a)


A
C
G
T
T Deletion
Exon 4B
Starts

(b)

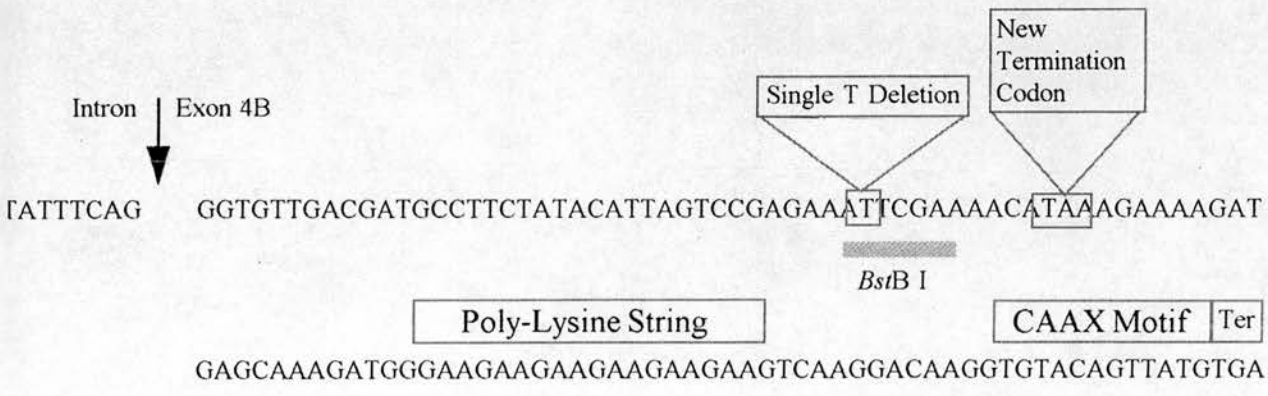
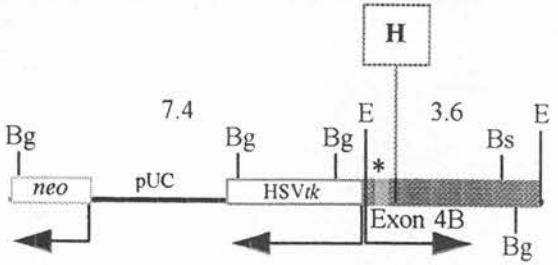

Intron
Exon 4B
Single T Deletion
New
Termination
Codon
ΓATTTCAG
GGTGTTGACGATGCCTTCTATACATTAGTCCGAGAAATTCGAAAACATAAAGAAAAGAT
BstB I
Poly-Lysine String
CAAX Motif
Ter
GAGCAAAGATGGGAAGAAGAAGAAGAAGAAGTCAAGGACAAGGTGTACAGTTATGTGA

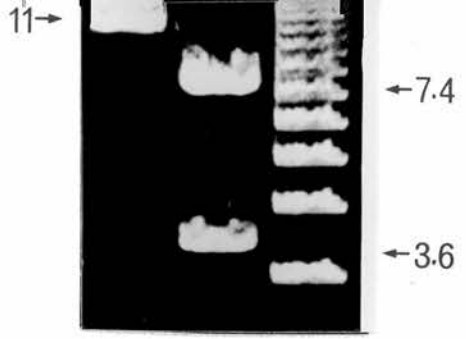
Figure 4-5: Vector pK4BΔPNT and its Targeting Structures

The figure shows, to scale, the structure of vector pK4BΔPNT, the wild-type murine K-*ras* locus in the exon 4B region, and the predicted insertion structure introduced at the locus with vector pK4BΔPNT. Because the introduced mutation is so small, at the scale shown the wild-type and hit and run mutant structures are identical. To the right of the vector structure is shown a restriction enzyme digestion of pK4BΔPNT. Lanes:1, *Hin*d III (linearised); 2, *Eco*R I; 3, DNA molecular size marker. As with the other vectors designed to target K-*ras* described in this thesis, pK4BΔPNT is based on pPNT and as such has a *neo* gene as a positive selectable marker and an HSV*tk* gene as a negative selectable marker (orientation as shown). The asterisk indicates the site of the point mutation. The sizes in kb obtained by *Eco*R I restriction enzyme digestion of pK4BΔPNT are shown.

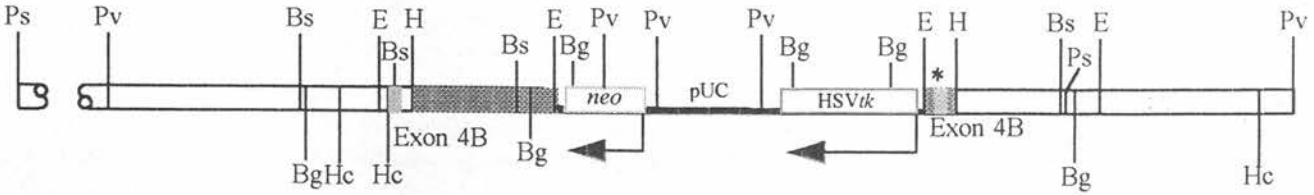
Selected restriction sites are shown: Bg, *Bgl* II; E, *Eco*R I; H, *Hin*d III; Bs, *Bst*B I; Ps, *Pst* I; Pv, *Pvu* II; Hc, *Hin*c II.

pK4BΔPNT

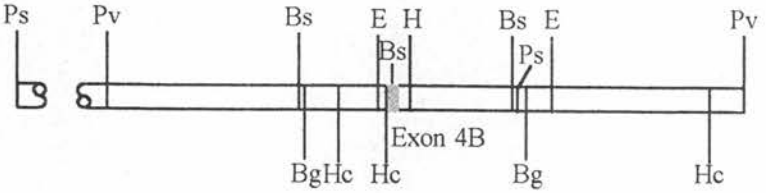




Insertion Structure



Wild-Type/Hit and Run Structure



density of $5x10^6$ cells/100mm plate. 29 clones designated HK1-HK29 inclusive were obtained from this experiment.

4.4.1 Screening of Clones HK1-HK29

The original objective was to screen these clones by Southern Analysis using probe 3 on separate *Pvu* II and *Bam*H I digests as above, but this probe is unreliable because of a relatively high level of non-specific hybridisation in comparison to specific hybridisation. Alternative screening strategies were therefore devised. However, use of the arms of the vector homology was unsuccessful because of very high background hybridisation. Faint and non-specific hybridisation was observed to an exon 1 genomic probe. No signal was obtained using the K-*ras* cDNA as a probe or the *Kpn* I-*Pst* I fragment derived from PBKT-142 as a 5' external probe to analyse a *Kpn* I digest. Oligonucleotide probing using the exon 1 PCR primer described in section 7.4.3 end-labelled with P-33 was similarly unsuccessful due to lack of signal at a high stringency wash (2xSSC at 56°C) or non-specific hybridisation at a low stringency wash (6xSSC at room temperature).

It was therefore concluded that the genomic sequence of K-*ras* in the region of exon 1 is refractory to reliable Southern analysis because of poor probe specificity. For this reason, a two-part strategy based on the PCR was adopted:

1) Screening for *neo* and the diagnostic mutation to test for the presence of vector sequence and the mutant K-*ras* exon 1 sequence in the genome.

2) Clones designated *neo*$^+$ in part 1 were then subjected to RT-PCR of K-*ras* RNA followed by the diagnostic PCR to ascertain presence or absence of the desired mutation at the K-*ras* locus, and confirming the generation of a functional mRNA product carrying the mutation from the gene.

Part 1 of the screen was carried out as described in section 7.3.6.3. Duplicate samples of each clone's genomic DNA were run in parallel, one of which was spiked with an equal quantity of genomic DNA from a clone known to be *neo*$^+$. Thus false negatives due to poisoning of the PCR by contaminants present in a particular DNA sample were avoided. The diagnostic PCR designed to observe and follow the point mutation was also run on each sample.

The second part of the screen relies on the fact that genomic DNA will not be susceptible to the PCR because the introns make the fragment to be amplified too large for the *Taq* polymerase enzyme (approximately 40kb). Thus the analysis works at the level of transcription and a positive result indicates a clone which carries the mutation and is correctly transcribing and splicing the gene with the desired mutation. For this analysis, ES cells from each clone were grown and RNA prepared and reverse transcribed as described for the analysis of splicing variant expression described in chapter 5. The PCR amplifying the K-*ras* cDNA was run on the sample. An aliquot was run on a 1.2% agarose gel and a plug of agarose from the K-*ras* cDNA band was taken with a Pasteur pipette and blown into 500µl ddH_2O and incubated at 55°C for 10 minutes. 1µl of this dilute solution of RT-PCR product was then used as template for the codon 12 mutation PCR described in section 7.3.6.2. The results obtained from this screen are shown in table 4-1.

From table 4-1 it may be seen that unfortunately no correctly-targeted clones carrying an activating mutation at codon 12 of exon 1 of the K-*ras* gene were obtained from this experiment.

Table 4-1

The table shows the results of the screening of clones HK1-29, obtained by a fast-track hit and run protocol with vector pKiTPNT, by a two-part PCR strategy.

Table 4-1

Screening of Codon 12 Activating Mutation Clones HK1-HK29

Genomic Screen	Wt/*Neo*-	Mutant/*Neo*	Wt/*Neo*+	Mutant/*Neo*+
N^{o} Clones	4	2	12	6
(There were additionally 4 clones carrying the mutation, and 1 wild-type, which were uncharacterised with respect to *neo*)				
RT-PCR Tested	2	2	4	6
RT-PCR: WT	2	2	4	6
RT-PCR: Mutant	0	0	0	0
(All 4 clones carrying the mutation, but with an unknown status with respect to *neo*, were also shown to be wild-type at the K-*ras* locus by RT-PCR)				

4.5 Experiments with pK4BΔPNT

4.5.1 Metabolic Cooperation In Ganciclovir Backselection

This vector, constructed as described in section 4.3, was electroporated into HM1 ES cells, and submitted to the fast-track protocol as described. However, backselection was carried out at different cell densities to monitor for distortion of the results by metabolic cooperation, as was suspected to be the case in experiments described previously. Cells were therefore plated out at different densities of cells per plate and the number of colonies obtained was counted. The results of this experiment are shown in table 4-2, and graphically in figure 4-6.

It was concluded that approximately 10^5 cells per 100mm plate is optimal for selection in ganciclovir, providing a workable compromise between the number of plates required for an experiment and the number of clones obtained, and this cell density was accordingly used for the repeated backselection of clone P66 described in section 3.3.3.

4.5.2 Screening of Clones obtained with Vector pK4BΔPNT

188 clones were picked from the backselection described above, from the low cell-density plates. To minimise labour in the screening process, the initial screen used was somewhat similar to that finally used for the clones obtained with vector pKiTPNT, namely RT-PCR of K-*ras* followed by enzymatic digestion of the product with *Bst*B I and agarose gel electrophoresis to separate the products. Loss of the *Bst*B I site would indicate a successful hit and run event producing a transcribed mRNA with a mutation in exon 4B as designed. The K-*ras* allele on the other chromosome provided a control for enzyme activity. Clones carrying such a mutation could then be further screened to check for the presence of anomalous events, by Southern analysis, dideoxysequencing or PCR as necessary, but of the 67 clones

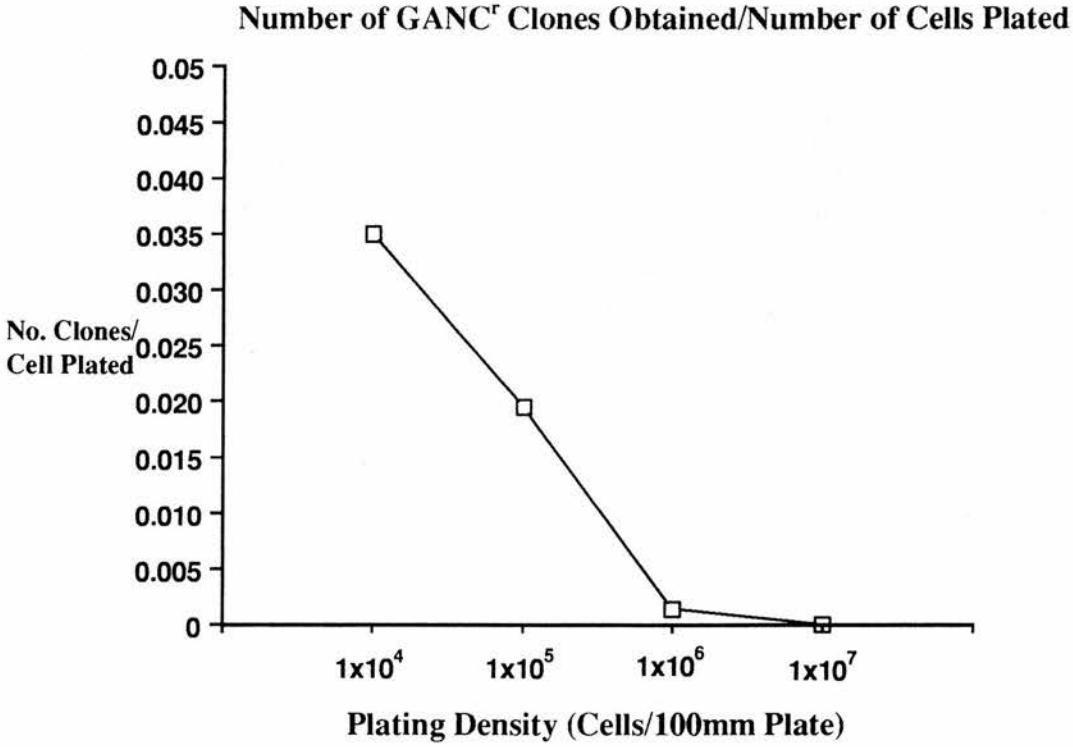
Table 4-2 and Figure 4-6: Variation in Cloning with Increasing Cell Density in the Presence of Ganciclovir

HM1 ES cells were transfected with vector pK4BΔPNT and subjected to G418 selection. Clones were pooled and the cells expanded to $2x10^8$ cells. The table shows the number of clones obtained by plating *neo*$^+$ cells at different cell densities per plate and subjecting the plated cells to ganciclovir selection. These data are also presented in figure 4-6 as the number of GANC^r clones obtained per original cell plated.

Table 4-2

Nº Cells/100mm Plate	Nº GANC^r Clones Obtained	Mean Nº of Clones/Cell plated
$1x10^4$	376 324	$3.50x10^{-2}$
$1x10^5$	2272 1620	$1.95x10^{-2}$
$1x10^6$	1248 1576	$1.41x10^{-3}$
$1x10^7$	0 0	0

Figure 4-6



successfully screened by this method none was found to be carrying the desired mutation.

4.6 Discussion

4.6.1 Targeting Efficiency of Vector pKiTPNT

The targeting efficiency of this vector is clearly low, because no correctly-targeted insertion events were observed. This is probably due to the nature of the homology carried by the vector. The length of homology is 2.8kb, which is relatively short by modern standards; many workers are currently using homologies in the region of 10kb in targeting constructs to optimise targeting efficiencies (see, for example, Rudolph *et al.*, 1993a, 1993b; Ernst *et al.*, 1994; Gorry *et al.*, 1994). A marked increase in targeting frequency was observed when the length of homology in the targeting vector was increased from 1.3kb to 6.8kb (Hasty *et al.*, 1991a). However, 2.8kb has been shown to be sufficient for measurable gene targeting to occur (Adair *et al.*, 1989; Schulman *et al.*, 1990; Hasty *et al.*, 1991a; Deng & Capecchi, 1992; Scheerer & Adair, 1994).

Targeting efficiency is also affected by small divergences in homology. It has been shown that targeting constructs built with homology regions derived from the same strain of mouse as the ES cell to be targeted can be much more efficient than those made from non-isogenic DNA (Te Riele *et al.*, 1992; Deursen & Wieringa, 1992; Deng & Capecchi, 1992). This reflects polymorphic differences in untranslated sequence between the two strains, which reduce the maximum length of uninterrupted homology of the targeting construct by mismatches. Elimination of such differences by the use of isogenic DNA increases the effective length of homology, and hence targeting efficiency, of the vector. The murine homology in this

case is derived from an AKR-RF/J hybrid mouse, whereas the ES cells are 129/Ola-derived, and therefore this may be a significant factor in this experiment.

Target genes also vary widely in their recorded targeting frequencies (Camerini-Otero & Kucherlapati, 1990), for unknown reasons. This phenomenon may be related to characteristics of the target locus such as its chromatin organisation. There are some indications that the targeting frequencies of *ras* genes are low: Other work in our group, some of which is described in this thesis, suggests this is the case, as have informal discussions with a group involved in similar work (T. Jacks group; personal communication) Targeting of the N-*ras* locus was observed to be very inefficient (a targeting frequency of 1/5000-7000 transfected cells; Cases & Dautry, 1992; Umanoff *et al.*, 1995) when targeted by PNS-replacement. Efficient targeting of human c-K-*ras*2 was observed (29% $G418^r$ colonies) but was carried out using a PNS replacement vector with 6.7kb of homology and a promoterless *neo* gene as the positive selectable marker on a transformed human cell line and therefore the relevance to the present results is dubious (Shirasawa *et al.* 1993). An attempt to target the H-*ras* gene was also unsuccessful (Crombie, R., Balmain, A., Clarke, A. R. & Hooper, M. L.; personal communication).

4.6.2 Study of Clones Obtained By Fast-Track Hit and Run With Vector pKiTPNT

Many problems were experienced during the screening of the potentially-targeted exon 1 clones, which is why further electroporations of the vector were not carried out to generate additional clones and increase the probability of obtaining a hit and run activated K-*ras* gene. Since the locus appeared to be extremely refractory to successful Southern analysis, an alternative analysis examining the K-*ras* locus at the level of transcription, namely RT-PCR, was chosen as the best way of circumventing the problems experienced. As well as showing the presence or otherwise of the

desired alteration, it also shows that the modified gene is still operational and that functions such as splicing have not been impaired by other mutations introduced by the hit and run method.

Unfortunately no correctly-targeted hit and run clones have been identified from the 29 obtained. The largest group (12 clones) contain *neo* but are negative for the mutation both at the genomic and transcribed levels. These are most probably clones which have incorporated a proportion of the vector sequence at a random location in the genome, but have excluded the fragment of homology incorporating the mutation during the integration event, i.e. have not integrated the very 5' end of the construct. However the possibility of a targeted event which has failed to recombine correctly and produced a gene which fails to transcribe a stable mRNA is not formally excluded by this analysis. Random integration is also the most likely explanation for the second largest group of clones, namely those which have incorporated both the *neo* gene and the point mutation into the genome but fail to express the mutation at the level of mRNA. Four clones were completely wild-type by the analyses carried out. These may be either random integration events largely or completely excised by deletion or they may be hit and run events which have reverted to wild type. A hit and run event which has resolved to wild type appears less probable, however, given that there is no unequivocal evidence for a targeting event in any of the other groups of clones. Again, random integration is the most likely explanation for clones which are mutation positive at the genomic level but are *neo*$^-$, with a deletion removing the selectable markers.

4.6.3 pK4BΔPNT as a Hit and Run Vector

pK4BΔPNT was designed and constructed as a hit and run vector whose purpose is to introduce a frame shift mutation leading to a premature stop codon in exon 4B of the murine K-*ras* proto-oncogene.

An assay based on PCR followed by restriction endonuclease digestion, described in section 4.5.2, has been designed and optimised which permits the identification of the mutant murine gene.

Although only 37% of clones picked in this experiment were successfully screened no clones characterised by the desired event were noted, as determined by K-*ras* RT-PCR followed by restriction endonuclease digestion of the products and agarose gel electrophoresis. A total of 67 clones were screened, and it is therefore not possible to rule out low-frequency targeting as in the other two K-*ras* experiments described. The genomic clone pY413 was the only available clone at the time of vector construction, and therefore the length of homology within the constructed vector is probably suboptimal. The mutation is necessarily at the 5' end of the vector homology, so that most revertant cross-overs occurring in a targeted clone with an insertion structure of the type shown in figure 4-5 are predicted to favour the restoration of the wild-type exon. Approximately 8% of cross-overs between homology duplications of this structure, calculated from a comparison of relative lengths of homology on each side of the mutation, are predicted to leave the mutation in the genome. However, branch migration events have been shown to be capable of drastically modifying the frequency at which the mutation is observed in each of the homology duplications (Hasty *et al.*, 1991c) and it was therefore anticipated that backselection would almost certainly occur on clones which have the mutation either in both duplications or in the 5' duplication as well as that shown in figure 4-5. For such clones the opposite applies: nearly all reversion events can be expected to retain the mutation.

Ignoring branch migration and heteroduplex repair-mediated conversion of the 5' homology duplication, the above calculation predicts that 8%, or about 5 clones out of 67, might be predicted to have retained the point mutation in K-*ras* exon 4B, but it

must be borne in mind that many of the G418^r clones in the backselection are likely to be random integrations, which have undergone mutations in the HSV*tk* gene to produce GANC^r clones, and therefore the expected number of targeted revertants is reduced.

4.6.4 Fast-Track Hit and Run

A potential criticism of this protocol is that the background level of mutation of the HSV*tk* gene actually observed would be too high to facilitate the isolation of targeted clones. As may be seen from figure 4-6, at a cell plating density of $1x10^4$ cells/plate $3.5x10^{-2}$ clones/plated cell were obtained. This corresponds to $7x10^6$ GANC^r clones from $2x10^8$ G418^r clones. Therefore at such a sufficiently low plating density that bystander killing by metabolic cooperation may be ignored, a much larger number of GANC^r clones was obtained than originally expected. The results of screening some of the clones obtained by this procedure with vectors pKiTPNT and pK4BΔPNT tend to agree with this criticism, and it may be that the level of competing illegitimate recombination events and mutations of the HSV*tk* gene make the experiment unworkable at the present locus. However, it will be remembered that in chapter 4, evidence was presented for a measurable level of events at the target locus by vector pKiD0PNT, even if the desired hit and run event was not observed. This is consistent with the idea that hit and run reversion events will be favoured over the survival of illegitimate recombination events by mutation but that because the vector fails to target as predicted, the desired event was not detected in this case.

It is likely that the level of background GANC^r clones in this experiment impeding the isolation of targeted clones is an artefact of the negative selectable marker, and that the premise of the experiment is correct. HSV*tk* cannot be positively selected in wild-type cells; in the vectors described in this thesis, *neo* was used as a positive selectable marker. Furthernore, there is also the possibility that mutations in the HSV*tk* gene

during propogation of the construct in *E. coli* will result in a cohort of clones resistant to gancicolvir. After transfection of the vector and before backselection, there is a significant time when HSV*tk* is not under selection-approximately 14 days during the growth of clones under positive selection in G418, followed by pooling (when each clone comprises approximately 1000-5000 individual cells), and then a further 4-5 days of expansion to a total cell number of $2x10^8$ cells prior to plating and backselection the following day. This corresponds to approximately 18 cell generations (our own observations; Hasty *et al.*, 1991c), and constitutes a problem because of the exponential growth of cells: a cell undergoing point mutation of the HSV*tk* gene during the first cell generation and growing at the same rate as other ES cells until backselection produces over 250,000 HSV*tk*$^-$ "daughter" clones upon backselection. Thus reversion events due to intra-chromosomal recombination (selected, predominantly, at the final cell generation when a single HSV*tk*$^-$ cell generates two clones) are swamped by the background of point mutations in the negative selectable marker occurring during positive selection and expansion. The conclusion from this experiment is therefore that HSV*tk*/GANC selection is an inappropriate negative selection method for the fast-track hit and run technique. The *hprt* gene, used as both a positive and negative selectable marker in *hprt*$^-$ ES cells, would represent a significant improvement on the present system, because HAT selection could be maintained during cloning of transfectants and cell expansion, thus eradicating the vast majority of spontaneous mutants. Enrichment for the desired event over the present strategy would therefore be significantly improved, and thus the isolation of the desired mutants is predicted to become a relatively trivial problem. Because there is effectively an additional selection for homologous recombination events at the target locus in addition to that for stable transfection events, the frequency of targeted recombinants should be very high and with the suggested refinement this strategy might have general applicability.

4.7 Conclusion

Two vectors were constructed with the objective of introducing point mutations into the murine K-*ras* gene. One such vector (pKiTPNT) was shown to have a very low targeting frequency so that it was not possible to carry out a backselection on a targeted clone. This experiment would be improved by the use of a vector with increased, isogenic homology to the target locus to improve the targeting efficiency and also by the development of efficient and specific screens for such clones. PCR between the vector and the external genomic sequence may provide a more efficient screen at this locus than Southern analysis, and with the advent of long-distance PCR protocols becomes a much more realistic prospect for the screening of ES cell clones for targeting events, because it is no longer necessary to keep one arm of the homology very short to permit PCR amplification of a diagnostic fragment (Barnes, 1994; Cheng *et al.*, 1994).

The targeting efficiency for vector pK4BΔPNT is unknown because this vector was only used in an experiment where the backselection was carried out on pooled clones. This vector should therefore next be used to electroporate ES cells and screen clones at the $G418^r$ stage to check the initial targeting frequency of this construct. Although such a site is at present unknown, a linearisation site closer to the middle of the region of homology than the *Hin*d III site originally chosen may prove useful in increasing the targeting efficiency (Dickinson *et al.*, 1993). If targeted insertions can be obtained, backselection can be expected to produce a low but observable number of mutation events. A higher efficiency of retention of the mutation is predicted if a clone which has the mutation in the 5' duplication of homology can be isolated. Should this prove not to be the case, the vector homology should be expanded on the 5' arm of the homology containing the mutation. This should improve both the targeting frequency and the frequency of reversions under backselection to the

mutant form. Use of the exon 4B reverse PCR primer described in section 7.4.3 combined with the "Vectorette" (ICI) single direction PCR-based genomic walking system using a proof-reading polymerase would be an appropriate method of deriving the required extra genomic sequences.

RT-PCR was a convenient and precise assay for point mutations introduced into the mouse genome, showing whether or not the experiment had introduced the designed mutation into the target gene, and that the modified gene still produced a functional mRNA molecule.

4.7.1 Mutations Introduced by Hit and Run

Although the conclusion is a somewhat negative one, it now appears likely that the lack of success in deriving insertion and hit and run revertant clones with vector pKiD0PNT described in the previous chapter is not due to the specific nature of the mutation (a large deletion) to be introduced, because similar results were obtained with vectors designed to create either a tiny (1bp) deletion or a missense mutation. It is therefore concluded that the vector architecture and the nature of the target locus are the most important factors in determining the targeting efficiency in an experiment and is furthermore predicted that there is no reason why the mutations described may not be created in the future by further optimisation of the experimental protocol.

CHAPTER 5-

VARIATION OF K-*RAS* EXON 4 SPLICING OPTIONS IN MURINE DEVELOPMENT

Variation of K-*Ras* Exon 4 Splicing Options in Murine Development

5.1 Introduction

As described in the Introduction, section 1.2.2, the K-*ras* proto-oncogene has two alternative C-terminal exons, designated 4A and 4B. Although not associated with a catalytic activity, these exons are regarded as essential to the function of the protein, in that they encode the region of the protein at which the molecule is linked to the inner surface of the plasma membrane. This has been shown to be necessary for normal function of *ras* (Willumsen *et al.*, 1984a, 1984b; Leevers & Marshall, 1992). The exons differ somewhat, in that exon 4A lacks the poly-lysine string which characterises exon 4B. Despite this, most studies, focusing on the behaviour of *ras* as an oncogene or an effector of cell signalling function, have not differentiated between the two forms of K-*ras*, or indeed usually between any members of the *ras* family, and so subdivisions of function between isoforms remain unresolved. For example, the v-*onc* form of K-*ras* is K-ras^{A} (McGrath *et al.*, 1983; Shimizu *et al.*, 1983b), but overexpression of K-ras^{B} has also been observed in transformed cells (George *et al.*, 1985). Relatively little is known of the more subtle, but presumably important, differences in function suggested by the conservation of the *ras* genes. It was therefore decided, in parallel with targeting experiments designed to separate the function of the two K-*ras* isoforms, that the representation of the two different forms of K-*ras* in different situations be investigated. Such a study would hopefully reveal any subdivisions of function.

The objective of the experiment was to observe any differences in the relative preferences of the two K-*ras* splicing variants in the mouse during development, either between stages of development of the embryo or between tissue type, which may indicate subdivision of function of one variant from another. The stratagem

adopted was that of reverse-transcription-polymerase-chain-reaction (RT-PCR), to measure the relative proportions of mRNA of each splicing variant, in a variety of different tissues at different stages of development. RNA was prepared from each tissue sample, and using oligo $dT_{12\text{-}18}$ as a primer cDNA was prepared. The cDNA was then used as template in a K-*ras*-specific PCR and the relative amounts of product of each splicing variant were determined. Determination of the proportions of the mRNA species was achieved by separating the cDNA products by agarose gel electrophoresis, Southern blotting the gel and hybridisation to a K-*ras* cDNA probe. Densitometry of the autoradiographs provided a simple method of determining relative proportions of each cDNA.

5.2 Results

5.2.1 Validation of the Analysis

It is clearly of prime importance that any differences in relative expression observed represent real differences in the tissue under study, and not artefacts of the PCR reaction or procedures used to analyse the products.

5.2.1.1 Primer/Template Sequence Considerations

The mRNAs of each splicing variant, as shown in the Introduction, both contain exons 1 and 4B. There is therefore no difference in binding affinity between the templates and primers for either of the 2 possible cDNA species which might lead to an increase in amplification efficiency for one variant at the expense of the other.

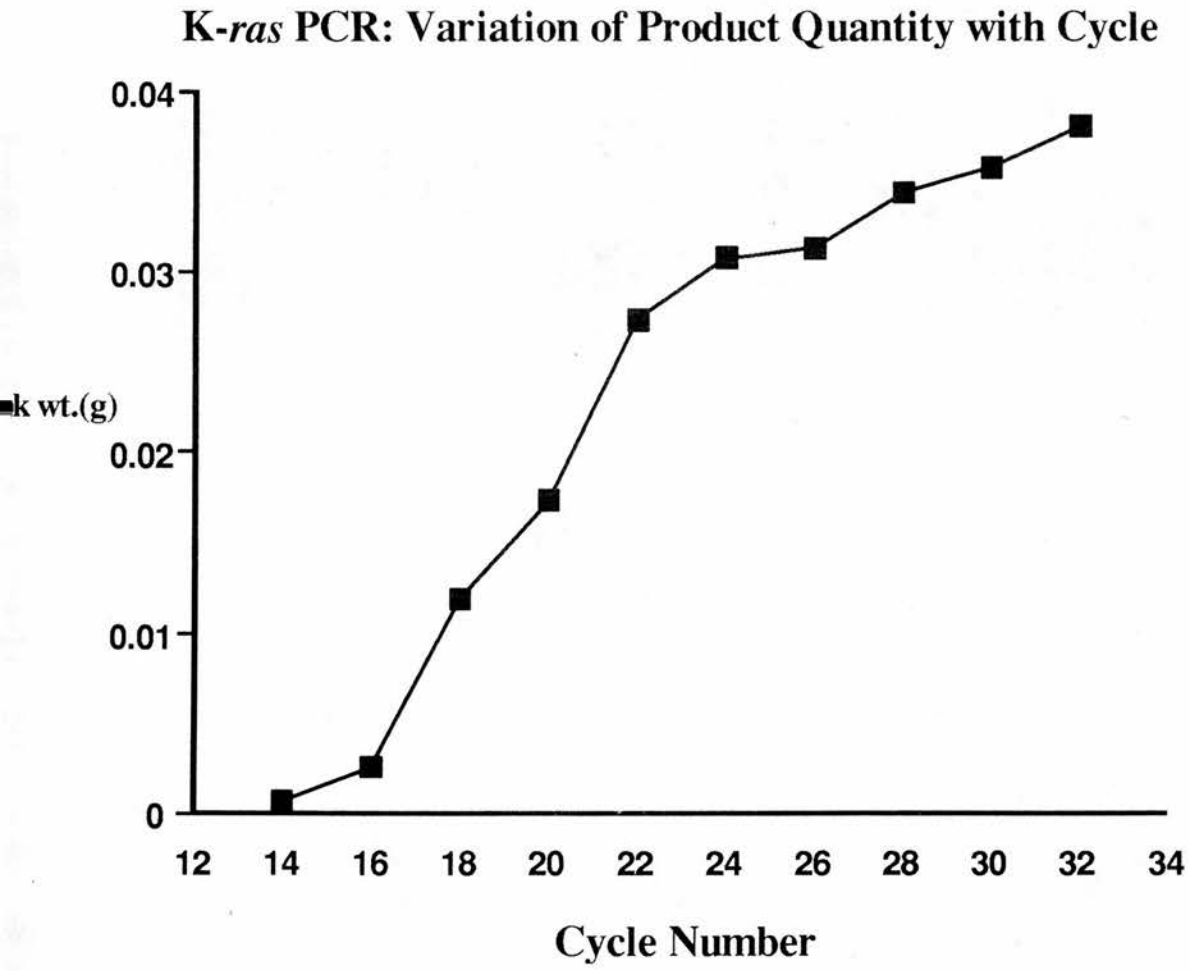
5.2.1.2 Cycle Number

Care was taken to avoid allowing the PCR to progress significantly past the logarithmic stage of reaction, and thus possibly introduce a limiting factor other than the quantity of template. Figure 5-1 shows the amount of product obtained, measured

Figure 5-1: Increase in Quantity of Specific Product With Increasing Cycle Number for PCR of K-*Ras* cDNA

The figure shows the increase in product of the K-*ras* PCR with increasing cycle number. This is measured as an increase in the size of the densitometric peak, representing the product at the specified cycle number, after Southern hybridisation as described in the text. Densitometry of the bands was used to generate a plot of relative peak absorbance against cycle number. After printing, each peak was cut out and the the mass of the peak, proportional to the amount of PCR product, was measured.

Figure 5-1


K-*ras* PCR: Variation of Product Quantity with Cycle
0.04
0.03
k wt.(g)
0.02
0.01
0
12
14
16
18
20
22
24
26
28
30
32
34
Cycle Number

densitometrically, versus cycle number of the PCR. All following PCRs were done at 21 cycles, which was shown empirically to produce a measurable quantity of product in most cases and yet just precedes the plateauing of the PCR.

5.2.1.3 Quantification of Products

In order to confirm that the relative fractions of each PCR product as measured by densitometry, corresponding to relative abundances of mRNA in a tissue, are an accurate reflection of the proportions of each mRNA species the experiment shown in figure 5-2 was performed. A total of 1pg of the two cloned K-*ras* cDNAs was mixed in different molar ratios as template, and then subjected to the PCR. The cDNAs were obtained by RT-PCR of mouse strain 129/Ola liver RNA using the primers described in section 7.4.3 and cloned into pUC19 (Yanisch-Perron *et al.*, 1985). Dideoxy sequencing revealed no differences from the published sequence (George *et al.*, 1985). The RT-PCR products were separated by agarose gel electrophoresis and Southern blotted. The PCR products were then hybridised to the K-*ras* cDNA containing exons 1, 2, 3, 4A and 4B, and the filter was autoradiographed. Autoradiographs were taken at several exposures, to avoid the introduction of errors as a consequence of saturation of the film.

Autoradiographs were analysed by scanning laser densitometry, and the relative band densities were compared with the relative proportions of each template initially used in each reaction.

As is shown in figure 5-2, there was very good agreement between the relative proportions of each template in the reaction and the relative intensities of each band, as measured on the autoradiograph by laser densitometry. A correlation coefficient between the proportion of each cDNA present as template and the density of the band representing it on the autoradiograph of greater than 0.999 was recorded.

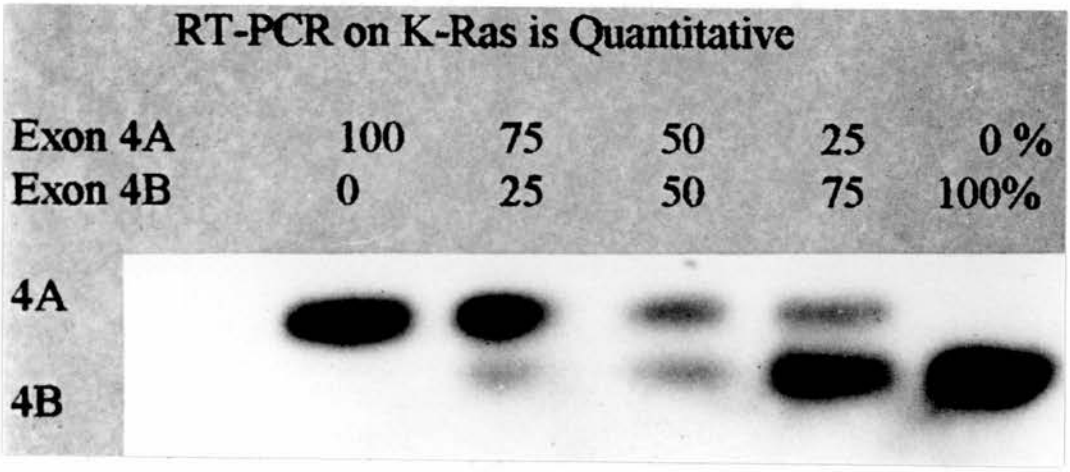
Figure 5-2: Quantity-Specific Nature of the PCR of K-*Ras* cDNAs

a) Southern analysis of K-*ras* PCR carried out on 1pg of control murine K-*ras* cDNAs, mixed in the molar quantities shown above the bands.

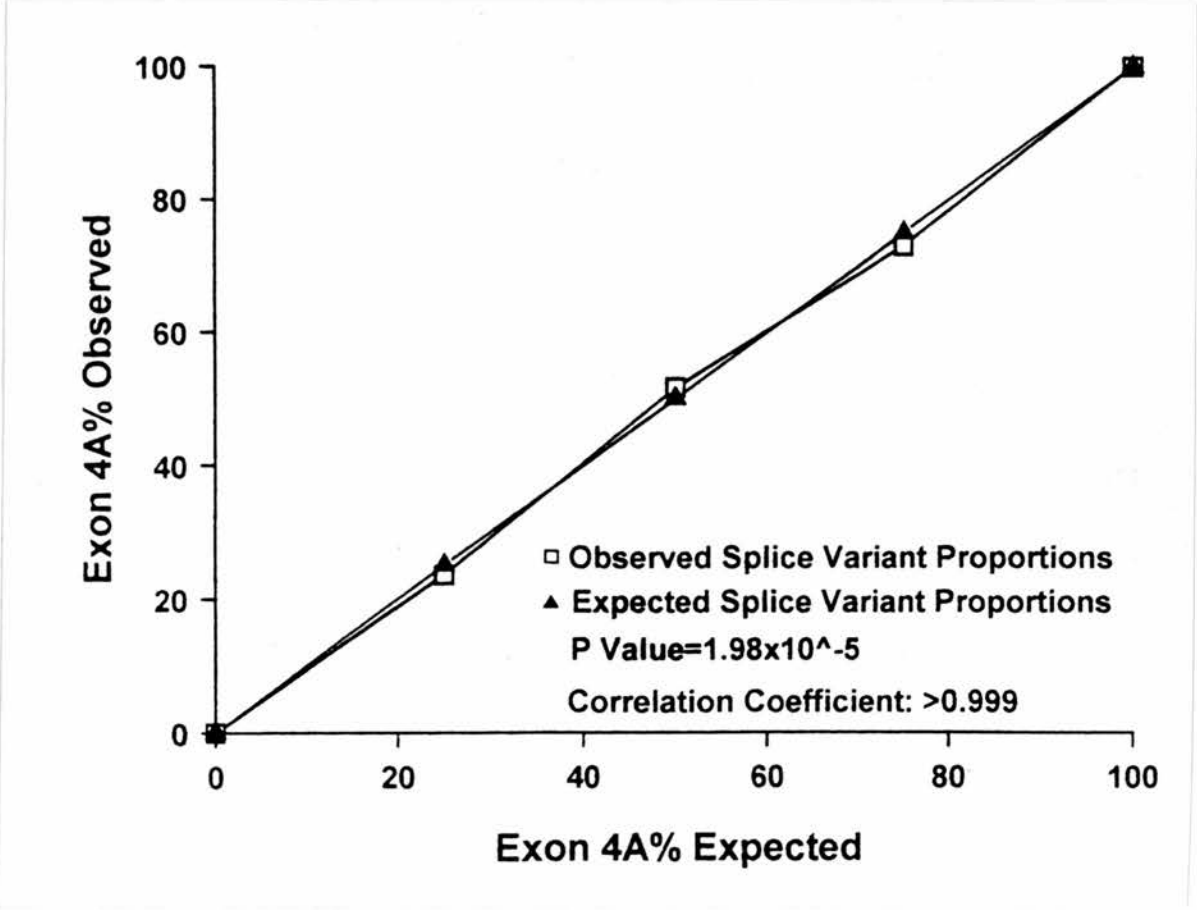
b) Graph of the observed splice variant proportion as a percentage of the total K-*ras* PCR product against the proportion expected based upon the actual ratios of template used in the PCR. Symbols: black triangle, expected values; white squares, observed values. The P value gives the significance of the correlation coefficient.

Figure 5-2

(a)


RT-PCR on K-Ras is Quantitative
Exon 4A 100 75 50 25 0 %
Exon 4B 0 25 50 75 100%
4A
4B

(b)


Exon 4A% Observed
Exon 4A% Expected
0
20
40
60
80
100
Observed Splice Variant Proportions
Expected Splice Variant Proportions
P Value=1.98x10^-5
Correlation Coefficient: >0.999

5.2.2 K-*Ras* Exon 4 Splicing Options in Murine Development

Balb/c mice were taken every three days of pregnancy from day 9 onwards to birth, and sacrificed. Embryos were dissected as far as technically possible at each stage of development. Embryos were not dissected earlier than day 9 post coitum because of the technical difficulties of dissection and the problem of separating maternal from embryonic tissue. Three to five samples of each tissue from different embryos were analysed separately. In addition, ES and EC cells were analysed as an approximation of early stage (day 3.5) development, although it should of course be borne in mind that these are actually cell lines adapted to *in vitro* conditions. The ES and EC cells were also subjected to induction of differentiation *in vitro* as described in chapter 7 and the differentiated populations studied.

Sample data, obtained by RT-PCR and Southern hybridisation to the K-*ras* cDNA, of K-*ras* transcripts from ES and EC cells and isolated mouse tissues are shown in figure 5-3.

Table 5-1 shows the results obtained for the exon 4 splice options analysis carried out on the developing mouse embryo. These data are also presented in a graphical format in figure 5-4, where any changes with time are perhaps more easily seen.

5.3 Discussion

5.3.1 PCR of Murine K-*ras* cDNA is Quantity-Specific

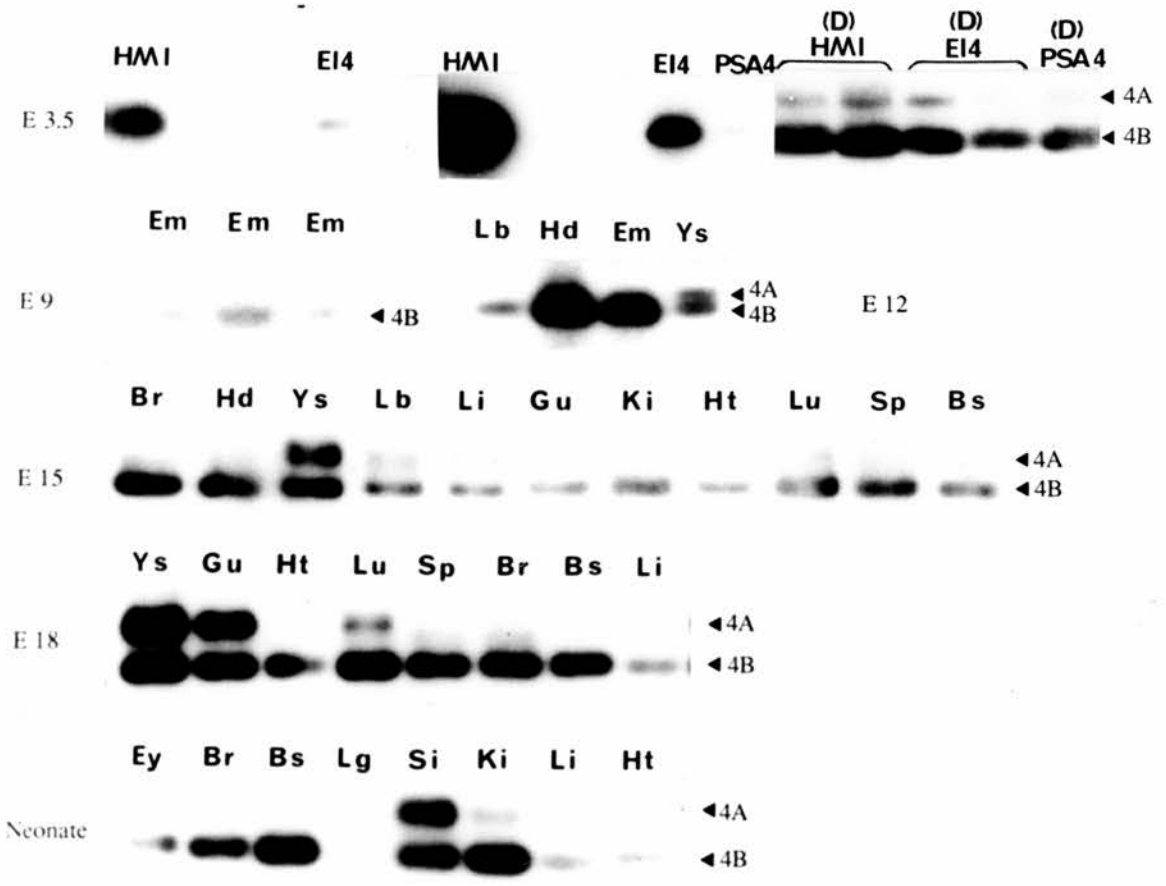
The major criticism of this type of analysis is that the data obtained, (i.e. measurements of the relative proportions of each K-*ras* splicing variant for a given tissue at a given stage in the development of the mouse embryo), may not accurately reflect exon 4 expression patterns in the mouse. Discrepancies may arise because of the introduction of artefacts or distortions at one or more of several stages: reverse

Figure 5-3: Tissue Specific K-*Ras* Splicing Variant Preferences in the Developing Mouse

The figure shows the results of Southern hybridisation of RT-PCR product of murine K-*ras* transcripts to the K-*ras* cDNA probe containing exons 1, 2, 3, 4A and 4B. Samples from various tissues at different stages of development are shown. The examples of undifferentiated ES and EC cells at the top row ("E 3.5") are duplicated at two different exposures of the autoradiograph to show that only the K-*ras*^{B} transcript was observed for undifferentiated E14, HM1 and PSA4 cells.

Figure 5-3

Tissue-Specific K-*ras* Exon 4B Splicing Variant Occurrence in the Developing Mouse Embryo



Key

Ex	Days of Embryoid Development
E14	Embryonic Stem Cell Line E14
HM1	Embryonic Stem Cell Line HM1
PSA4	Embryonal Carcinoma Cell Line PSA4
(D)	Differentiated as cell aggregates for 4-5 weeks

Br	Brain	Bs	Brain Stem	Em	Embryo
Ey	Eye	Gu	Gut	Hd	Head
Ht	Heart	Ki	Kidney	Lb	Limbs
Li	Liver	Lg	Large Intestine	Lu	Lungs
Sp	Spine	Si	Small Intestine	Ys	Yolk Sac

Table 5-1: Tissue Specific K-*ras* Splicing Variant Preferences in the Developing Mouse

The table shows the relative proportions of K-*ras* transcripts K-*ras*^{A} and K-*ras*^{B} expressed in each tissue type, as a percentage of the entire K-*ras* mRNA measured. Each value is given as a mean of up to 5 samples.

Notes:

1) Stage of development is given in days; "N" denotes a neonatal sample.

2) The number of samples is shown in column **n.**

3) "Diff. 1" indicates differentiation stage 1; see chapter 7.

4) "Diff. 2" indicates differentiation stage 2; see chapter 7.

Tissue	Day	n	4B Mean (%)	4A Mean (%)
E14	3.5	1	100	0
E14 (Diff 2)	3.5	2	90	10
HM1	3.5	1	100	0
HM1 (Diff 1)	3.5	1	53	47
HM1 (Diff 2)	3.5	4	85	15
PSA4 [EC]	-	1	100	0
PSA4 [EC] (Diff 2)	-	2	79	21
Embryo	9	5	100	0
Embryo+Yolk Sac	9	2	100	0
Embryo	12	5	99	1
Gut	15	4	77	23
Gut	18	4	60	40
Small Intestine	N	4	48	52
Large Intestine	N	3	50	50
Stomach	N	3	70	30
Liver	12	4	89	11
Liver	15	4	60	40
Liver	18	4	77	23
Liver	N	3	70	30
Kidney	15	4	86	14
Kidney	N	4	85	15
Heart	15	4	97	3
Heart	18	4	99	1
Heart	N	4	85	15
Lungs	15	4	72	29
Lungs	18	4	75	25
Lungs	N	4	66	34
Spine	15	5	87	13
Spine	18	3	97	3
Spine	N	4	90	10
Brain	15	5	99	1
Brain	18	4	97	3
Brain	N	4	76	24
Brain Stem	15	3	91	9
Brain Stem	18	4	98	2
Brain Stem	N	4	98	2
Eye	15	4	95	5
Eye	N	4	94	5
Head	12	5	96	4
Head	15	5	99	1
Head	N	1	100	0
Limbs	12	2	100	0
Limbs	15	5	71	29
Blood	15	1	92	8
Yolk Sac	12	5	56	44
Yolk Sac	15	5	54	46
Yolk Sac	18	4	44	56

Figure 5-4: Tissue Specific K-*ras* Splicing Variant Preferences in the Developing Mouse

The Figure shows the relative proportions of K-*ras* transcripts K-*ras*^{A} (red) and K-*ras*^{B} (blue) expressed in each tissue type, as a percentage of the entire K-*ras* mRNA measured. Each value is given as a mean of up to 5 samples.

Notes:

1) Stage of development is given in days; "N" denotes a neonatal sample.

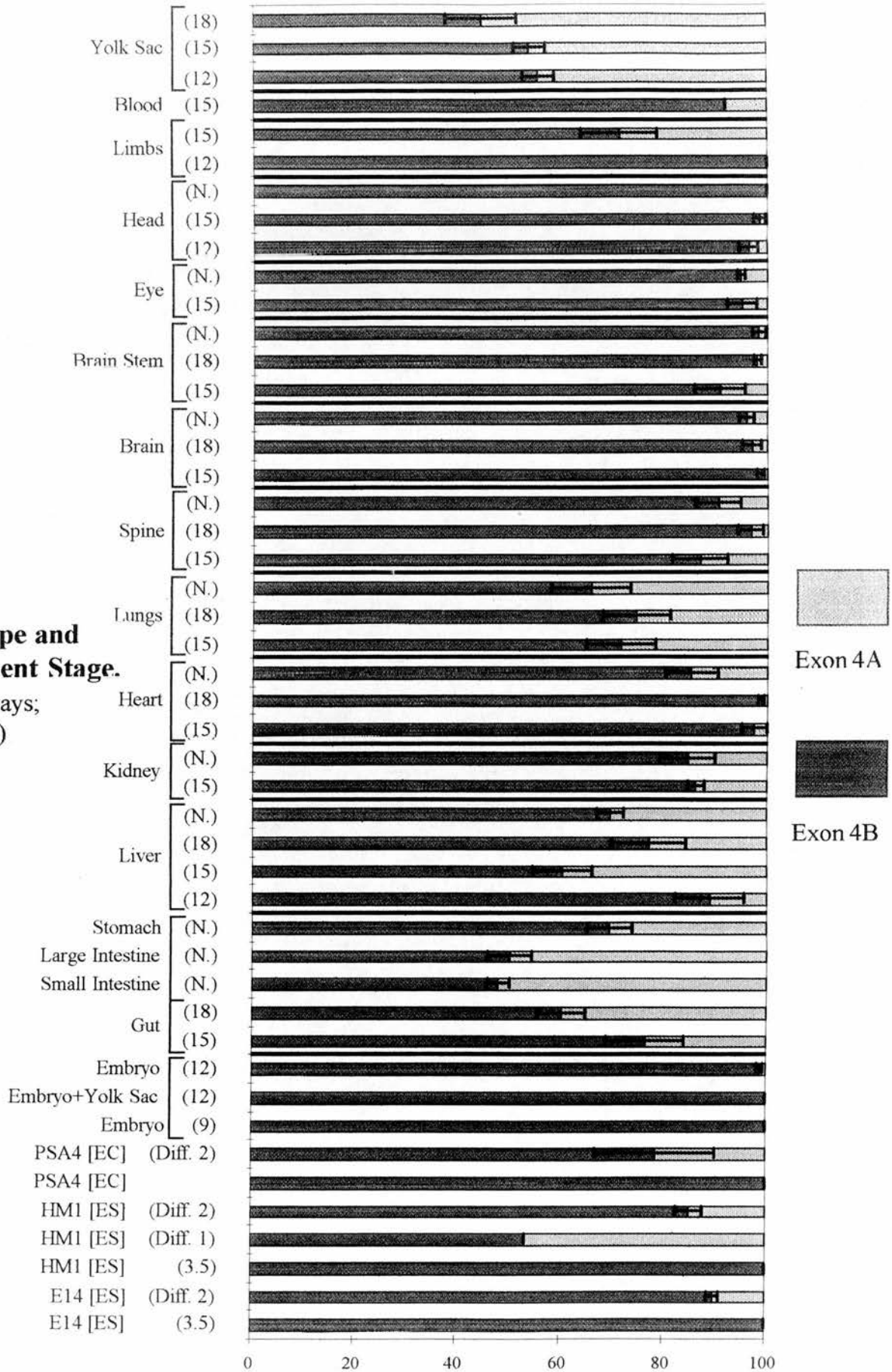
2) Error bars represent standard error of the mean for the tissue type and development stage.

3) "Diff. 1" indicates differentiation stage 1; see chapter 7.

4) "Diff. 2" indicates differentiation stage 2; see chapter 7.

ure 5-4

K-*ras* Exon 4 Splicing Variant Expression in the Mouse During Development



Type and
pment Stage.
r=Days;
-nate)

transcription, PCR, Southern analysis, autoradiography. It was therefore essential to optimise the analysis before proceeding to the mouse, and this is the purpose of the experiments described in section 5.2.1.

The control experiments described in sections 5.2.1.2 and 5.2.1.3 were carried out using 0.1-1pg of linear K-*ras* cDNAs mixed either in equimolar ratios (5.2.1.2) or in varying molar ratios (5.2.1.3). This amount was chosen for the optimisation experiments because it was likely to be representative of the quantities of K-*ras* mRNA/cDNA recovered from tissue samples.

Experiment 5.2.1.2 showed that the amount of product continues to increase during the PCR up to approximately 22-23 cycles, and then the plateau effect begins to be seen as the amount of new product synthesised begins to level off. 21 cycles was therefore selected as the optimum number for the analysis because this is high enough to give an easily detectable quantity of product and yet remain within the exponential region of the PCR, and thus avoid any potential problems with preferential amplification which may occur as the reaction begins to saturate. The cDNA is single stranded and so only one PCR primer can bind the template on the first round of the PCR. Consequently, at the start of the analysis there is one round of arithmetic rather than geometrical amplification of the target sequence. Therefore 21 cycles of PCR on the tissue-derived cDNA more properly equates to 20 cycles of PCR in the control experiment, thus placing the system effectively in the middle of the exponential range of the PCR.

Experiment 5.2.1.3 showed that there was good agreement between the starting ratios of the two variants at the beginning of the PCR and the ratios measured after PCR, Southern blotting and autoradiography. Results gained from murine tissues should therefore accurately reflect the relative amounts of each cDNA obtained after reverse transcription.

It should be borne in mind that this experiment is not a truly quantitative PCR in the classical sense, because the objective is not to compare amounts of product between samples, but rather to quantify the ratio of the two variants both present in the same sample. For this reason it was unnecessary to include an internal standard in the PCR reaction tube, but simply to show that measured ratios correspond well with the ratios of starting template.

The only source of error unaccounted for in the above control experiments is the starting material itself, i.e. RNA from tissue samples. Degradation of RNA by poor handling can be expected to reduce yield, although the use of RNAse-free materials and the use of liquid nitrogen to store RNA samples ameliorates this. However, would one mRNA species be significantly more stable than the other? Such a situation would result in poor estimates of relative expression by overestimating one splicing variant at the expense of the other. This was thought unlikely here however because of the relatively small size difference in the mRNA species (under 5%, George *et al.*, 1985) and the fact that the exon producing the size difference is internal to the primers. Thus exonuclease-mediated degradation can be expected to affect the quantity each variant obtained equally. Because both the mRNA species have the same untranslated sequences and encode products of almost identical size (189 amino acids for K-*ras*^{A} and 188 amino acids for K-*ras*^{B}) the "physiological" stability of each mRNA species in the cell is almost certainly identical (Sachs, 1993).

5.3.2 K-*Ras* is Represented in All Tissues Studied

In agreement with previous work (Muller *et al.*, 1983; Leon *et al.*, 1987), K-*ras* expression was observed in all embryonic tissues studied. The only exception to this is blood, where it was rarely observed, but this was due to the difficulty of obtaining adequate quantities of embryonic material.

This analysis does not formally show that K-*ras* is expressed in every cell type, because not every tissue type was examined, and most tissues comprise more than one cell type, but it appears likely from these observations that the K-*ras* gene is ubiquitously expressed at the tissue level in the mouse.

5.3.3 K-*Ras* Exon 4B is Always Expressed in Cells, Whereas Exon 4A is Not

As can be seen from the results in table 5-1 and figures 5-3 and 5-4, the expression of K-*ras* mRNA isoforms varies between tissue types. The mRNA of K-*ras*^{B} is present in all tissues studied. Furthermore, this form is almost always the most abundant, whatever the tissue. The only exceptions to this rule, neonatal small intestine and day 18 yolk sac, are where the values for exon 4A and exon 4B are sufficiently close that they are within the errors of the experiment and therefore no significance can be concluded from this.

K-*ras*^{A} is usually expressed in lower amounts, and in some cases such as neural tissue and muscle tissue is observed in very low quantities indeed (<10%). Further analysis is required to show whether such cell types in fact produce only the exon 4B transcript and the exon 4A transcript observed merely represents contamination from other cell types present in the sample.

5.3.4 Within a Tissue, K-*Ras* Splicing Variant Expression Does Not Usually Vary Significantly During Development

As is seen from the graphical presentation of the data, within any tissue type there is usually no obvious trend with respect to the relative proportions of each K-*ras* mRNA species over time. The Mann-Whitney non-parametric test was used to test the significance or otherwise of differences between proportions of K-*ras*^{B} in similar tissue samples at different stages of development.

There are few exceptions to the finding that there are no significant differences in K-*ras* exon 4 splicing variant expression over time within the same tissue type. In the heart, a slight elevation in K-*ras*^{A} (from extremely low levels to 15% of the total) was observed between day 18 embryos and newborn animals. In the liver between days 12 and 15 of development, an elevation in K-*ras*^{A} (from approximately 10% to 40% of the total) was observed. However, in both these cases the importance of these findings is cast into doubt by the lack of a significant difference between either of these stages of development and any of the others measured for the same tissue. They are therefore likely to be stochastic artefacts of the analysis.

In the case of the gut, there may be a slight increase in K-*ras*^{A} during development. There was no significant difference between large and small intestine of the neonate, and so these values were pooled to form a group of 7 observations designated "neonatal gut". Significant increases in K-*ras*^{A} compared to K-*ras*^{B} were observed between day 15 gut and neonatal gut and between day 18 gut and neonatal gut. Although it is obviously important to avoid over-interpretation of the data, this observation may reflect a small but significant change in K-*ras* expression patterns in the gut during the later stage of development, when the gut is undergoing active differentiation processes (see next section; Kaufman, 1992).

At day 9, no K-*ras*^{A} expression was observed in whole embryos, but within certain individual tissues (e.g. yolk sac and liver) at day 12, expression of this variant was observed. Thus it may be that for approximately the first half of development, K-*ras*^{B} is the only significant K-*ras* isoform expressed. However, a finer-level analysis of early-stage embryos is required to show this, as it may be that small amounts of K-*ras*^{A} present in particular cell types at this stage of development were not detectable in the analysis due to an excess of competing K-*ras*^{B}, present in all tissues.

In the majority of tissues studied significant modulation of K-*ras* expression patterns during development was not observed. These tissues included the embryonic brain, brain stem, spine, eye, heart, lungs, kidney, liver and yolk sac. As already seen however, there are frequently significant differences between tissue types. This result is consistent with the hypothesis that a cell type's pattern of K-*ras* expression depends on the cell type and that proliferation of the tissue during growth does not usually alter this; rather it is more likely that expression of K-*ras* is an early event in development. The fact that K-*ras* is expressed, albeit as K-*ras*^{B} only, in ES cells, (see also Pal *et al.*, 1993) is consistent with this hypothesis. It is therefore suggested that the K-*ras* splicing pattern for a particular cell type is probably determined at or before the time of differentiation into that cell type.

5.3.5 K-*ras* Splicing Variant Exon 4A is Associated with Differentiation

The most significant findings from this work are those pertaining to the study of ES cells and their differentiation. Three cell lines were studied-the ES cell lines E14 and HM1, and the EC cell line PSA4. PSA4 is an EC cell line which shows a wide spectrum of cell types upon *in vitro* differentiation (Table 1, and references therein, of the Appendix, Silver *et al.*, 1983). In addition to analysis of the parent cells, the cells were subjected to the induction of differentiation *in vitro* as described in chapter 7 and the differentiated derivatives were also studied.

The major difference between the use of cell lines and tissues derived from a mouse is that prior to any *in vitro* differentiation treatment the cells are a pure population, with little or no contamination of the population by other minority cell types as is present in tissues derived from the mouse. Therefore it was concluded that in ES and EC cells, the exon 4B splicing variant is the only K-*ras* transcript expressed. However, upon differentiation *in vitro*, K-*ras*^{A} expression rapidly (within days of the removal of

LIF) becomes upregulated to comprise an appreciable proportion of the total K-*ras* mRNA as the cells lose their pluripotent phenotype and become committed to various pathways of differentiation. Although not formally shown, it is likely that the relatively wide variation in 4A:4B ratios recorded for individual populations of differentiated ES and EC cells is a reflection of the extent of differentiation and the proportions of particular lineages which developed from the pluripotent cells in the different experiments. It is interesting to note that no evidence for the presence of the K-*ras*^{A} transcript was observed at day 9 of embryogenesis. Although this remains to be confirmed by a finer-level analysis, it may suggest, in combination with the results obtained from the *in vitro* differentation of pluripotent cells, that induction of K-*ras*^{A} transcription is a relatively late event in embryogenesis (between days 9 and 12).

As described above, some evidence for a role of K-*ras*^{A} in gut development and differentiation was also observed in this study.

Is the upregulation of K-*ras*^{A} expression an initiating event in the formation of different tissues from undifferentiated precursors, or merely an associated observation? The fact that studies of *ras* on cell lines (Hagag *et al.*, 1986; Szerberenyi *et al.*, 1990) suggest an important role for *ras* in differentiation may argue for the former proposition: the presence of multiple splicing variants with subtly different functions provides another level of control over *ras*-mediated activation/differentiation of the cell. The different splicing variants may interact with different upstream and downstream molecules to activate similar, but not necessarily the same, genes in response to *ras* activation. This is also consistent with the other major observation of this work, that splicing variant usage does not usually change over time within a tissue type. It is suggested that once a cell is committed to a certain lineage, K-*ras* splicing patterns tend to remain constant as one of the functions of the molecule is to maintain the appropriate differentiated phenotype of

the cell by activating the right genes in the right proportions. This may be upset in neoplasms by mutation of *ras*, and indeed may be affected by the type of neoplasm (tissue type) in question.

Further analysis of individual cell types derived from differentiated ES and EC cell precursors, together with a finer-level analysis of the early-stage embryo, will be important in verifying the association of K-*ras* expression patterns with pathways of differentiation. Comparison of the ability of ES cells carrying homozygous mutations forcing the exclusive expression of either K-ras^B or K-ras^A to differentiate along particular lineages as compared to their wild-type controls, and an analysis of the same mutations in the context of the mouse embryo will formally show the significance of varying K-*ras* expression patterns in particular developmental processes. Since it is already apparent that quantitative differences in K-*ras* exon 4 variant expression are seen between different tissues, it is suggested that there may also be phenotypic differences observed in mice heterozygous for mutations deleting either K-ras^B or K-ras^A as the appropriate quantitative balance between the two isoforms of K-*ras* is perturbed.

Should mice carrying mutations deleting either K-ras^B or K-ras^A prove to be viable, an analysis of the profile of K-*ras* mutations in tumours in these animals and of the spectrum of tumours which may be induced may also prove informative in the study of the involvement of K-*ras* in malignancy.

CHAPTER 6-
CONCLUDING DISCUSSION

Concluding Discussion

This study has examined some aspects of the hit and run gene targeting technique as a method for the introduction of small mutations into the mammalian genome, and some aspects of the function and expression of the K-*ras* proto-oncogene *in vivo*. An important issue when attempting a hit and run experiment is the accuracy with which the homologous recombination event between a chromosome and the introduced plasmid is accomplished. A low level of fidelity in this process might engender the introduction of undesired mutations together with the intended modification. This question was investigated in the context of the *hprt* gene (chapter 2). It was hypothesised that an error introduced into the target locus during a hit and run experiment would most likely be introduced during the initial insertion ("hit") step of the procedure, since unexpected mutations have previously been observed in gene targeting experiments (Thomas & Capecchi, 1986; Doetschman *et al.*, 1988; Thompson *et al.*, 1989; Brinster *et al.*, 1989), but a high level of accuracy has been observed in intra-chromosomal recombination events upon which the second, "run", step of the procedure depends (Stachelek & Liskay, 1988). Therefore a strategy based upon the correction of the $hprt^{b\text{-}m3}$ allele by an insertion vector was adopted to study this problem. In addition to the required correcting sequences, a *neo* gene was included in the targeting construct to permit the recovery of all "targeted" clones, whether they were faithful or unfaithful recombinants. A comparison of Southern analysis and phenotypic analysis (assessed by the ability of recombinant clones to survive in HAT medium) was used to study the clones for evidence of errors occurring during HR and resulting in a HAT^s phenotype. The *hprt* enzyme activity of targeted clones was also analysed. A wider standard deviation of *hprt* enzyme activities for targeted clones as compared with wild-type controls was expected if small errors in recombination resulted in clones with an attenuated level of *hprt* activity which were still HAT^r. Although the results were somewhat clouded by a reduced level of *hprt* enzyme activity in the targeted clones, probably due to the presence of interfering bacterial

plasmid sequences in vector pDMW101neo (Melton, 1990), it has been possible to make certain conclusions regarding the accuracy of HR events. No evidence was obtained for the presence of deletions at the site of integration, as has been previously reported (Doetschman *et al.*, 1988; Thompson *et al.*, 1989). In the present study, there was no reason why clones would not be identified in which deletions had occurred at the 3' end of the integration site (as well as the 5' end as previously reported), resulting in ablation of *hprt* function, because clone isolation was independent of gene function. None of the clones shown to be targeted by Southern analysis were found to be HAT^{s}. It was also observed that the mean level of *hprt* enzyme activity of a targeted clone was independent of the method of selection of the clone, which is inconsistent with a high rate of error in the HR process. It was therefore possible to set a upper limit on the inaccuracy of gene targeting, despite the relatively small sample size, which is compatible with the successful execution of hit and run gene targeting experiments.

Experiments were also carried out which had the objective of introducing small, defined mutations into the murine K-*ras* gene, by hit and run (chapters 3 and 4). These mutations were a large deletion covering the promoter/exon 0 region, a missense mutation creating an activating mutation, and a deletion of a single base-pair leading to a frameshift and premature termination of translation of (specifically) K-ras^{B}. A low targeting efficiency was observed. In the case of the vector designed to create a deletion, only one clone which had undergone a recombination event at the target locus was obtained, and this was subsequently shown to have been generated by an unpredicted event. The event leading to the structure suggested is thought to have involved the combined concatenation of two vector molecules and an intra-vector recombination event to produce a more complicated targeting plasmid structure prior to insertion. Upon integration of the rearranged construct into the target locus by homologous recombination, a complex insertion structure resulted, containing a triplication of homology and two copies of the positive selectable marker, but not the negative selectable marker. This created multiple possibilities for intra-chromosomal

recombination events upon backselection, some of which were observed, and some of which (including, unfortunately, the desired deletion) were not.

Similarly, no targeted clones, suggesting a very low targeting frequency were observed for the construct designed to introduce an activating mutation at codon 12 of the K-*ras* gene. Further problems were experienced with this experiment, in that it proved difficult to screen the clones obtained by Southern blot. Therefore, a "fast-track" protocol was designed both to accelerate the production of and increase the probability of obtaining the desired revertant mutation in a hit and run experiment. This protocol was adopted with this vector and the other vectors designed to introduce mutations in the K-*ras* gene. The fast-track protocol showed very little improvement over the conventional approach with the selectable markers in use (*neo* and HSV*tk*) due to an unacceptably high level of background mutation at the HSV*tk* gene. It is suggested however, that this may be avoided by the use of an *hprt* minigene as both a positive and negative selectable marker, which would eliminate the majority of background clones obtained in this type of experiment. The problems experienced in the screening of clones obtained with vector pKiTPNT were solved by the use of RT-PCR to generate a substrate which could be subjected to diagnostic restriction enzyme digestion, to directly analyse the mRNA of the target gene. This strategy was also used to provide a rapid screen for clones obtained with the vector designed to introduce a premature termination codon into the K-*ras* gene. No such clones were obtained, probably because of the problem of HSV*tk* mutation described above, although refinements to this experiment have also been suggested. The screen at the mRNA level based upon RT-PCR represents a generalised approach to the screening of targeted mutations which is unaffected by limitations of Southern analysis, such as sensitivity and the size of changes introduced into the target locus which may be studied.

The expression of the two alternative K-*ras* exon 4 splicing variants in the developing mouse has also been studied (chapter 5), by means of an RT-PCR strategy which

amplifies both of the K-*ras* cDNA molecules in proportion to their original relative amounts. This allowed a direct comparison of the ratios of the two mRNA species from different tissues at different stages of development of the mouse to be made. From the results obtained in this study, it is apparent that there is a significant, defined subdivision of function of the two splicing variants of K-*ras in vivo*. In agreement with previous work (Muller *et al.*, 1983; Leon *et al.*, 1987; Pal *et al.*, 1993) K-*ras* expression was observed at all stages of development and in all tissues studied, but these studies have not sought to differentiate between K-*ras*^{A} and K-*ras*^{B}. Here, however, it was shown that whilst K-*ras*^{B} is ubiquitously expressed, K-*ras*^{A} has a more limited pattern of expression, and where present is usually the minor component of the K-*ras* mRNA in a tissue. There was little evidence for changes in K-*ras*^{A} and K-*ras*^{B} during the growth of differentated organs, suggesting that the K-*ras* expression pattern for a particular cell- or tissue-type is an event determined early in development. An association of K-*ras* splicing variant expression patterns with the differentiation of cell types was demonstrated, based upon the observation that ES or EC cells, of different lines, normally expressed exclusively K-*ras*^{B} but upon *in vitro* differentiation a significant and rapid upregulation of K-*ras*^{A} expression was observed.

CHAPTER 7-
MATERIALS AND METHODS

Materials and Methods

7.1 Manipulation of DNA

7.1.1 Large-Scale Preparation of Plasmid DNA

Two 20µl aliquots of competent *Escherichia coli* cells (>2x10^7 CFU/µg pUC19), strain DH5α (Life technologies) were removed from -70°C storage and thawed on ice. 1µl (about 1µg) of DNA solution was added to one of the tubes of *E. coli* cells and the tubes incubated on ice for 30 minutes. The cells were heat-shocked at 42°C for 40 seconds and placed on ice. 80µl of S.O.C. medium was added to the tube and the cells were incubated at 37°C for one hour in a shaking incubator (225 rpm). The transformation was plated out onto an L-amp plate (disposable 100mm bacterial culture dish containing L-broth supplemented with 50µg/ml ampicillin and 1.2% bactoagar) and incubated overnight at 37°C.

S.O.C. Medium		**L-Broth**	
Bactotryptone	20g/l	Bactotryptone	10g/l
Yeast Extract	5g/l	Yeast Extract	5g/l
NaCl, $MgCl_2$, $MgSO_4$	10mM each	NaCl	0.17mM
KCl	2.5mM	in ddH_2O, autoclaved	
Glucose	20mM	Supplemented with 50µg/ml ampicillin where stated	
in ddH_2O, filter sterilised		Supplemented with1.2% w/v Bactoagar for plates	

The next day a colony was picked from the plate transformed with plasmid DNA (the other plate should give no colonies) and was placed in 10ml L-broth supplemented with ampicillin and the cells grown for about 8 hours in a shaking incubator at 37°C. This culture was added to a 500ml flask of L-broth with ampicillin and incubated overnight in a shaking incubator at 37°C, 225 rpm.

The following morning the cells were harvested by centrifugation at 4000xg for 10 minutes at 4°C. The supernatant was discarded and the cells lysed by resuspending in 10ml Solution 1 with 5mg/ml lysozyme and incubating at room temperature for 5 minutes.

Solution 1:	Glucose	50mM
	Tris.HCl pH 8.0	25mM
	EDTA	10mM

20ml of fresh Solution 2 was then added, mixed by inversion and the lysate was incubated on ice for 10 minutes.

Solution 2:	NaOH	0.2M
	SDS	1% w/v

15ml Solution 3 was added, the lysate mixed by inversion and incubated on ice for 10 minutes.

Solution 3:	Potassium Acetate	3M
	Glacial Acetic Acid	11.5% v/v

The precipitated proteins were removed by centrifugation at 48000xg for 20 minutes at 4°C and the DNA was precipitated from the supernatant by addition of 0.6 volumes of propan-2-ol, mixing by inversion and incubating at room temperature for 15 minutes. The DNA precipitate was recovered by centrifugation at 17000xg for 30 minutes at room temperature. The supernatant was discarded, and the pellet washed with 70% ethanol and then dried under vacuum. The DNA was resuspended in 4ml TE buffer.

TE Buffer:	Tris.HCl pH 8.0	10mM
	EDTA	1mM

4g of caesium chloride and 0.4ml 10mg/ml ethidium bromide were then added and the DNA solution placed in a polythene tube. The tube was centrifuged in an evacuated ultracentrifuge overnight (Beckman, rotor V65ti) at $2x10^5$xg, room

temperature. The DNA band was visualised with long-wave U.V. radiation, and recovered with a syringe and wide-bore needle. 6-10 extractions (two more extractions than was sufficient to completely remove the red colour from the solution) with CsCl-saturated propan-2-ol were then carried out to remove the ethidium bromide from the plasmid. The plasmid DNA was precipitated by combining 0.45ml DNA solution with 0.5ml H_2O and 0.54ml propan-2-ol and centrifugation at 13000rpm for 2 minutes in a microfuge. The supernatant was discarded and the pellet air-dried and finally resuspended in 0.5ml TE.

7.1.2 Small Scale Preparation of Plasmid DNA

This method is a modification of the procedures described by Birnboim & Doly (1979) and Ish-Horowicz & Burke (1981), described in Sambrook *et al.* (1989). Solutions1, 2 and 3 were as described in section 7.1.1. A single bacterial colony from a 100µl transformation was transferred to 1ml L-broth supplemented with 50µg/ml ampicillin and incubated with shaking overnight at 37°C. The cells were harvested by centrifugation in a microfuge for 30 seconds at 4°C and the supernatant discarded. The pellet was then resuspended in 100µl cold solution 1 (omitting lysozyme), mixing by vortexing. 200µl fresh solution 2 was added and mixed by inversion. 150µl solution 3 was then added, the lysate again vortexed and incubated on ice for 5 minutes. The tube was centrifuged for 5 minutes in a microfuge, and the supernatant transferred to a fresh tube. An equal volume of 50% phenol: 50% chloroform was added, mixing by vortexing, and then microfuging for a further 2 minutes at 4°C. The aqueous layer was removed and the DNA precipitated by addition of 2 volumes of ethanol. After 2 minutes, the tube was microfuged for 5 minutes at 4°C and the supernatant discarded. The pellet was air-dried and redissolved in 50µl TE pH8.0 containing 20µg/ml DNAse-free pancreatic RNAse. Inserts of clones were then

checked by restriction digestion and agarose gel electrophoresis, and dideoxy sequencing, prior to transformation and large-scale preparation.

7.1.3 Preparation of DNA Fragments using Powdered Glass

The Geneclean Kit from Bio 101, La Jolla, USA was used for the preparation of DNA fragments for ligations, probes, etc. Briefly, The desired amount of DNA was digested with the appropriate restriction endonuclease, and separated on a 0.8% agarose-TBE gel. The band was visualised quickly with U.V. light and excised from the gel with a scalpel. 0.5 volumes of Bio 101's modifier solution for TBE gels and 4.5 volumes of 6M NaI solution were added and the gel dissolved at 55°C. 10µl, or a larger amount if appropriate, of Glassmilk^{TM} were added and the sample incubated on ice for 5-15 minutes. The glass was then recovered by centrifugation for 10 seconds in a microfuge and washed 3 times with Bio 101's NaCl/Ethanol/H_2O wash. The DNA fragment was then eluted from the glass into TE at 55°C for 5 minutes.

7.1.4 DNA Ligation

Following fragment isolation, appropriate amounts of fragments calculated according to Sambrook *et al.*, 1989, from Dugaiczyk *et al.* 1975 (typically a total of about 150ng DNA) were mixed with 0.4 Weiss units T4 DNA ligase and ligation buffer and incubated at room temperature for 4 hours.

Ligation Buffer:	Tris.HCl pH 7.8	50mM
(Final conc.)	$MgCl_2$	10mM
	DTT	1mM
	ATP	1mM
	BSA	100µg/ml
	Hexamminecobalt (III) chloride	1mM

7.1.5 Klenow Reaction

DNA fragments were sometimes rendered blunt-ended by the use of Klenow fragment (Sambrook *et al.*, 1989). After digestion, the solution was made to 25μM for each of the dNTPs and 1 unit klenow enzyme added. The mixture was incubated at 30°C for 15 minutes and then the enzyme was deactivated by heating to 75°C for 10 minutes. The DNA was then dialysed against water for 3-5 minutes prior to further manipulations.

7.1.6 Dephosphorylation of Vectors

To increase the efficiency of ligation reactions by preventing re-circularization of the vector, 1 unit Calf Intestinal Phosphatase (Boehringer Mannheim) was added to the linearised vector with CIP Buffer and the mixture incubated at 37°C for 30 minutes. The vector was then gel purified prior to ligation.

CIP Buffer:	Tris.HCl pH 9.0	50mM
(Final conc.)	$MgCl_2$	1mM
	$ZnCl_2$	0.1mM
	Spermidine	1mM

7.1.7 DNA Sequence Analysis

7.1.7.1 DNA Template Preparation

DNA was sequenced directly from plasmid (Hattori & Sakaki, 1986), templates being prepared as follows: 1-2 pmoles (usually about 2-4μg) plasmid DNA were diluted to 18μl with ddH_2O and denatured by the addition of 2μl 2M NaOH and incubation for 5 minutes at room temperature. 8μl 5M ammonium acetate (pH 7.5, filter sterilised) was added and the DNA precipitated by the addition of 100μl ethanol and incubating the tube at -70°C for 5 minutes. The denatured plasmid was recovered by centrifugation at 13000rpm for 15 minutes in a microfuge at 4°C and the supernatant discarded. The pellet was washed with 500μl 75% ethanol, microfuged for 2 minutes

at 4°C and dried under vacuum. The prepared template was finally resuspended in 7µl ddH_2O and sequenced.

7.1.7.2 Dideoxy Sequencing

Chain-termination sequencing (Sanger *et al.*, 1977) was carried out using the Sequenase® Version 2.0 system from United States Biochemical, Cleveland, USA, according to the manufacture's instructions. Briefly, template prepared as described in section 7.1.7.1 was mixed with 1µl (about 0.5pmol) sequencing primer and 2µl 5x reaction buffer and the mixture incubated at 65°C for 2 minutes before being permitted to cool to below 30°C over a period of 30 minutes to anneal the primer to the template.

Reaction Buffer (5x):	Tris.HCl pH 7.5	0.2M
	$MgCl_2$	0.1M
	NaCl	0.25M

The template/primer mix was then placed on ice and 1µl 0.1M DTT, 2µl labelling mix diluted 1:5 in ddH_2O, 0.5µl [α-^{35}S] dATP (10µCi/µl, Amersham) and 2µl Sequenase enzyme (diluted 1:8 in enzyme dilution buffer) were added, mixed and the reaction incubated at room temperature for 2-5 minutes.

Labelling Mix:	dGTP, dCTP, dTTP	7.5µM each
Enzyme Dilution Buffer:	Tris.HCl pH 7.5	10mM
	DTT	5mM
	BSA	0.5mg/ml

Reactions were then teminated by placing 3.5µl reaction into each of four tubes warmed to 37°C, containing one of the four (A, C, G or T) termination mixes. The tubes were incubated a further 3-5 minutes at 37°C and then reaction was halted by the addition of 4µl Stop Solution.

Sequencing reactions were denatured at 90°C for 2 minutes and immediately run on a 6% polyacrylamide, 8M urea 0.4mm sequencing gel prewarmed to 50°C. (Either 1x TBE or a 0.5-5x TBE gradient gel). Upon completion of elecrophoresis, the gel was fixed for 20 minutes in 10% methanol, 10% acetic acid, blotted onto Whatman 3MM paper, dried and exposed to autoradiography film overnight.

Termination Mixes:	dATP, dCTP, dGTP, dTTP	80µM each
	One of ddATP, ddCTP, ddGTP, ddTTP	80µM
	NaCl	50mM
Stop Solution:	Formamide	95% v/v
	EDTA	20mM
	Bromophenol Blue	0.5g/l
	Xylene Cyanol FF	0.5g/l

7.1.8 Targeting Constructs

Targeting vectors are described in their respective chapters. The PGK*neo* component of vector pDWM101neo was derived from pSPGKneo (a gift of Hein te Riele, Amsterdam). This construct is shown in figure 7-1 (a).

The constructs designed for targeting of the K-*ras* gene were based upon pPNT (Tybulewicz *et al.*, 1991), a pUC-based cassette (Yanisch-Perron *et al.*, 1985) containing the *neo* gene as a positive selectable marker (Southern & Berg, 1982) and the HSV*tk* gene (Colbere-Garapin *et al.*, 1979) as a negative selectable marker. Both genes are under the control of the phosphogycerate kinase (PGK) promoter and polyadenylation sequences (Adra *et al.*, 1987). Vector pPNT is shown in figure 7-1 (b).

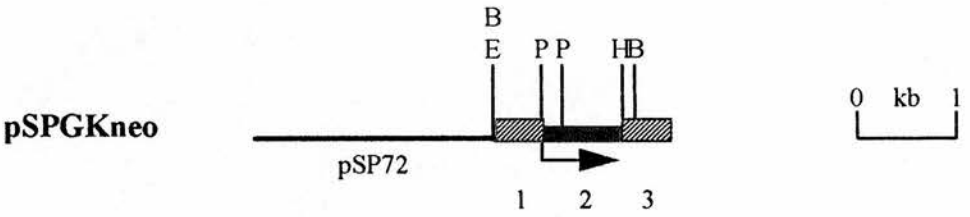
Figure 7-1: Targeting and Selection Constructs

a) **pSPGKneo**. This construct was used as the donor of the 1.5kb *Bgl* II *neo* fragment cloned into vector pDWM101 to generate pDWM101neo (see chapter 2). Selected restriction sites are shown: B, *Bgl* II; H, *Bam*H I; P, *Pst* I; E, *Eco*R I. The numbered blocks represent, respectively: 1, Phosphoglycerate Kinase (PGK) promoter; 2, *neo* gene coding sequence with the orientation and translated region shown by an arrow; 3, PGK polyadenylation sequence. The bacterial plasmid sequence pSP72 is shown as a black line.

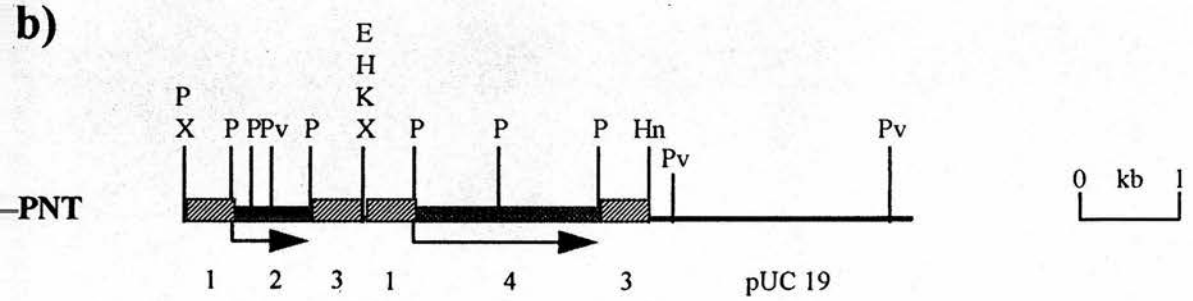
b) **pPNT**. This construct forms the basis of vectors pKiD0PNT, pKiTPNT and pK4BΔPNT (see chapters 3 and 4). Selected restriction sites are shown: H, *Bam*H I; P, *Pst* I; Pv, *Pvu* II; E, *Eco*R I; K, *Kpn* I; Hn, *Hin*d III; X, *Xba* I. The numbered blocks represent, respectively: 1, PGK promoter; 2, *neo* gene coding sequence with the orientation and translated region shown by an arrow; 3, PGK polyadenylation sequence; 4, Herpes Simplex Virus thymidine kinase (HSV*tk*) gene coding sequence with the orientation and translated region shown by an arrow . The bacterial plasmid sequence pUC 18 is shown as a black line.

Figure 7-1

(a)


B
E
P P
H B
pSPGKneo
pSP72
1 2 3
0 kb 1

b)


E
H
K
P
X
P P Pv P
X
P
P
P
Hn
Pv
Pv
PNT
1 2 3 1 4 3
pUC 19
0 kb 1

7.2 Culture of ES Cells and EC Cells

7.2.1 Cell Lines

The following cell lines were used where mentioned in the appropriate chapters:

Cell Line	ES / EC Cell Line	Mouse Strain	Reference
E14	ES	129/Ola	Handyside *et al.*,1989
E14TG2a	ES	129/Ola	Hooper *et al.*, 1987
HM1	ES	129/Ola	Magin *et al.*,1992
PSA4	EC	129/Sv	Martin & Evans, 1975

7.2.2 Routine Culture of ES and EC cells

7.2.2.1 Media

ES cells were cultured in Complete Medium (CM5:5) supplemented with LIF (Leukaemia-inhibiting Factor) (Williams *et al.*, 1988, Smith *et al.*, 1988, Moreau *et al.*, 1988) prepared by transfection of Cos-7 cells (Chen and Okyama, 1987) and 0.1mM β-mercaptoethanol or with CM5:5 supplemented with 60% CM5:5 conditioned by Buffalo Rat Liver (BRL) (Smith and Hooper, 1987) cells, and β-mercaptoethanol. Rarely, a feeder layer was used.

Complete Medium (CM5:5)

Glasgow's Modified Eagle's Medium (GMEM) (Life technologies)*	1x
Foetal Calf Serum & Neonatal Calf Serum	5% v/v each
Non-essential amino acids (Life Technologies):	
Glycine, L-alanine, L-aspartic acid, L-asparagine and L-glutamic acid	0.1mM each
L-proline and L-serine	0.2mM each
Sodium Pyruvate	1.0mM

* 1x GMEM includes L-Glutamine at 2.0mM and 2.5g/l Sodium Bicarbonate, but these were added separately to the prepared medium

For preparation of BRL-conditioned medium, a confluent monolayer of BRL cells in a 175cm^2 flask was incubated with 60ml CM5:5 for 7 days and then the medium was aspirated off and filtered through a sterile 0.8μm filter before storage at -20°C.

To prepare feeder cells, a just-subconfluent layer of primary fibroblasts was cultured for 2 hours in the presence of 10μg/ml mitomycin-C in medium and then washed 3 times in PBS. The cells were removed from the flask by trypsinisation and seeded out at $1x10^6$ cells/ $25cm^2$ flask.

For differentiation experiments, media CMβ (CM5:5 supplemented with β-mercaptoethanol but not LIF) and EFNβ (Medium as above containing β-mercaptoethanol but lacking non-essential amino acids, sodium pyruvate and LIF) were used as described.

7.2.2.2 Serum

Two types of serum were used, foetal calf serum (FCS) and neonatal calf serum (NCS). Batches were purchased from various manufacturers and tested before use by being used to prepare standard medium, with LIF. ES cells were passaged by standard trypsinisation and stopping the trypsin with the medium containing the tested sera; this medium was then used to feed the culture of ES cells (10^3 cells in a 6cm culture dish) for 7 days, re-feeding after 3 days. The cells were stained with 1.5g/l Leishman's stain in methanol, and serum batches were deemed satisfactory if cloning efficiencies of 20-30%, predominantly of undifferentiated cells, were seen.

7.2.2.3 Culture Conditions

ES Cells were routinely cultured by growing in CM 5:5 supplemented with LIF and β -mercaptoethanol on tissue culture treated plastic (Falcon, Nunc or Costar) coated with swine skin gelatin (Sigma). A solution of 10g/l gelatin in ddH_2O was prepared and autoclaved twice. This was then diluted 1:10 in ddH_2O and tissue culture plasticware was coated by covering the surface with the 1g/l gelatin solution and leaving the plate or flask at room temperature for a minimum of 15 minutes. Cultures were kept at a temperature of 37°C, in a humidified atmosphere of 5% CO_2 , 95% air.

Medium was changed as appropriate, usually every day or every other day, and cells were passaged, usually about 1:6-1:10, just before confluence was reached.

7.2.2.4 Passaging ES Cells

To subculture ES cells, the cells were fed with fresh medium about 2 hours prior to passaging. After this time, the medium was aspirated and the cells were washed twice with sterile phosphate buffered saline (PBS, obtained as a tablet concentrate from Flow Laboratories or 10x concentrate from Life Technologies).

Phosphate-Buffered Saline:	NaCl	8g/l
(Calcium- and Magnesium-free)	KCl	0.2g/l
	Na_2HPO_4	1.44g/l
	KH_2PO_4	0.24g/l

The PBS was aspirated and 1ml TVP for a $25cm^2$ flask was added and the cells incubated at 37°C for 3-5 minutes.

TVP:	Trypsin (Flow Laboratories)	0.25g/l
	Chick Serum (Flow Laboratories)	10g/l
	EDTA (Na_2 salt)	1mM
	in Ca/Mg-free PBS, filter-sterilised	

The cells were then disaggregated by a sharp tap to the flask and a minimum of an equal volume of medium was added. The cells were then diluted as appropriate with medium and replated on fresh plasticware coated with 0.1% gelatin. The ES cells were refed after 2 hours.

7.2.2.5 Freezing Cells

ES and EC cells when not required were stored in the vapour over liquid nitrogen. Cells were trypsinised as described in section 7.2.2.4, washed once by neutralising the TVP with 10-20ml CM5:5 and centrifuging the cell suspension for 5 minutes at 1000rpm, and then resuspended in freezing medium. The cells were then cooled at an

approximate rate of 1°C/minute to -70°C overnight, and moved to storage over liquid nitrogen the next day.

Freezing Medium:	CM5:5	80% v/v
	FCS	10% v/v
	DMSO	10% v/v

(rendered slightly acid with sterile 1M HCl and filter-sterilised)

To bring frozen cells back to culture, a freezing vial was removed from liquid nitrogen and thawed in a 37°C water bath as quickly as possible. The cells were washed in 10-20ml CM 5:5 before replating.

7.2.3 Transformation of ES Cells by Electroporation

ES cells were transformed by electroporation, using a BioRad Gene Pulser. 10^8 cells (approximately one confluent 175 cm^2 flask) were trypsinised as above and harvested by centrifugation in a 20ml sterile universal for 5 minutes at 1000rpm. The cells were resuspended in 0.6ml PBS containing 150µg linearised plasmid DNA (verified by agarose gel electrophoresis). The cell/DNA suspension was placed in a 1ml electroporation cuvette (BioRad, path length 0.4cm) and an electric pulse (3µF, 800V, time constant 0.1s) was applied. The ES cells were immediately transferred to 100ml CM5:5 and plated out in 90mm tissue culture dishes. The following day, the medium was aspirated and selective medium added.

7.2.4 Selection of Transformed ES Cell Clones

ES cells plated out at a density of 10^7 cells/90mm plate (unless stated otherwise) were subjected to selection for 10-14 days, until clones were visible to the naked eye (1-2mm in diameter) and were then picked into individual wells of a 24-well plate and disaggregated to permit expansion.

On achieving confluence, wells were divided into two parts. Half the cells were frozen, and the other half were used to prepare genomic DNA for screening of the clones.

Selective Media

CM5:5 supplemented with LIF and β-mercaptoethanol, supplemented with:

(1)	HAT	Hypoxanthine (Sigma) Thymidine (Sigma) Aminopterin (Sigma)	0.1mM 20μM 0.8μM
(2)	6-TG	2-amino-6-mercaptopurine (Sigma)	10μg/ml
(3)	G418	Geneticin Sulphate (Life Technologies)	300μg/ml
(4)	Ganciclovir	Dihydroxy-propoxy-Guanine	2μM

7.2.5 *In-Vitro* Differentiation of ES and EC Cells

$2.5x10^6$ cells were seeded out onto a 60mm tissue culture dish and cultured in EFNβ for 3 days. Clumps of cells were then dislodged from the dish by directing a jet of 5ml medium across the bottom of the dish 3 times, and harvesting the cells by allowing them to sediment out under gravity. The cells clumps were then transferred with a wide-bore pipette to a 100mm dish previously coated with agarose: Solutions of 1% agarose (Miles) and 2% agarose in PBS were autoclaved. A 100mm culture dish was coated with a layer of 2% agarose, removing the excess, and when the layer had set at room temperature, a 1% agarose layer was applied onto this. 10ml EFNβ was added and the agarose layers equilibrated overnight with medium at 37°C, 5% CO_2 in air. The cells were then cultured on agarose for about 6 days with EFNβ, changing the medium every other day, until aggregates of cells with defined cavities were observed. This was defined as "differentiation stage 1". Aggregates were then harvested by sedimentation under gravity, and seeded out onto 4 non-gelatinised 60mm culture dishes, and cultured as a suspension in medium CMβ, changing the

medium as often as required, for 4-5 weeks. The end of this period was defined as "differentiation stage 2". Samples of cells were taken at the end of both stages for the K-*ras* PCR analysis described in section 7.4.3.

7.3 Analysis of ES Cell Clones

7.3.1 Preparation of ES Cell Genomic DNA

ES cells were harvested by brief centrifugation in a microfuge, and then 0.5ml of lysis buffer, supplemented with 600µg proteinase K (Sigma), was added. The cells were incubated overnight at 37°C with shaking. The lysate was extracted twice with an equal volume of phenol to remove the proteins, and then with an equal volume of chloroform. The aqueous phase was taken and an equal volume of propan-2-ol was added to precipitate the DNA. The tube was centrifuged for 2 minutes at 13000rpm in a microfuge and the propan-2-ol/water mix discarded. The DNA pellet was then briefly vacuum-dried and resuspended in 100-500µl (depending on yield) TE buffer.

Lysis Buffer:		
	Tris.HCl pH 7.5	50mM
	EDTA	50mM
	NaCl	100mM
	DTT	5mM
	Spermidine	2mM
	SDS	1% w/v

7.3.2 Restriction Digestion and Agarose Gel Electrophoresis of Genomic DNA

Approximately 10µg (30µl) ES cell clone DNA was digested overnight at 37°C with a 3-fold excess of restriction enzyme in the manufacturer's recommended buffer supplemented with 1mM spermidine. 10% by volume loading dye was then added and the digested DNA fragments were separated by agarose gel electrophoresis in a 0.8% w/v gel in 1x TBE buffer at 40-100mA. The DNA was visualised by staining

for 20 minutes in 0.5μg/ml ethidium bromide and then destained for 20 minutes in dsH_2O.

Loading Dye (10x):	Bromophenol Blue	4g/l
	Ficoll 400	300g/l
TBE Buffer:	Tris-Borate	90mM
	EDTA	2mM
	Made to pH 8.0	

Scaled-down versions of the above procedure were used for routine preparation of DNA fragments for probes, ligations, to estimate DNA concentration, etc.

7.3.3 Southern Analysis of ES Cell Clones

7.3.3.1 Southern Transfer of Digested DNA

Transfer of DNA to hybridisation filters was carried out essentially as described by Southern (1975), with modifications recommended by the membrane manufacturers (Amersham). Briefly, the DNA was partially depurinated by treating the gel with 0.25M HCl until the tracking dye had changed colour, and then rinsed in ddH_2O for 30 minutes. This process is reported to enhance the transfer of larger fragments (Sambrook *et al.*, 1989). A capillary blot was then set up: A tray was filled with 0.4M NaOH, and a platform suspended above it. A wick made of Whatman 3MM paper was laid over this and the gel placed on top. A piece of Amersham Hybond N+ nylon membrane was soaked in 0.4M NaOH and then placed on the gel. Another piece of Whatman 3MM paper, also soaked in 0.4M NaOH was placed over this and then a layer of about 2 inches of paper towels. A weight of approximately 1kg was placed on top. The area surrounding the gel was covered with Saran-Wrap to prevent short-circuiting of the solution resulting in a poor DNA transfer. Gels were blotted overnight.

7.3.3.2 Hybridisation of Southern Blotted DNA

Hybridisations were carried out in a Hybaid hybridisation oven with rotisserie in glass tubes at 65°C. A membrane was wetted in 2xSSC and then rolled up in a fine mesh and placed in the tube. The 2xSSC was then replaced with 25ml Prehybridisation/Hybridisation Solution.

20x Saline Sodium Citrate (SSC)

NaCl	3M
Sodium Citrate	0.3M

Prehybridisation/Hybridisation Solution

Dextran Sulphate	100g/l
SSC	6x
SDS	1% w/v

250µg/ml sonicated salmon sperm DNA (denatured by boiling for 5 minutes) was added immediately before addition of the solution to the hybridisation tube. The nylon filter was prehybridised at 65°C for a minimum of 2 hours prior to adding the probe. Filters were hybridised overnight and then washed the next morning, rinsing out the hybridisation bottle with 2xSSC followed by 4 washes for 15 minutes in 0.1xSSC, 1%SDS at 65°C. The bottle was rinsed again with 2xSSC and the filter wrapped in Saran-Wrap and placed in an autoradiograph cassette with 2 intensifying screens and a piece of autoradiography film (Kodak or Agfa). The cassette was placed at -70°C to expose.

Filters were stripped to be re-probed according to Amersham's directions, briefly, a 30-minute wash at 45°C in 0.4M NaOH followed by a 15-minute wash at 45°C in 0.1xSSC, 0.1% w/v SDS, 0.2M Tris.HCl pH 7.5.

7.3.3.3 Preparation of DNA Probes

DNA probes were prepared by random priming of DNA fragments prepared by restriction endonuclease digestion of the appropriate plasmid, agarose gel

electrophoresis of the digested fragments and use of the Geneclean kit (Bio 101) to purify the desired fragment. 25ng of linear DNA was denatured by boiling for 5 minutes. In the case of the Amersham Megaprime kit, the DNA was then snap-cooled on ice for 5 minutes, but this is not necessary when using the Amersham Rediprime kit. Both kits were used according to the manufacturer's instructions. Briefly, after denaturing the DNA the deoxynucleotides, reaction buffer and 2 units of klenow enzyme were added, either as solutions in the case of the Megaprime kit, or simply resuspended in the solution of denatured probe in the case of the Rediprime kit. 50μ Ci α-^{32}P labelled dCTP (Amersham) was added and the mixture incubated at 37°C for 10 minutes (Rediprime) or 30 minutes (Megaprime). Rediprime kit reactions were then halted by the addition of 1/10 volume of 0.2M EDTA. Probes were separated from free nucleotides using a Pharmacia Nick Column (Sepharose G50) and then denatured by boiling for 5 minutes and snap freezing on ice for 5 minutes. The freshly-labelled probe was then added to the prehybridised filters.

7.3.6 Screening of ES Cell Clones by the Polymerase Chain Reaction

PCR reactions were carried out using Taq DNA polymerase and associated reagents from Life Technologies, deoxynucleotides from Pharmacia and oligonucleotide primers from Oswell on a Hybaid Omnigene heating block.

7.3.6.1 Technical Considerations Pertaining to PCR Conditions

PCR primers where appropriate are cited from the work of their original designers. Other primers were designed by assigning the regions of the gene to be amplified, and selecting a sequence of about 20 base pairs. If possible, a C:G pair would be chosen for the extended end of the primer. Using the rule of thumb that the addition of a A:T base pair would increase melting temperature by 2°C and the addition of a C:G base pair would increase melting temperature by 4°C (Thein & Wallace, 1986), the

melting temperatures of the two primers was kept close, and in the region of 55°C-65°C. Suggested primer sequences were checked using the GCG v7.0 DNA sequence analysis software "squiggles" to ensure the primer would not be excessively prone to intra-molecular hybridisation which could interfere with the annealing step of the PCR reaction.

Reaction programs were devised to minimise the denaturing time, as *Taq* DNA polymerase has a reduced half-life at high temperatures. Initially annealing temperatures were set at 5°C below the calculated melting temperature of the template/primer hybrid, and refined empirically from this starting point. Extension times were set bearing in mind that the maximum rate of synthesis of *Taq* DNA polymerase is approximately 1000 base pairs per minute.

7.3.6.2 Codon 12 Activating Mutation Diagnostic PCR

This PCR introduced a diagnostic *Hph* I restriction endonuclease site in the product if the murine K-*ras* proto-oncogene template contained an activating glycine→ aspartic acid mutation in codon 12 (Kumar & Dunn, 1989). The primers are shown below:

Forward Strand: 5' > ACT TGT GGT GGT TGG AGG TG < 3'

Reverse Strand: 5' > TCC ACA AAG TGA TTC TGA AT < 3'

A 50µl reaction was set up, consisting of 50pmoles each primer, Life Technologies Polymerase Reaction Buffer (20mM Tris.HCl pH 8.4, 50mM KCl final concentration), 1.5mM $MgCl_2$, 100µM each for dATP, dCTP, dGTP and dTTP, 5µl genomic DNA prepared as above as template and 2 units of Taq DNA polymerase. 50µl of paraffin oil was layered over the top of the reaction mixture, and the reaction was run according to the program on the following page.

1 Cycle	94°C	3 minutes
32 Cycles	94°C	30 seconds
	58°C	30 seconds
	72°C	30 seconds
1 Cycle	72°C	10 minutes

The PCR product was then digested with 2 units of *Hph* I (New England Biolabs) at 37°C overnight and the digested products separated by electrophoresis in 1xTBE, 20% polyacrylamide to detect the presence or absence of the *Hph* I site. The PCR product is a 75-mer of part of K-*ras* exon 1 which may be cloven into a pair of fragments 46bp and 29bp in size by *Hph* I if there is a G→A substitution in the first position of codon 12.

7.3.6.3 Neomycin Resistance Gene PCR

To detect the presence of the *neo* gene in an ES cell's genome, this PCR was used. The primers are shown below:

Forward Strand: 5' > GCG ATG CCT GCT TGC CGA < 3'

Reverse Strand: 5' > GAA GGC GAT AGA AGG CGA < 3'

A 50µl reaction was set up, consisting of 45pmoles of each primer, Life Technologies Polymerase Reaction Buffer (20mM Tris.HCl pH 8.4, 50mM KCl final concentration), 1.5mM $MgCl_2$, 20µM each for dATP, dCTP, dGTP and dTTP, 5µl genomic DNA prepared as above as template and 2 units of Taq DNA polymerase. 50µl of paraffin oil was layered over the top of the reaction mixture, and the reaction was run according to the following program:

1 Cycle	93°C	3 minutes
35 Cycles	93°C	1 minute
	60°C	1 minute

	72°C	1 minute
1 Cycle	72°C	10 minutes

An aliquot of the reaction product was then run on a 2% agarose gel; presence of a 215bp band indicated presence of the *neo* gene.

7.3.6.4 K-*Ras* Exon 4B PCR

This PCR was designed to provide an easy diagnostic test for the point mutation which was designed to place a premature nonsense codon in exon 4B of the murine K-*ras* gene. The exon is amplified and digested with the restriction endonuclease BstBI. Loss of the restriction site indicates that the mutation is present.

The reverse strand primer was that described in section 7.4.3, and the forward strand primer is shown below:

Forward Strand: >5' GGA ATT CCT ATA TTT CAG GGT GTT GAC GAT <3'

This primer contains an *Eco*R I restriction endonuclease site at the 5' end should it ever be desired to clone the PCR product. A 100µl reaction was set up consisting of 50pmoles of each primer, Life Technologies Polymerase Reaction Buffer (20mM Tris.HCl pH 8.4, 50mM KCl final concentration), 1.5mM $MgCl_2$, 20µM each for dATP, dCTP, dGTP and dTTP, 10µl genomic DNA prepared as above as template and 2 units of Taq DNA polymerase. 100µl of paraffin oil was layered over the top of the reaction mixture, and the reaction was run according to the following program:

1 Cycle	94°C	3 minutes
30 Cycles	94°C	30 seconds
	55°C	30 seconds
	72°C	30 seconds
1 Cycle	72°C	10 minutes

The PCR product was extracted with chloroform to remove any paraffin oil and digested at 65°C with the restriction enzyme *Bst*B I (New England Biolabs) in the manufacturer's recommended buffer overnight. The digested products were run on a 2% agarose gel to check for loss of the *Bst*B I site; the full-length PCR product is 117bp in size, and is cloven into a pair of fragments 50bp and 67bp in size if the *Bst*B I site diagnostic of the wild-type allele is present.

7.4 Reverse Transcription PCR of K-*Ras* mRNA

7.4.1 Sample Preparation

Mice were killed and dissected in PBS. Tissue samples were placed in cryo-preservation tubes, snap-frozen in liquid nitrogen and stored in the vapour over liquid nitrogen until required.

7.4.2 Preparation of RNA using TRIzol Reagent

When working with RNA all plasticware was either guaranteed RNAse-free by the manufacturer, or immersed overnight in ethanol containing 0.1% v/v diethyl pyrocarbonate prior to autoclaving. 0.1% v/v DEPC was added to ddH_2O prior to autoclaving, and this was used for the making of all solutions.

TRIzol™ Reagent (Life Technologies) was used for the preparation of RNA according to the manufacturer's instructions. TRIzol is a mono-phasic solution of phenol and guanidine isothiocyanate, and based on the method of Chomczynski and Sacchi (1987). Briefly, a small sample of tissue was mixed with 1ml TRIzol, homogenised and incubated at room temperature for 5 minutes. 0.2ml chloroform was added, the mixture inverted several times and then incubated at room temperature for 2-3 minutes. The tube was centrifuged at 13000rpm in a microfuge for 15 minutes at 4°C and the aqueous phase recovered. 0.5ml propan-2-ol was

added to precipitate the RNA which was recovered by microfuging at 13000rpm for 10 minutes at 4°C. The supernatant was discarded and the RNA pellet washed with 1ml 75% ethanol, then microfuged at 6500rpm for 5 minutes at 4°C. The ethanol was discarded and the RNA pellet was briefly dried under vacuum, prior to being resuspended in 20µl RNAse-free ddH_2O. The pellet was pipetted a few times and then incubated at 50°C for 10 minutes to aid solvation.

RNA was quantified spectrophotometrically by taking readings at 260nm and 280nm. An $O.D._{260}$ of 1 corresponds to 40µg/ml RNA. The ratio of $O.D._{260}/O.D._{280}$ was calculated and deemed satisfactory if it was in the range 1.8-2.0. A ratio significantly lower than 1.8 suggests contamination of the RNA with phenol or protein; higher than 2.1 that the RNA is degraded and unsuitable for reverse transcription reactions. For samples used in the experiments described in chapter 5, RNA quantities are necessarily very small because of the amount of source material available, and therefore representative samples from a set of preparations were measured as it was impossible to test every sample to be used.

7.4.3 Reverse Transcription

Approximately 1-5µg RNA was made up to 11µl with RNAse-free ddH_2O and 1µl (0.5µg) oligo $(dT)_{12\text{-}18}$ primer (Life Technologies) was added. The mixture was incubated at 70°C for 10 minutes and then chilled on ice. The reaction was then made to 20µl with Reaction Mix, microfuged briefly and incubated at room temperature for 10 minutes, then 37°C for 60 minutes, and finally terminated at 90°C for 10 minutes.

Reaction Conditions:		
	Tris.HCl pH 8.3	50mM
	KCl	75mM
	$MgCl_2$	3mM
	DTT	10mM
	dNTPs	0.5mM

The reaction was chilled on ice for 10 minutes, 1-4 units RNAse H were added and the reaction incubated at 37°C for 20 minutes.

7.4.4 K-*Ras* PCR of Murine cDNA

10µl cDNA was used as template in the PCR reaction. The primers are shown below:

Forward: 5' > GGA ATT CCG CCT GCT GAA AAT GAC TGA GT < 3'

Reverse: 5' > CGG GAT CCC GTG TAC ACC TTG TCC TTG ACT T < 3'

These primers incorporate an *Eco*R I (Forward) and a *Bam*H I (Reverse) restriction endonuclease site to permit easy cloning of the products when required. A 50µl reaction was set up, consisting of 5pmoles each primer, Life Technologies Polymerase Reaction Buffer (20mM Tris.HCl pH 8.4, 50mM KCl final concentration), 1.5mM $MgCl_2$, 100µM each of dATP, dCTP, dGTP and dTTP and 2 units of Taq DNA polymerase. 50µl of paraffin oil was layered over the top of the reaction mixture, and the reaction was run according to the program overleaf.

1 Cycle	94°C	3 minutes
21 Cycles	94°C	1 minute
	55°C	1 minute
	72°C	2 minutes
1 Cycle	72°C	10 minutes

The PCR generates two products 687bp and 565bp in size, representing K-*ras*^{A} and K-*ras*^{B} respectively. This PCR was also used, at 32 cycles, for the preparation of murine cDNAs to be cloned as probes, or to provide template cDNAs from clones to be screened by other methods, e.g. that described in section 7.3.6.

7.5 *Hprt* Enzyme Analysis

7.5.1 Protein Extract Preparation

A confluent 75cm^2 or 162cm^2 flask of ES cells was harvested as described, and the cells washed twice in 10ml CM5:5, twice in 10ml PBS and then stored as a pellet at -70°C until required. The pellet was resuspended in 0.5ml Extraction Buffer and subjected to 3 freeze/thaw cycles, with agitation, in a propan-2-ol/dry ice and 37°C water bath, respectively. The lysate was then cleared by centrifugation for 15 minutes at 30,000xg and the supernatant protein extract was retained and stored at -70°C until required.

Extraction Buffer:	Sodium Phosphate	50mM
	β-mercaptoethanol	10mM
	pH 7.0	

7.5.2 Protein Quantification

The Bio-Rad protein assay, (Bradford 1976), was used according to the manufacturer's instructions. Briefly, volumes of ES cell extract (1-20µl) or known quantities of BSA as a standard protein (0-1.4mg/ml) were diluted to 0.1ml with extraction buffer, and 5ml 1:5 diluted and filtered dye reagent was added and the solution mixed by inversion. After a minimum of 5 minutes at room temperature, the OD_{595} was read versus the protein blank. Multiple linear regression was used to calculate protein concentrations.

7.5.3 HPRT Enzyme Activity Assay (Gillin *et al.*, 1972)

50µl reactions were set up in eppendorf tubes, consisting of 25µl 2xReaction Mix, 1-10µg protein extract, and ddH_2O to a total volume of 50µl. Identical controls for each tube also containing 30mM EDTA were run in parallel. Reactions were

staggered at 20 second intervals, being started by the addition of either water (blanks) or protein extract (samples) to the pre-incubated at 37°C reaction tube.

2xRM:	100µl	1M	Tris.HCl pH 7.4
	200µl	1mM	Hypoxanthine
	500µl	2mM	Phosphoribosyl Pyrophosphate (freshly prepared)
	100µl	0.1M	$MgCl_2$
	50µl	50µCi/ml	8-^{14}C Hypoxanthine (Amersham)
	50µl		ddH_2O

Reactions were incubated for 40 minutes and then stopped by the addition of 1ml ice-cold TB (samples) or 1ml TB-E (EDTA controls).

TB:	Tris.HCl pH 7.0	1mM
	KCl	1mM
	EDTA	1.5mM
TB-E:	Tris.HCl pH 7.0	1mM
	KCl	1mM

The sample was then filtered through a Whatman DE82 ion exchange disc with a Millipore vacuum manifold. The tube was rinsed out with 1ml TB-E and this also washed through the filter disc. The filter was washed twice with 5ml TB-E and then dried overnight. The following day, the filters were counted in a scintillation counter.

7.6 Chimaera Production

7.6.1 Mouse Strains

The ES cells used (E14 and HM1) are both derived from the 129/Ola strain of mouse, which carries the mutant alleles c^{ch} and *p* and as such give rise to a light yellow mouse with pink eyes. They have the *GPI-1s*^{a} isotype . Host blastocysts were F2 crosses of CBAxC57BL/6, and thus give dark mice (25% black and 75% grey), CBA being agouti and C57BL/6 black (*a/a*). They are homozygous for the *GPI-1s*^{b}

isotype. Chimaerism in the coat could thus be estimated from the degree of light coat colour on a dark mouse, and in other tissues from the relative proportion of *GPI-1s*^{a}.

7.6.2 Collection of Embryos

Pregnant mice were killed 3.5 days post-coitum by cervical dislocation. The uterine horns were dissected out, and the blastocysts flushed out by the insertion of a 25 gauge needle into the horn and injecting 1ml CM5:5 medium. The blastocysts were collected in a 60mm tissue culture dish, from which they were picked out using a binocular microscope and mouth-controlled pasteur pipette into a drop of medium. This was kept under liquid paraffin at 37°C, 5% CO_2 in air until the embryos were injected.

7.6.3 Injection of Blastocysts

About 1ml CM5:5 was placed at the bottom of a 90mm plastic petri dish and covered with liquid paraffin. The blastocysts and a single-cell suspension of ES cells, prepared by routine trypsinisation and resuspension in CM 5:5, were blown into this using a mouth-controlled pasteur pipette. Injections were carried out using a Leitz Labovert FS micromanipulator with a binocular microscope. About 12 ES cells, picked for good morphology, were injected into each blastocyst, using a hand-made, rounded-end glass pipette to hold the blastocyst and a narrower glass needle, pulled on a Camden Instruments computer controlled electrode puller (model 773), and given a "pen-nib" shaped point by hand, to pick up the ES cells themselves and inject them through the zona pellucida and trophoblast layer of the blastocyst into the blastocoel. The injected embryos were returned to the incubator, where they re-expanded, until transfer to the foster-mother.

7.6.4 Return of Blastocysts to Mothers

Pseudo-pregnant mice were anaesthetized by the intra-peritoneal injection of 0.1ml/10g body weight Hypnorm/Hypnovel mix (50µg/ml fentanyl citrate, 1.7mg/ml fluanisone,1.7mg/ml midazolam hydrochloride). Under a binocular microscope, the appropriate area of the back was sterilised with 70% ethanol, and the end of the uterine horn was carefully exposed, being located by its proximity to the slightly darker ovary and the associated fat pad. A small hole was introduced into the uterus wall with a 25 gauge needle, and then the blastocysts (8-12/each uterine horn) were blown into the lumen through this incision using a mouth-controlled pasteur pipette. The major incision was then sutured using a stitch for the peritoneum and 2 or 3 stitches, as required, for the skin. The mouse was then provided with a separate cage, and kept warm over night.

7.7 Glucose Phosphate Isomerase Isozyme Analysis

The ubiquitous GPI-1 enzyme has two isoforms which can be distinguished electrophoretically. The *Gpi-1s*^{a} isoform is present in strain 129 (from which the ES cell lines used are derived), and the *Gpi-1s*^{b} isoform in CBA and C57BL/6 (F2 hybrids of which were used as host blastocysts). As described in Ansell and Micklem (1986), this may be exploited to identify contribution of ES cell-derived tissue in a chimaeric tissue of interest. 20mg tissue was homogenised with 50µl sample buffer and samples stored at -20°C until required.

Sample Buffer:	Triethanolamine-HCl pH 7.6	50mM
	Dithioerythritol	2mM
	BSA	0.5mgl/ml
	Digitonin	1.6mM

50ml Supre Haem buffer (Helena Laboratories, #5802) was placed in each electrode reservoir of the gel tanks and wet Whatman 3MM paper wicks placed over the supports to contact the buffer. Cellulose acetate sheets were pre-soaked in the supre

haem buffer for 20 minutes, 8µl sample was loaded at the anode of the sheet with an applicator loop and the sample electrophoresed for 90 minutes at 4°C, constant voltage 350V. The sheet was stained by pouring on 1.2% agarose in water containing 2.5ml Stain Solution and incubating in a dark box until bands appeared.

Stain Solution:	0.4ml	Tris.HCl pH 8.0
(Per sheet)	1.0ml	Assay Stock
	15µl	Glucose-6-phosphate dehydrogenase
	0.5ml	*10mg/ml Methyl thiazolyl tetrazolium
	0.5ml	*2.5mg/ml Phenazine methosulphate
	* Added just before use	
Assay Stock:	0.3mg/ml	NADP
	6.5mg/ml	Fructose-6-Phosphate
	in 1M Tris.HCl pH 8.0, stored at -20°C	

The acetate plate was fixed and stored in a solution of 15% glycerol, 3% acetic acid, and photographed.

7.8 Appendix: Abbreviations

6-TG	6-Thioguanine
bp	Base Pairs
BRL	Buffalo Rat Liver
CM	Complete Medium
ddH_2O	Distilled, De-ionised Water
DMSO	Dimethyl Sulphoxide
dNTP(s)	Deoxyribonucleotide Triphosphate(s)
DNA	Deoxyribose Nucleic Acid
DTT	Dithiothreitol
EC cell(s)	Embryonal Carcinoma cell(s)
ECR	Extra-Chromosomal Recombination
EDTA	Ethylene Diamine Tetra-Acetic Acid
EFNβ	Eagle's medium, with Foetal and New-born calf sera, supplemented with β-mercaptoethanol
ES cell(s)	Embryonal Stem cell(s)
FCS	Foetal Calf Serum
GANC	Ganciclovir
GAP	GTPase Activating Protein
GMEM	Glasgow's Modified Eagle's Medium
HAT	Hypoanthine-Aminopterin-Thymidine
hprt	Hypoxanthine-guanine Phosphoribosyl Transferase

HR	Homologous Recombination
ICR	Intra-Chromosomal Recombination
kb	1000 Base Pairs
K-*ras*	Kirsten *ras* proto-oncogene
LIF	Leukaemia Inhibitory Factor
mRNA	Messenger Ribose Nucleic Acid
NaCl	Sodium Chloride
NCS	Newborn Calf Serum
PBS	Phosphate Buffered Saline
PCR	Polymerase Chain Reaction
PNS	Positive-Negative Selection
RNA	Ribose Nucleic Acid
RT-PCR	Reverse Transcription-Polymerase Chain Reaction
SDS	Sodium Dodecyl Sulphate
SSC	Saline Sodium Citrate

CHAPTER 8-
REFERENCES

References

Adair, G. M., Nairn, R. S., Wilson, J. H., Seidman, M. M., Brotherman, K. A., McKinnon, C. & Scheerer, J. B. (1989). Targeted homologous recombination at the endogenous adenine phosphoribosyltransferase locus in Chinese hamster cells. *Proceedings of the National Academy of Sciences of the USA* **86** 4574-4578.

Adra, C. N., Boer, P. H. & McBurney, M. W. (1987). Cloning and expression of the mouse PGK-1 gene and the nucleotide sequence of its promoter. *Gene* **60** 65-74.

Ahmad, K. & Naz, R. K. (1993). Presence and possible role of c-ras and nuclear (c-fos and c-jun) proto-oncogene products in preimplantation embryonic development in mice. *Molecular Reproduction and Development* **36** 297-306.

Al-Shawi, R., Burke, J., Wallace, H., Jones, C., Harrison, S., Buxton, D., Maley, S., Chandley, A. & Bishop, J. O. (1991). The herpes simplex virus type 1 thymidine kinase is expressed in the testes of transgenic mice under the control of a cryptic promoter. *Molecular and Cellular Biology* **11** 4207-4216.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. & Watson, J. D. (1994). In *Molecular Biology of the Cell*, 3rd edn. pub. Garland, New York & London.

Ansell, J. D. & Micklem, H. S. (1986) Genetic markers for following cell populations. Handbook of Experimental Immunology, 4th edn. vol. 2, chapter 56. Ed. D. M. Weir, L.A. Herzenberg, C. D. Blackwell & L. A. Herzenberg. Oxford: Blackwell Scientific Publications.

Askew, G. R., Doetschman, T. & Lingrel, J. B. (1993). Site-directed point mutations in embryonic stem cells: a gene targeting tag-and-exchange strategy. *Molecular and Cellular Biology* **13** 4115-4124.

Avruch, J., Zhang, X-f. & Kyriakis, J. M. (1994). Raf meets Ras: completing the framework of a signal transduction pathway. *Trends in Biochemical Sciences* **19** 279-283.

Ayares, D., Chekuri, L., Song, K.-Y., Kucherlapati, R. (1986). Sequence homology requirements for intermolecular recombination in mammalian cells. *Proceedings of the National Academy of Sciences of the USA* **83** 5199-5203.

Balmain, A., Ramsden, M., Bowden, G. T. & Smith, J. (1984). Activation of the mouse cellular Harvey-*ras* gene in chemically induced benign skin papillomas. *Nature* **307** 658-660.

Barnes, W. M. (1994). PCR amplification of up to 35-kb DNA with high fidelity and high yield from λ bacteriophage templates. *Proceedings of the National Academy of Sciences of the USA* **91** 2216-2220.

Bar-Sagi, D. & Feramisco, J. (1986). Induction of membrane ruffling and fluid phase pinocytosis in quiescent fibroblasts by *ras* proteins. *Science* **233** 1061-1068.

Barbacid, M. (1987). *ras* genes. *Annual Review of Biochemistry* **56** 779-827.

Bautista, D. & Shulman, M. J. (1993). A hit-and-run system for introducing mutations into the Ig H chain locus of hybridoma cells by homologous recombination. *Journal of Immunology* **151** 1950-1958.

Beddington, R. S. P. & Robertson, E. J. (1989). An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development* **105** 733-737.

Beitel, G. J., Clark, S. G. & Horvitz, H. R. (1990). *Caenorhabditis elegans ras* gene *let-60* acts as a switch in the pathway of vulval induction. *Nature* **348** 503-509.

Bernetgrandaud, A., Ouazana, R., Morle, F. & Godet, J. (1992). A method for improving the efficiency of the positive negative selection used to isolate homologous recombinants. *Nucleic Acids Research* **20** 6417-6418.

Birnboim, H. C. & Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research* **7** 1513-1523

Boguski, M. S. & McCormick, F. (1993). Proteins regulating Ras and its relatives. *Nature* **366** 643-654.

Bollag, G. & McCormick, F. (1991). Regulators and effectors of *ras* proteins. *Annual Review of Cell Biology* **7** 601-632.

Bollag, R. J., Waldman, A. S. & Liskay, R. M. (1989) Homologous recombination in mammalian cells. *Annual Review of Genetics* **23** 199-225.

Bos, J. L. (1989). *ras* oncogenes in human cancer: a review. *Cancer Research* **49** 4682-4689.

Bourne, H. R., Sanders, D. A. & McCormick, F. (1991). The GTPase superfamily: conserved structure and molecular mechanism. *Nature* **349** 117-127.

Bowtell, D., Fu, P., Simon, M. & Senior, P. (1992). Identification of murine homologues of the *Drosophila Son of Sevenless* gene: Potential activators of Ras. *Proceedings of the National Academy of Sciences of the USA* **89** 6511-6515.

Bradford, M. M. (1976). A rapid and sensitive method for quantitation of microgram quantities of protein utilising the principle of protein-dye binding. *Analytical Biochemistry* **72** 248-254.

Bradley, A. (1987). Production and analysis of chimaeric mice. In *Teratocarcinomas and embryonic stem Cells-A Practical Approach*, ed. E. J. Robertson, pp. 113-151. Oxford: IRL Press.

Bradley, A. (1993). Site directed mutagenesis in the mouse. *Recent Progress in Hormone Research* **48** 237-251.

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

Brewer, L. M. & Brown, N. A. (1992). Distribution of $p21^{ras}$ in postimplantation rat embryos. *The Anatomical Record* **234** 443-451.

Brinster, R. L., Braun, R. E., Lo., D., Avarbock, M. R., Oram, F. & Palmiter, R. D. (1989). Targeted correction of a major histocompatibility class II E_{α} gene by DNA microinjected into mouse eggs. *Proceedings of the National Academy of Sciences of the USA* **86** 7087-7091.

Cai, H., Szeberenyi, J. & Cooper, G. M. (1990). Effect of a dominant inhibitory Ha-*ras* mutation on mitogenic signal transduction in NIH3T3 cells. *Molecular and Cellular Biology* **10** 5314-5323.

Camerini-Otero, R. D. & Kucherlapati, R. (1990). Right on Target. *The New Biologist* **2** 337-341.

Capon, D. J., Seeburg, P. H., McGrath, J. P., Hayflick, J. S., Edman, U., Levinson, A. D. & Goeddel, D. V. (1983) Activation of Ki-*ras*2 gene in human colon and lung carcinomas by two different point mutations. *Nature* **304** 507-513.

Cases, S. & Dautry, F. (1992) Inactivation of the murine N-*ras* gene by gene targeting. *Oncogene* **7** 2525-2528.

Cavalli-Sforza, L. L. & Lederberg, J. (1956). Isolation of pre-adaptive mutants in bacteria by sib selection. *Genetics* **41** 367-381.

Chang, E. H., Furth, M. E., Scolnick, E. M. & Lowy, D. R. (1982). Tumorigenic transformation of mammalian cells induced by a normal human gene homologous to the oncogene of Harvey murine sarcoma virus. *Nature* **297** 479-483.

Chardin, P., Camonis, J. H., Gale, N. W., Van Aelst, L., Schlessinger, J., Wigler, M. H. & Bar-Sagi, D. (1993). Human Sos1: A guanine nucleotide exchange factor for Ras that binds to GRB2. *Science* **260** 1338-1343.

Charron, J., Malynn, B. A., Fisher, P., Stewart, V., Jeanotte, L., Goff, S. P., Robertson, E. J. & Alt, F. W. (1992). Embryonic lethality in mice homozygous for a targeted disruption of the N-*myc* gene. *Genes and Development* **6** 2248-2257.

Chen, C. & Okayama, H. (1987). High-efficiency transformation of mammalian cells by plasmid DNA. *Molecular and Cellular Biology* **7** 2745-2752.

Cheng, S., Fockler, C., Barnes, W. M. & Higuchi, R. (1994). Effective amplification of long targets from cloned inserts and human genomic DNA. *Proceedings of the National Academy of Sciences of the USA* **91** 5695-5699.

Chesa, P. G., Rettig, W. J., Melamed, M. R., Old, L. J. & Niman, H. L. (1987). Expression of p21ras in normal and malignant human tissues: lack of association with proliferation and malignancy. *Proceedings of the National Academy of Sciences of the USA* **84** 3234-3238.

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

Colbere-Garapin, F., Chousterman, S., Horodniceau, F., Kourilsky, P. & Garapin, A.-C. (1979). Cloning of the active thymidine kinase gene of herpes simplex virus type 1 in *Escherichia coli* K-12. *Proceedings of the National Academy of Science of the USA* **76** 3755-3759.

Colledge, W. H., Ratcliff, R., Foster, D., Williamson, R. & Evans, M. J. (1992). Cystic fibrosis mouse with intestinal obstruction. *Lancet* **340** 680.

Cui, K.-H., Putland, R. A., Seamark, R. F. & Matthews, C. D. (1993). Precise sex selected births of mice following single cell embryo biopsy and Y-linked testis-specific gene analysis. *Human Reproduction* **8** 621-626.

Davis, A. C., Wims, M. & Bradley, A. (1992). Investigation of coelectroporation as a method for introducing small mutations into embryonic stem cells. *Molecular and Cellular Biology* **12** 2769-2776.

DeFeo, D., Gonda, M. A., Young, H. A., Chang, E. H., Lowy, D. R., Scolnick, E. M. & Ellis, R. W. (1981). Analysis of two divergent rat genomic clones homologous to the transforming gene of Harvey murine sarcoma virus. *Proceedings of the National Academy of Science of the USA* **78** 3328-3332.

DeGregori, J., Russ, A., von Melchner, H., Rayburn, H., Priyaranjan, P., Jenkins, N. A., Copeland, N. G. & Ruley, H. E. (1994). A murine homolog of the yeast RNA1 gene is required for postimplantation development. *Genes and Development* **8** 265-276.

Deng, C. & Capecchi, M. R. (1992). Re-examination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus. *Molecular and Cellular Biology* **12** 3365-3371.

Deng, C., Thomas, K. R. & Capecchi, M. R. (1993). Location of crossovers during gene targeting with insertion and replacement vectors. *Molecular and Cellular Biology* **13** 2134-2140.

Detlott, P. J., Lewis, J., John, S. W. M., Shehee, W. R., Langenbach, R., Maeda, N. & Smithies, O. (1994). Deletion and replacement of the mouse adult β-globin genes by a "plug and socket" repeated targeting strategy. *Molecular and Cellular Biology* **14** 6936-6943.

Dickinson, P., Kimber, W. L., Kilanowski, F. M., Stevenson, B. J., Porteous, D. J. & Dorin, J. R. (1993). High frequency gene targeting using insertional vectors. *Human Molecular Genetics* **2** 1299-1302.

Dixon, W. J. (1953). Processing data for outliers. *Biometrics* **9** 74-89.

Doetschman, T., Gregg, R.G., Maeda, N., Hooper, M. L., Melton, D. W., Thompson, S. & Smithies, O. (1987). Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature* **330** 576-578.

Doetschman, T., Maeda, N & Smithies, O. (1988). Targeted Mutation of the *Hprt* gene in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences of the USA* **85** 8583-8588.

Dorin, J. R., Dickinson, P., Alton, E. W. F. W., Smith, S. N., Geddes, D. M., Stevenson, B.J., Kimber, W. L., Fleming, S., Clarke, A. R., Hooper, M. L., Anderson, L., Beddington, R. S. P. & Porteous, D. J. (1992). Cystic-fibrosis in the mouse by targeted insertional mutagenesis. *Nature* **359** 211-215.

Dorin, J. R., Inglis, J. D. & Porteous, D. J. (1989). Selection for precise chromosomal targeting of a dominant marker by homologous recombination. *Science* **243** 1357-1360.

Downward, J., Graves, J. D., Warne, P. H., Rayter, S. & Cantrell, D. A. (1990). Stimulation of p21*ras* upon T-cell activation. *Nature* **346** 719-723.

Dugaiczyk, A., Boyer, H. W. & Goodman, H. M. (1975). Ligation of *Eco*R I endonuclease-generated DNA fragments into linear and circular structures. *Journal of Molecular Biology* **96** 171-184.

Ellis, J. & Bernstein, A. (1989). Gene targeting with retroviral vectors: recombination by gene conversion into regions of nonhomology. *Molecular and Cellular Biology* **9** 1621-1627.

Ellis, R. W., DeFeo, D., Shih, T. Y., Gonda, M. A., Young, H. A., Tsuchida, N., Lowy, D. R. & Scolnick, E. M. (1981). The p21 *src* genes of Harvey and Kirsten sarcoma viruses originate from divergent members of a family of normal vertebrate genes. *Nature* **292** 506-511.

Ernst, M., Gearing, D. P. & Dunn, A. R. (1994). Functional and biochemical association of Hck with the LIF/IL-6 receptor signal transducing subunit gp130 in embryonic stem cells. *EMBO Journal* **13** 1574-1584.

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

Feig, L. A. (1993). The many roads that lead to Ras. *Science* **260** 767-768.

Feig, L. A. (1994). Guanine-nucleotide exchange factors: a family of positive regulators of Ras and related GTPases. *Current Opinion in Cell Biology* **6** 204-211.

Feramisco, J. R., Gross, M., Kamata, T., Rosenberg, M. & Sweet, R. W. (1984). Microinjection of the oncogene form of the human H-*ras* (T24) protein results in rapid proliferation of quiescent cells. *Cell* **38** 109-117.

Feuerstein, J., Kalbitzer, H. R., John, J., Goody, R. S. & Wittinghofer, A. (1987a). Characterisation of the metal-ion-GDP complex at the active sites of transforming and non-transforming p21 proteins by observation of the 17O-Mn superhyperfine coupling and by kinetic methods. *European Journal of Biochemistry* **162** 49-55.

Feuerstein, J., Goody, R. S. & Wittinghofer, A. (1987b). Preparation and characterization of nucleotide-free and metal ion-free p21 "apoprotein". *Journal of Biological Chemistry* **262** 8455-8458.

Field, J., Broek, D., Kataoka, T. & Wigler, M. (1987). Guanine nucleotide activation of, and competition between, RAS proteins from *Saccharomyces cerevisiae*. *Molecular and Cellular Biology* **7** 2128-2133.

Field, J., Nikawa, J., Broek, D., MacDonald, B., Rodgers, L., Wilson, I. A., Lerner, R. A. & Wigler, M. (1988). Purification of a *Ras*-responsive adenylyl cyclase complex from *Saccharomyces cerevisiae* by use of an epitope addition method. *Molecular and Cellular Biology* **8** 2159-2165.

Fiering, S., Kim, C. G., Epner, E. M. & Groudine, M. (1993). An "in-out" strategy using gene targeting and FLP recombinase for the functional dissection of complex DNA regulatory elements: Analysis of the β-globin locus control region. *Proceedings of the National Academy of Sciences of the USA* **90** 8469-8473.

Finney, R. E. & Bishop, J. M. (1993). Predisposition to neoplastic transformation caused by gene replacement of H-*ras*1. *Science* **260** 1524-1527.

Folger, K. R., Thomas, K. & Capecchi, M. R. (1985). Nonreciprocal exchanges of information between DNA duplexes coinjected into mammalian cell nuclei. *Molecular and Cellular Biology* **5** 59-69.

Fox, N., Damjanov, I., Martinez-Hernandez, A., Knowles, B. B. & Solter, D. (1981). Immunohistochemical localization of the early embryonic antigen (SSEA-1) in postimplantation mouse embryos and fetal and adult tissues. *Developmental Biology* **83** 391-398.

Frohman, M. A. & Martin, G. R (1990). Detection of homologous recombinants. In *PCR Protocols: a guide to methods and applications*, pp. 228-236. Ed. M. A. Innis, D. H. Gelfand, J. J. Sninsky & T. J. White. San Diego: Academic Press, Inc.

Fujiyama, A. & Tamanoi, F. (1986). Processing and fatty acid acylation of RAS1 and RAS2 proteins in *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the USA* **83** 1266-1270.

Furth, M. E., Aldrich, T. H. & Cordon-Cardo, C. (1987). Expression of *ras* proto-oncogene proteins in normal human tissues. *Oncogene* **1** 47-58.

Gardner, R. L. (ed.) (1983). *Embryonic and Germ Cell Tumours in Man and Animals. Cancer Surveys,* vol. 2, no. 1. Oxford, U. K.: Oxford University Press.

Gearing, D. P. & Bruce, A. G. (1992). Oncostatin-M binds the high-affinity leukaemia inhibitory factor receptor. *New Biologist* **4** 61-65.

George, D. L., Scott, A. F., Trusko, S., Glick, B., Ford, E. & Dorney, D. J. (1985) Structure and expression of amplified cKi-*ras* gene sequences in Y1 mouse adrenal tumor cells. *EMBO Journal* **4** 1199-1203.

Gibbs, J. B., Sigal, I. S., Poe, M. & Scolnick, E. M. (1984). Intrinsic GTPase activity distinguishes normal and oncogenic *ras* p21 molecules. *Proceedings of the National Academy of Sciences of the USA* **81** 5704-5708.

Gibbs, J. B., Marshall, M. S., Scolnick, E. M., Dixon, R. A. & Vogel, U. S. (1990). Modulation of guanine nucleotides bound to *Ras* in NIH3T3 cells by oncogenes, growth factors, and the GTPase activating protein (GAP). *Journal of Biological Chemistry* **265** 20437-20442.

Gillin, F. D., Roufa, D. J., Beaudet, A. L. & Caskey, C. T. (1972) 8-Azaguanine resistance in mammalian cells I. Hypoxanthine-Guanine Phosphoribosyltransferase. *Genetics* **72** 239-252.

Ginsburg, M., Snow, M. L. H. & McLaren, A. (1990). Primordial germ cells in the mouse embryo during gastrulation. *Development* **110** 521-528.

Glomset, J. A. & Farnsworth, C. C. (1994). Role of protein modification reactions in programming interactions between Ras-related GTPases and cell membranes. *Annual Review of Cell Biology* **10** 181-205.

Gorry, P., Lufkin, T., Dierich, A., Rochette-egly, C., Decimo, D., Dolle, P., Mark, M., Durand, B. & Chambon, P. (1994). The cellular retinoic acid binding protein I is dispensable. *Proceedings of the National Academy of Sciences of the USA* **91** 9032-9036.

Gossett, L. A., Zhuang, W. & Olson, E. N. (1988). Dexamethasone-dependent inhibition of differentiation of C2 myoblasts bearing steroid-inducible N-*ras* oncogenes. *Journal of Cell Biology* **106** 2127-2138.

Gossler, A. Doetschman, T., Korn, R., Serfling, E. & Kemler, R. (1986). Transgenesis by means of blastocyst-derived embryonic stem cell lines. *Proceedings of the National Academy of Sciences of the USA* **83** 9065-9069.

Goyette, M. Petropoulos, C. J., Shank, P. R. & Fausto, N. (1983). Expression of a cellular oncogene during liver regeneration. *Science* **219** 510-512.

Green, M. C. (1989). In *Genetic Variants and Strains of the Laboratory Mouse*. Ed. M. F. Lyon & A. G. Searle. 2nd edn. Oxford University Press, London.

Gu, H., Zou, Y.-R. & Rajewsky, K. (1993). Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-*loxP*-mediated gene targeting. *Cell* **73** 1155-1164.

Guerrero, I., Villasante, A., Corces, V. & Pellicer, A. (1984). Activation of a c-K-*ras* oncogene by somatic mutation in mouse lymphomas induced by gamma radiation. *Science* **225** 1159-1162.

Gulbins, E., Coggeshall, M., Baier, G., Katzav, S., Burn, P. & Altman, A. (1993). Tyrosine kinase-stimulated guanine nucleotide exchange activity of Vav in T cell activation. *Science* **260** 822-825.

Hagag, N., Halegoua, S. & Viola, M. (1986). Inhibition of growth factor-induced differentiation of PC12 cells by microinjection of antibody to *ras* p21. *Nature* **319** 680-682.

Hall, A. (1994). A biochemical function for Ras-at last. *Science* **264** 1413-1414.

Han, M. & Sternberg, P. W. (1990). *let-60*, a gene that specifies cell fates during C. elegans vulval induction, encodes as *ras* protein. *Cell* **63** 921-931.

Hancock, J. F., Magee, A. I., Childs, J. E., Marshall, C. J. (1989). All *ras* proteins are polyisoprenylated but only some are palmitoylated. *Cell* **57** 1167-1177.

Hancock, J. F., Paterson, H. & Marshall, C.J. (1990). A polybasic domain or palmitoylation is required in addition to the CAAX motif to localize p21*ras* to the plasma membrane. *Cell* **63** 133-139.

Hancock, J. F., Cadwallader, K. & Marshall, C. J. (1991). Methylation and proteolysis are essential for efficient membrane binding of prenylated $p21^{K\text{-}ras(B)}$. *EMBO Journal* **10** 641-646.

Handyside, A. H., O'Neill, G. T., 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 Archives of Developmental Biology* **198** 48-56.

Hanson, K. D. & Sedivy, J. M. (1995). Analysis of biological selections for high-efficiency gene targeting. *Molecular and Cellular Biology* **15** 45-51.

Hardy, K., Carthew, P., Handyside, A. H. & Hooper, M. L. (1990). Extragonadal teratocarcinoma derived from embryonal stem cells in chimaeric mice. *Journal of Pathology* **160** 71-76.

Harvey, J. J. (1964). An unidentified virus which causes the rapid production of tumours in mice. *Nature* **204** 1104-1105.

Hasty, P., Rivera-Perez, J. & Bradley, A. (1991a). The length of homology required for gene targeting in embryonic stem cells. *Molecular and Cellular Biology* **11** 5586-5591.

Hasty, P., Rivera-Perez, J., Chang, C. & Bradley, A. (1991b). Target frequency and integration pattern for insertion and replacement vectors in embryonic stem cells. *Molecular and Cellular Biology* **11** 4509-4517.

Hasty, P., Ramirez-Solis, R., Krumlauf, R. and Bradley, A. (1991c). Introduction of a subtle mutation into the *Hox-2.6* locus in embryonic stem cells. *Nature* **350** 243-246.

Hasty, P., Rivera-Perez, J. & Bradley, A. (1992). The role and fate of DNA ends for homologous recombination in embryonic stem cells. *Molecular and Cellular Biology* **12** 2464-2474.

Hasty, P. & Bradley, A. (1993). Gene targeting vectors for mammalian cells. In *Gene Targeting-A Practical Approach.* Ed. A. L. Joyner, IRL Press, Oxford, UK.

Hattori, M. & Sakaki, P. (1986). Dideoxy sequencing method using denatured plasmid templates. *Analytical Biochemistry* **152** 232-238.

Hayashi, K. (1992). PCR-SSCP: A method for detection of mutations. *Genetic Analysis-Techniques and Applications* **9** 73-79.

Hoffman, E. K., Trusko, S. P., Freeman, N. & George, D. L. (1987). Structural and functional characterization of the promoter region of the mouse c-Ki-*ras* gene. *Molecular and Cellular Biology* **7** 2592-2596.

Hoffman, E. K., Trusko, S. P., Murphy, M. & George, D. L. (1990). An S1 nuclease-sensitive homopurine/homopyrimidine domain in the c-Ki-*ras* promoter interacts with a nuclear factor. *Proceedings of the National Academy of Sciences of the USA* **87** 2705-2709.

Hogan, B., Costantini, F. & Lacy, E. (1986). *Manipulating the mouse embryo: A laboratory manual.* Cold Spring Harbor, N. Y.: Cold Spring Harbor Laboratory.

Holliday, R. (1964). A mechanism for gene conversion in fungi. *Genetical Research* **5** 282-304.

Hooper, M. L. (1992) *Embryonal Stem Cells: Introducing Planned Changes into the Animal Germline.* Modern Genetics, vol. 1 (Ed. Evans, H. J.). Harwood Academic Publishers, Chur, Switzerland.

Hooper, M. L., 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.

Horie, K. & Shimada, K. (1994). Gene targeting by a vector with hairpin-shaped oligonucleotide caps. *Biochemistry and Molecular Biology International* **32** 1041-1048.

Horie, K., Nishiguchi, S., Maeda, S. & Shimada, K. (1994). Structures of replacement vectors for efficient gene targeting. *Journal of Biochemistry* **115** 477-485.

Ip, N. Y., Nye, S. H., Boulton, T. G., Davis, S., Taga, T., Li, Y. P., Birren, S. J., Yasukawa, K., Kishimoto, T., Anderson, D. J., Stahl, N. & Yancopoulos, G. D. (1992). CNTF and LIF act on neuronal cells via shared signalling pathways that involve the IL-6 signal transducing receptor component gp130. *Cell* **69** 1121-1132.

Ish-Horowicz, D. & Burke, J. F. (1981) Rapid and efficient cosmid cloning. *Nucleic Acids Research* **9** 2989-2998.

Itzhaki, J. E. & Porter, A. C. G. (1991). Targeted disruption of a human interferon-inducible gene detected by secretion of human growth hormone. *Nucleic Acids Research* **19** 3835-3842.

Jaenisch, R. (1988). Transgenic animals. *Science* **240** 1468-1475.

Jackson, J. H., Li, J. W., Buss, J. E., Der, C. J. & Cochrane, C. G. (1994). Polylysine domain of K-ras 4B protein is crucial for malignant transformation.

Jasin, M. & Berg, P. (1988). Homologous integration in mammalian cells without target gene selection. *Genes and Development* **2** 1353-1363.

Jeanotte, L., Ruiz, J. C. & Robertson, E. J. (1991). Low level of Hox 3.1 gene expression does not preclude the use of promoterless vectors to generate a targeted gene disruption. *Molecular and Cellular Biology* **11** 5578-5585.

John, J., Frech, M. & Wittinghofer, A. (1988). Biochemical properties of Ha-ras encoded p21 mutants and mechanism of the autophosphorylation reaction. *Journal of Biological Chemistry* **263** 11792-11799.

John, J., Sohmen, R., Feuerstein, J., Linke, R., Wittinghofer, A. & Goody, R. S. (1990). Kinetics of interaction of nucleotides with nucleotide-free H-ras p21. *Biochemistry* **29** 6058-6065.

Johnson, P. & Friedman, T. (1990). Limited bidirectional activity of 2 housekeeping gene promoters-human HPRT and PGK. *Gene* **88** 207-213.

Johnson, R. S., Sheng, M., Greenberg, M. E., Kolodner, R. D., Papaioannou, V. E. & Spiegelman, B. M. (1989) Targeting of nonexpressed genes in embryonic stem cells via homologous recombination. *Science* **245** 1234-1236.

Jordano, J. & Perucho, M. (1986). Chromatin structure of the promoter region of the human c-K-ras gene. *Nucleic Acids Research* **14** 7361-7378.

Jordano, J. & Perucho, M. (1988). Initial characterization of a potential transcriptional enhancer for the human c-K-*ras* gene. *Oncogene* **2** 359-366.

Joyner, A. L., Skarnes, W. C. & Rossant, J. (1989). Production of a mutation in mouse *En-2* gene by homologous recombination in embryonic stem cells. *Nature* **338** 153-156.

Jung, S., Rajewsky, K. & Radbruch, A. (1993). Shutdown of class switch recombination by deletion of a switch region control element. *Science* **259** 984-987.

Kadonaga, J. T., Jones, K. A. & Tijian, R. (1986). Promoter-specific activation of RNA polymerase II transcription by Sp1. *Trends in Biochemical Sciences* **11** 20-23.

Kamata, T. & Feramisco, J. R. (1984). Epidermal growth factor stimulates the guanine nucleotide binding activity and phosphorylation of *ras* oncogene proteins. *Nature* **310** 147-150.

Katsuki, M., Sato, M., Kimura, M., Yokoyama, M., Kobayashi, K. & Nomura, T. (1988). Conversion of normal behaviour to shiverer by myelin basic protein antisense cDNA in transgenic mice. *Science* **241** 593-595.

Kaufman, M. H. (1992). *The Atlas of Mouse Development.* Academic Press Ltd: London, UK.

Kim, H.-S. & Smithies, O. (1988). Recombinant fragment assay for gene targeting based on the polymerase chain reaction. *Nucleic Acids Research* **16** 8887-8903.

Kim, H.-S., Popovich, B. W., Shehee, W. R., Shesely, E. G.& Smithies, O. (1991). Problems encountered in detecting a targeted gene by the polymerase chain reaction. *Gene.* **103** 227-233.

Kirsten, W. H. & Mayer, L. A. (1967). Morphologic responses to a murine erythroblastosis virus. *Journal of the National Cancer Institute* **39** 311-335.

Konecki, D. S., Brennand, J., Fuscoe, J. C., Caskey, C. T. & Chinault, A. C. (1982). Hypoxanthine-guanine phosphoribosyl transferase genes of mouse and Chinese-hamster - construction and sequence analysis of cDNA recombinants. *Nucleic Acids Research* **10** 6763-6775.

Kotsinas, A., Spandidos, D. A., Romanowski, P. & Wyllie, A. H. (1993). Relative expression of wild-type and activated Ki-*ras*2 oncogene in colorectal carcinomas. *International Journal of Oncology* **3** 841-845.

Kumar, R. & Dunn, L. L. (1989). Designed diagnostic restriction fragment length polymorphisms for the detection of point mutations in *ras* oncogenes. *Oncogene Research* **1** 235-241.

Kumar, S. & Simons, J. P. (1993). The effects of terminal heterologies on gene targeting by insertion vectors in embryonic stem cells. *Nucleic Acids Research* **21** 1541-1548.

Lacal, J. C., Srivastava, S. K., Anderson, P. S. & Aaronson, S. A. (1986). *Ras* p21 proteins with high or low GTPase activity can efficiently transform NIH/3T3 cells. *Cell* **44** 609-617.

Leevers, S. J. & Marshall, C. J. (1992). Activation of extracellular signal-regulated kinase, ERK2, by p21 *ras* oncoprotein. *EMBO Journal* **11** 569-574.

Leon, J., Guerrero, I. & Pellicer, A. (1987). Differential expression of the *ras* gene family in mice. *Molecular and Cellular Biology* **7** 1535-1540.

Liskay, R. M., Letsou, A. & Stachelek, J. L. (1987). Homology requirement for efficient gene conversion between duplicated chromosomal sequences in mammalian cells. *Genetics* **115** 161-167.

Lockett, T. J. & Sleigh, M. J. (1987). Oncogene expression in differentiating F9 mouse embryonal carcinoma cells. *Experimental Cell Research* **173** 370-378.

Lowy, D. R. & Willumsen, B. M. (1993). Function and regulation of ras. *Annual Review of Biochemistry* **62** 851-891.

Magin, T. M., McWhir, J. & Melton, D. W.. (1992). A new mouse embryonic stem cell line with good germ line contribution and gene targeting frequency. *Nucleic Acids Research* **20** 3795-3796.

Manne, V., Bekesi, E. & Kung, H. F. (1985). Ha-*ras* proteins exhibit GTPase activity: point mutations that activate Ha-*ras* gene products result in decreased GTPase activity. *Proceedings of the National Academy of Sciences of the USA* **82** 376-380.

Mansour, S. L., Thomas, K. R. & Capecchi, M. R. (1988). Disruption of the proto-oncogene *int-2* in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. *Nature* **336** 348-352.

Martegani, E., Vanoni, M. Zippel, R., Coccetti, P. ,Brambilla, R., Ferrari, C., Sturani, E. & Alberghina, L. (1992). Cloning by functional complementation of a mouse cDNA encoding a homologue of CDC25, a *Saccharomyces cerevisiae* RAS activator. *EMBO Journal* **11** 2151-2157.

Martin, G. R. & Evans, M. J. (1975). The formation of embryoid bodies *in vitro* by homogeneous embryonal carcinoma cell cultures derived from single cells. *Teratomas and Differentiation*, ed. M. I. Sherman & D. Solter, pp.169-187. New York: Academic Press.

Martin, G. R. (1980). Teratocarcinoma and mammalian embryogenesis. *Science* **209** 768-776.

Martin, G. R. (1981) Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proceedings of the National Academy of Sciences of the USA* **78** 3585-3588.

Matsui, Y., Toksoz, D., Nishikawa, S., Nishikawa, S.-I., Williams, D., Zsebo, K. & Hogan, B. L. M. (1991). Effect of *Steel* factor and leukaemia inhibitory factor on murine primordial germ cells in culture. *Nature* **353** 750-752.

Matsui, Y., Zsebo, K. & Hogan, B. L. M. (1992). Derivation of pluripotent embryonic stem cells from murine primordial germ cells in culture. *Cell* **70** 841-847.

McCarrick, J. W., Parnes, J. R., Seong, R. H., Solter, D. & Knowles, B. B. (1993). Positive-negative selection gene targeting with the diphtheria toxin A-chain gene in mouse embryonic stem cells. *Transgenic Research* **2** 183-190.

McCormick, F. (1993). How receptors turn Ras on. *Nature* **363** 15-16.

McGrath, J. P., Capon, D. J., Smith, D. H., Chen, E. Y., Seeburg, P. H., Goeddel, D. V. & Levinson, A. D. (1983) Structure and organization of the human Ki-*ras* proto-oncogene and a related processed pseudogene. *Nature* **304** 501-506.

McGrath, J. P., Capon, D. J., Goeddel, D. V. & Levinson, A. D. (1984). Comparative biochemical properties of normal and activated human *ras* p21 protein. *Nature* **310** 644-649.

McKusick, V. A. (1978). *Mendelian Inheritance in Man.* 5th edn., John Hopkins Press, Baltimore.

Melton, D. W. (1990). The use of gene targeting to develop animal models for human genetic diseases. *Biochemical Society Transactions* **18** 1035-1039.

Melton, D. W., McEwan, C., McKie, A. B. & Reid, A. M. (1986). Expression of the mouse HPRT gene: deletional analysis of the promoter region of an X-chromosome linked housekeeping gene. *Cell* **44** 319-328.

Melton, D. W., Konecki, D. S., Brennand, J. & Caskey, C. T. (1984). Structure, expression and mutation of the hypoxanthine phosphoribosyltransferase gene. *Proceedings of the National Academy of Sciences of the USA* **81** 2147-2151.

Meselson, M. S. & Radding, C. M. (1975). A general model for genetic recombination. *Proceedings of the National Academy of Sciences of the USA* **72** 358-361.

Miller, C. C. J., McPheat, J. C. & Potts, W. (1992). Targeted integration of the *Ren-1D* locus in mouse embryonic stem cells. *Proceedings of the National Academy of Sciences of the USA* **89** 5020-5024.

Mintz, B. & Fleischman, R. A. (1981). Teratocarcinomas and other neoplasms as developmental defects in gene expression. *Advances in Cancer Research* **34** 211-278.

Moens, C. B., Auerbach, A. B., Conlon, R. A., Joyner, A. L. & Rossant, J. (1992). A targeted mutation reveals a role for N-*myc* in branching morphogenesis in the embryonic mouse lung. *Genes and Development* **6** 691-704.

Mombaerts, P., Clarke, A. R., Hooper, M. L. & Tonegawa, S. (1991). Creation of a large genomic deletion at the T-cell antigen receptor β-subunit locus in mouse embryonic stem cells by gene targeting. *Proceedings of the National Academy of Sciences of the USA* **88** 3084-3087.

Moodie, S. A. & Wolfman, A. (1994). The 3Rs of life: Ras, Raf and growth regulation. *Trends in Genetics* **10** 44-48.

Moreau, J-F., Donaldson, D. D., Bennett, F., Witek-Giannotti, J. A., Clark, S. C. & Wong, G. G. (1988) Leukaemia inhibitory factor is identical to the myeloid growth factor human interleukin for DA cells. *Nature* **336** 690-692.

Mulcahy, L. S., Smith, M. R. & Stacey, D. W. (1985). Requirement for *ras* proto-oncogene function during serum stimulated growth of NIH3T3 cells. *Nature* **313** 241-243.

Muller, R., Slamon, D. J., Tremblay, J. M., Cline, M. J. & Verma, I. M. (1982). Differential expression of cellular oncogenes during pre- and post-natal development in the mouse. *Nature* **299** 640-645.

Muller, R., Slamon, D. J., Adamson, E. D., Tremblay, J. M., Muller, D., Cline, M. J. & Verma, I. M. (1983). Transcription of c-*onc* genes c-*ras*^{Ki} and c-*fms* during mouse development. *Molecular and Cellular Biology* **3** 1062-1069.

Munir, M. I., Rossiter, B. J. F. & Caskey, C. T. (1990). Antisense RNA production in transgenic mice. *Somatic Cell and Molecular Genetics* **16** 383-394.

Neal, S. E., Eccleston, J. F., Hall. A. & Webb, M. R. (1988). Kinetic analysis of the hydrolysis of p21N-Ras - The basal GTPase mechanism. *Journal of Biological Chemistry* **263** 19718-19722.

Newman, C. M. H. & Magee, A. I. (1993). Posttranslational processing of the ras superfamily of small GTP-binding proteins. *Biochimica et Biophysica Acta* **1155** 79-96.

Nickoloff, J. A. & Reynolds, R. J. (1990). Transcription stimulates homologous recombination in mammalian cells. *Molecular and Cellular Biology* **10** 4837-4845.

Nickoloff, J. A. (1992). Transcription enhances intrachromosomal recombination in mammalian cells. *Molecular and Cellular Biology* **12** 5311-5318.

O' Brien, S. J. (1984). In *Genetics Maps*, pp.1-584. Cold Spring Harbor, NY: Cold Spring Harbor Lab.

O'Neal, W. K., Hasty, P., McCray, P. B., Jr., Casey, B., Rivera-Perez, J., Welsh, M. J., Beaudet, A. L. & Bradley, A. (1993). A severe phenotype in mice with a duplication of exon 3 in the cystic fibrosis locus. *Human Molecular Genetics* **2** 1561-1569.

Olson, E. N., Spizz, G. & Tainsky, M. A. (1987). The oncogenic forms of N-*ras* of H-*ras* prevent skeletal myoblast differentiation. *Molecular and Cellular Biology* **7** 2104-2111.

Orr-Weaver, T., Szostak, J. W. & Rothstein, R. J. (1981). Yeast transformation: a model system for the study of recombination. *Proceedings of the National Academy of Sciences of the USA* **78** 6354-6358.

Orr-Weaver, T., Szostak, J. W. (1985). Fungal recombination. *Microbiological Reviews* **49** 33-58.

Pal, S. K., Crowell, R., Kiessling, A. A. & Cooper, G. M. (1993). Expression of proto-oncogenes in mouse eggs and pre-implantation embryos. *Molecular Reproduction and Development* **35** 8-15.

Palmiter, R. D. & Brinster, R. L. (1986). Germline transformation of mice. *Annual Review of Genetics* **20** 465-499.

Papaionnou, V. E. & Rossant, J. (1983). Effects of the embryonic environment on proliferation and differentiation of embryonal carcinoma cells. In *Cancer Surveys*, vol. 2, no. 1, pp. 165-183 (ed. R. L. Gardner). Oxford, U. K.: Oxford University Press.

Porfiri, E., Evans, T., Chardin, P. & Hancock, J. F. (1994). Prenylation of Ras proteins is required for efficient hSOS1-promoted guanine nucleotide exchange. *Journal of Biological Chemistry* **269** 22672-22677.

Quaife, C. J.,Pinkert, C. A., Ornitz, D. M., Palmiter, R. D. & Brinster, R. L. (1987). Pancreatic neoplasia induced by *ras* expression in acinar cells of transgenic mice. *Cell* **48** 1023-1034.

Ramirez-Solis, R., Zheng, H., Whiting, J., Krumlauf, R. & Bradley, A. (1993). *Hoxb-4 (Hox2.6)* Mutant mice show homeotic transformation of a cervical vertebra and defects in the closure of the sternal rudiments. *Cell* **73** 279-294.

Resnick, J. L., Bixler, L. S., Cheng, L. & Donovan, P. J. (1992). Long-term proliferation of mouse primordial germ cells in culture. *Nature* **359** 550-551.

Robertson, E. J. (1987). Embryo-derived stem cell lines. In *Teratocarcinomas and Embryonic Stem Cells-A Practical Approach*, ed. E. J. Robertson, pp. 71-112. Oxford: IRL Press.

Robertson, E. J., Kaufman, M. H., Bradley, A., Evans, M. J. (1983). Isolation, properties and karyotype analysis of pluripotential (EK) cell lines from normal and parthenogenetic embryos. In *Teratocarcinoma Stem Cells: Cold Spring Harbor Conferences on Cell Proliferation*, vol. 10, pp. 647-663. Ed. Silver, L. M.., Martin, G. R. & Strickland, S. Cold Spring Harbor, N. Y.: Cold Spring Harbor Laboratory.

Robertson, E. J., Bradley, A., Keuhn, M. & Evans, M. J. (1986). Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. *Nature* **323** 445-448.

Rothstein, R. (1991). Targeting, disruption, replacement and allele rescue: integrative DNA transformation in yeast. *Methods in Enzymology* **194** 281-301.

Rubinstein, M., Japon, M. A. & Low, M. J. (1993). Introduction of a point mutation into the mouse genome by homologous recombination in embryonic stem cells using a replacement type vector with a selectable marker. *Nucleic Acids Research* **21** 2613-2617.

Rubnitz, J. & Subramani, S. (1984). The minimum amount of homology required for homologous recombination in mammalian cells. *Molecular and Cellular Biology* **4** 2253-2258.

Rudnicki, M. A., Reuhl, K. R. & McBurney, M. W. (1989). A transfected H-*ras* oncogene does not inhibit differentiation of cardiac and skeletal muscle from embryonal carcinoma cells. *Biochemistry and Cell Biology* **67** 590-596.

Rudolph, U., Brabet, P., Kaplan, J. Hasty, P., Bradley, A. & Birnbauer, L. (1993a). Targeting of the $G_{i2}\alpha$ gene in ES cells with replacement and insertion vectors. *Journal of Receptor Research* **13** 619-637.

Rudolph, U. Brabet, P., J. Hasty, P., Bradley, A. & Birnbauer, L. (1993b). Disruption of the $G_{i2\alpha}$ locus in embryonic stem cells and mice: a modified hit and run strategy with detection by a PCR dependent on gap repair. *Transgenic Research* **2** 345-355.

Sachs, A. B. (1993). Messenger RNA degradation in eukaryotes. *Cell* **74** 413-421.

Sakagami, K., Tokinaga, Y., Yoshikura, H. & Kobayashi, I. (1994). Homology-associated nonhomologous recombination in mammalian gene targeting. *Molecular and Cellular Biology* **91** 8527-8531.

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

Sanger, F., Niklen, S. & Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the USA* **74** 5463-5467.

Satoh, T., Nakamura, S. & Kaziro, Y. (1987). Induction of neurite formation in PC12 cells by microinjection of proto-oncogenic Ha-*ras* protein preincubated with guanosine-5'-*O*-(3-thiotriphosphate). *Molecular and Cellular Biology* **7** 4553-4556.

Satoh, T., Endo, M., Nakafaku, M., Akiyama, T., Yamamoto, T. & Kaziro, Y. (1990a). Accumulation of $p21^{ras}$ GTP in response to stimulation with epidermal growth factor and oncogene products with tyrosine kinase activity. *Proceedings of the National Academy of Sciences of the USA* **87** 7926-7929.

Satoh, T., Endo, M., Nakafaku, M.,. Nakamura, S. & Kaziro, Y. (1990b). Platelet-derived growth factor stimulates formation of active $p21^{ras}$ GTP complex in Swiss mouse 3T3 cells. *Proceedings of the National Academy of Sciences of the USA* **87** 5993-5997.

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

Scheerer, J. B. & Adair, G. M. (1994). Homology dependence of targeted recombination at the Chinese hamster *aprt* locus. *Molecular and Cellular Biology* **14** 6663-6673.

Schwartzenberg, P. L., Goff, S. P., Robertson, E. J. (1989). Germ-line transmission of a *c-abl* mutation produced by gene disruption in ES cells. *Science* **246** 799-803.

Scolnick, E. M., Papageorge, A. G. & Shih, T. Y. (1979). Guanine nucleotide-binding activity as an assay for *src* protein of rat-derived sarcoma viruses. *Proceedings of the National Academy of Sciences of the USA* **76** 5355-5359.

Sedivy, J. M. & Sharp, P. A. (1989). Positive genetic selection for gene disruption in mammalian cells by homologous recombination. *Proceedings of the National Academy of Sciences of the USA* **86** 227-231.

Selfridge, J., Pow, A. M., McWhir, J., Magin, T. M. & Melton, D. W. (1992). Gene targeting using a mouse HPRT minigene/HPRT-deficient embryonic stem cell system: Inactivation of the mouse ERCC-1 gene. *Somatic Cell and Molecular Genetics* **18** 325-336.

Serwe, M. & Sablitzky (1993). V(D)J recombination in B cells is impaired but not blocked by targeted deletion of the immunoglobulin heavy chain intron enhancer. *EMBO Journal* **12** 2321-2327.

Sheardown, S. A. & Hooper, M. L. (1992). A relationship between gap junction-mediated intercellular communication and the *in vitro* developmental capacity of murine embryonic stem cells. *Experimental Cell Research* **198** 276-282.

Shehee, W. R., Oliver, P. & Smithies, O. (1993). Lethal thalassemia after insertional disruption of the mouse major β-globin gene. *Proceedings of the National Academy of Sciences of the USA* **90** 3177-3181.

Shih, T. Y., Weeks, M. O., Young, H. A. & Scolnick, E. M. (1979). Identification of a sarcoma virus-coded phosphoprotein in nonproducer cells transformed by Kirsten or Harvey murine sarcoma viruses. *Virology* **96** 64-79.

Shih, T. Y., Papageorge, A. G., Stokes, P. E., Weeks, M. O. & Scolnick, E. M. (1980). Guanine nucleotide-binding and autophosphorylating activities associated with the $p21^{src}$ protein of Harvey murine sarcoma viruses. *Nature* **287** 686-691.

Shih, T. Y., Weeks, M. O., Gruss, P., Dhar, R., Oroszlan, S, & Scolnick, E. M. (1982). Identification of a precursor in the biosynthesis of the p21 transforming protein of Harvey murine sarcoma virus. *Journal of Virology* **42** 253-261.

Shimizu, K., Goldfarb, M., Suard, Y., Perucho, M., Li, Y., Kamata, T. Feramisco, J., Stavnezer, E., Fogh, J. & Wigler, M. H. (1983a). Three human transforming genes are related to the viral *ras* oncogenes. *Proceedings of the National Academy of Sciences of the USA* **80** 2112-2116.

Shimizu, K., Birnbaum, D., Ruley, M. A., Fasano, O., Suard, Y., Edlund, L., Taparowsky, E., Goldfarb, M. & Wigler, M. (1983b) Structure of the Ki-*ras* gene of the human lung carcinoma cell line Calu-1. *Nature* **304** 497-500.

Shirasawa, S., Furuse, M., Yokoyama, N. & Sasazuki, T. (1993) Altered growth of human cancer cell lines disrupted at activated Ki-*ras*. *Science* **260** 85-88.

Shou, C., Farnsworth, C. L., Neel, B. G. & Feig, L. A. (1992). Molecular cloning of cDNAs encoding a guanine-nucleotide releasing factor for Ras p21. *Nature* **358** 351-354.

Shulman, M. J., Nissen, L. & Collins, C. (1990). Homologous recombination in hybridoma cells: dependence on time and fragment length. *Molecular and Cellular Biology* **10** 4466-4472.

Silver, L. M.., Martin, G. R. & Strickland, S. (editors, 1983). *Teratocarcinoma Stem Cells. Cold Spring Harbor Conferences on Cell Proliferation,* vol. 10. Cold Spring Harbor, N. Y.: Cold Spring Harbor Laboratory.

Simon, N. A., Bowtell, D. D. L., Dodson, G. S., Laverty, T. R. & Rubin, G. M. (1991). Ras1 and a putative guanine nucleotide exchange factor perform crucial steps in signalling by the sevenless protein tyrosine kinase. *Cell* **67** 701-716.

Sinn, E., Muller, W., Pattengale, P., Tepler, I., Wallace R. & Leder, P. (1987). Coexpression of MMTV/v-Ha-*ras* and MMTV/c-myc genes in transgenic mice: synergistic action of oncogenes *in vivo*. *Cell* **49** 465-475.

Slamon, D. J. & Cline, M. J. (1984). Expression of cellular oncogenes during embryonic and fetal development of the mouse. *Proceedings of the National Academy of Sciences of the USA* **81** 7141-7145.

Smith, A. G. & Hooper, M. L. (1987) Buffalo Rat Liver cells produce a diffusible activity which inhibits the differentiation of murine embryonal carcinoma and embryonic stem cells. *Developmental Biology* **121** 1-9.

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

Smith, A. J. H. & Kalogerakis, B. (1991). Detection of gene targeting by co-conversion of a single nucleotide change during replacement recombination at the immunoglobulin μ heavy chain locus. *Nucleic Acids Research* **19** 7161-7170.

Smithies, O., Gregg, R. G., Boggs, S. S., Koralewski, M. A. & Kucherlapati, R. S. (1985). Insertion of DNA sequence into the human chromosomal beta-globin locus by homologous recombination. *Nature* **317** 230-234.

Snouwaert, J. N., Brigman, K. K., Latour, A. M., Malouf, N. N., Boucher, R. C., Smithies, O. & Koller, B. H. (1992). An animal-model for cystic-fibrosis made by gene targeting. *Science* **257** 1083-1088.

Solter, D. & Damjanov, I. (1979). Teratocarcinoma and the expression of oncodevelopmental genes. *Methods in Cancer Research, vol. 18*, pp. 277-332. Ed. W. H. Fischman & H. Busch. New York: Academic Press.

Sorrentino, V., McKinney, M. D., Giorgi, M., Geremia, R. & Fleissner, E. (1988). Expression of cellular proto-oncogenes in the mouse male germ line: a distinctive 2.4-kilobase *pim*-1 transcript is expressed in haploid postmeiotic cells. *Proceedings of the National Academy of Sciences of the USA* **85** 2191-2195.

Southern, E. M. (1975). Detection of specific sequences among DNA fragments separated by separated by gel electrophoresis. *Journal of Molecular Biology* **98** 503.

Southern, P. J. & Berg, P. (1982).Transformation of mammalian cells to antibiotic resistance with a bacterial gene under control of the SV40 early region promoter. *Journal of Molecular and Applied Genetics* **1** 327-341.

Spandidos, D. A. & Dimitrov, T. (1985). High expression of ras p21 protein in normal mouse heart tissues. *Bioscience Reports* **5** 1035-1039.

Stacey, A., Schnieke, A., McWhir, J., Cooper, J., Colman, A. & Melton, D. W. (1994). Use of double-replacement gene targeting to replace the murine α-lactalbumin gene with its human counterpart in embryonic stem cells and mice. *Molecular and Cellular Biology* **14** 1009-1016.

Stacey, D. W. & Kung, H. F. (1984). Transformation of NIH3T3 cells by microinjection of Ha-*ras* p21 protein. *Nature* **310** 508-511.

Stachelek, J. L. & Liskay, R. M. (1988). Accuracy of intrachromosomal gene conversion in mouse cells. *Nucleic Acids Research* **16** 4069-4076.

Stanton, B. R., Perkins, A. S., Tessarollo, L., Sassoon, D. A. & Parada, L. F. (1992). Loss of N-*myc* function results in embryonic lethality and failure of the epithelial component of the embryo to develop. *Genes and Development* **6** 2235-2247.

Steeg, C. M., Ellis, J. & Bernstein, A. (1990). Introduction of specific point mutations into RNA polymerase II by gene targeting in mouse embryonic stem cells: Evidence for a DNA mismatch repair system. *Proceedings of the National Academy of Sciences of the USA* **87** 4680-4684.

Stevens, L. C. (1983). The origin and development of testicular, ovarian and embryo-derived teratomas. In *Teratocarcinoma Stem Cells: Cold Spring Harbor Conferences on Cell Proliferation*, vol. 10, pp. 23-36. Ed. Silver, L. M.., Martin, G. R. & Strickland, S. Cold Spring Harbor, N. Y.: Cold Spring Harbor Laboratory.

Stewart, C. L., Gadi, I. & Batt, H. (1994). Stem cells from primordial germ cells can re-enter the germline. *Developmental Biology* **161** 626-628.

Stewart, T. A. & Mintz, B. (1981). Successive generations of mice produced from an established culture line of euploid teratocarcinoma cells. *Proceedings of the National Academy of Sciences of the USA* **78** 6314-6318.

Stewart, T. A. & Mintz, B. (1982). Recurrent germline transmission of the teratocarcinoma genome from the METT-1 culture line to progeny *in vivo*. *Journal of Experimental Zoology* **224** 465-469.

Stinnakre, M. G., Evans, M. J., Willison, K. R. & Stern, P. L. (1981). Expression of Forssman antigen in the post-implantation mouse embryo. *Journal of Embryology and Experimental Morphology* **61** 117-131.

Stout, J. T. & Caskey, C. T. (1985). HPRT: Gene structure, expression, and mutation. *Annual Review of Genetics* **19** 127-148.

Stutts, P. & Brockman, R. W. (1963). A biochemical basis for resistance of L1210 mouse leukaemia to 6-thioguanine. *Biochemical Pharmacology*. **12** 97-104.

Sweet, R. W., Yokoyama, S., Kamata, T. Feramisco, J. R., Rosenberg, M. & Gross, M. (1984). The product of *ras* is a GTPase and the T24 oncogenic mutant is deficient in this activity. *Nature* **311** 273-275.

Szeberenyi, J., Cai, H. & Cooper, G. M. (1990). Effect of a dominant inhibitory Ha-*ras* mutation on neuronal differentiation of PC12 cells. *Molecular and Cellular Biology* **10** 5324-5332.

Szybalski, W. & Szybalska, E. H. (1962). Drug sensitivity as a genetic marker for human cell lines. *University of Michigan Medical Bulletin* **28** 277-293.

Tamanoi, F., Walsh, M., Kataoka, T. & Wigler, M. (1984). A product of the yeast *Ras2* gene is a guanine nucleotide binding protein. *Proceedings of the National Academy of Sciences of the USA* **81** 6924-6928.

Tanaka, T., Ida, N., Shimoda, H., Waki, C., Slamon, D. J. & Cline, M. J. (1986). Organ specific expression of ras oncoproteins during growth and development of the rat. *Molecular and Cellular Biochemistry* **70** 97-104.

Tatchell, K., Chaleff, D. T., DeFeo-Jones, D. & Scolnick, E. M. (1984). Requirement of either of a pair of *ras*-related genes of *Saccharomyces cerevisiae* for spore viability. *Nature* **309** 523-527.

Te Riele, H., Robanus Maandag, E. & Berns, A. (1992). Highly efficient gene targeting in embryonic stem cells via homologous recombination with isogenic DNA constructs. *Proceedings of the National Academy of Sciences of the USA* **89** 5128-5132.

Temeles, G. L., Gibbs, J. B., D'Alonzo, J. S., Sigal, I. S. & Scolnick, E. M. (1985). Yeast and mammalian *ras* proteins have conserved biochemical properties. *Nature* **313** 700-703.

Thein, S. L. & Wallace, R.B. (1986) The use of synthetic oligonucleotides as specific hybridisation probes in the diagnosis of genetic disorders. *Human Genetic Diseases: A Practical Approach*, pp. 30-50. Ed. K. E. Davis. IRL Press, Herndon, Virginia.

Thomas, K. R. & Capecchi, M. R. (1986). Introduction of homologous DNA-sequences into mammalian cells induces mutations in the cognate gene. *Nature* **324** 34-38.

Thomas, K. R. & Capecchi, M. R. (1987). Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* **51** 503-512.

Thomas, K. R., Folger, K. R. & Capecchi, M. R. (1986). High frequency targeting of genes to specific sites in the mammalian genome. *Cell* **44** 419-428.

Thomas, K. R., Deng, D. & Capecchi, M. R. (1992). High frequency gene targeting in embryonic stem cells by using sequence replacement vectors. *Molecular and Cellular Biology* **12** 2919-2923.

Thompson, S., Clarke, A. R., Pow, A.M., Hooper, M. L. & 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.

Trahey, M. & McCormick, F. (1987). A cytoplasmic protein stimulates normal N-ras p21 GTPase, but does not affect oncogenic mutants. *Science* **238** 542-545.

Tybulewicz, V. L. J., Crawford, C. E., Jackson, P. K., Bronson, R. T. & Mulligan, R. C. (1991). Neonatal lethality and lymphopenia in mice with a homozygous disruption of the *c-abl* proto-oncogene. *Cell* **65** 1153-1163.

Ulsh, L. & Shih, T. Y. (1984). Metabolic turnover of human c-*ras*H p21 protein of EJ bladder carcinoma and its normal cellular and viral homologues. *Molecular and Cellular Biology* **4** 1647-1652.

Umanoff, H., Edelmann, W., Pellicer, A. & Kucherlapati, R. (1995). The murine N-*ras* gene is not essential for growth and development. *Proceedings of the National Academy of Sciences of the USA* **92** 1709-1713.

Valencia, A., Chardin, P., Wittinghofer, A. & Sander, C. (1991). The *ras* protein family: Evolutionary tree and role of conserved amino acids. *Biochemistry* **30** 4637-4648.

Valencius, V. and Smithies, O. (1991a). Double-strand gap repair in a mammalian gene targeting reaction. *Molecular and Cellular Biology* **11** 4389-4397.

Valencius, V. and Smithies, O. (1991b). Testing an "In-Out" targeting procedure for making subtle genomic modifications in mouse embryonic stem cells. *Molecular and Cellular Biology*. **11** 1402-1408.

Van Deursen, J. & Wieringa, B. (1992). Targeting of the creatine kinase M gene in embryonic stem cells using isogenic and nonisogenic vectors. *Nucleic Acids Research* **20** 3815-3820.

Voet, D. & Voet, J. G. (1990). Nucleotide metabolism. In *Biochemistry*, chapter 26, pub. John Wiley & Sons, Inc., New York.

Wake, C. T., Vernaleone, F. & Wilson, J. H. (1985). Topological requirements for homologous recombination among DNA molecules transfected into mammalian cells. *Molecular and Cellular Biology* **5** 2080-2089.

Waldman, A. S. & Liskay, R. M. (1987). Differential effects of base-pair mismatch on intrachromosomal versus extrachromosomal recombination in mouse cells. *Proceedings of the National Academy of Sciences of the USA* **84** 5340-5344.

Waldman, A. S. & Liskay, R. M. (1988). Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology. *Molecular and Cellular Biology* **8** 5350-5357.

Wei, W., Mosteller, R. D., Sanyal, P., Gonzales, E., McKinney, D., Dasgupta, C., Li, P., Liu, B., X. & Broek, D. (1992). Identification of a mammalian gene structurally and functionally related to the CDC25 gene of *Saccharomyces cerevisiae*. *Proceedings of the National Academy of Sciences of the USA* **89** 7100-7104.

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. & Gough, N. M. (1988) Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336** 684-688.

Willingham, M. C., Pastan, I., Shih, T. Y. & Scolnick, E. M. (1980). Location of the *src* gene product of the Harvey strain of MSV to plasma membrane of transformed cells by electron microscopic immunocytochemistry. *Cell* **19** 1005-1014.

Willumsen, B. M., Chritsensen, A., Hubbert, N. L., Papgeorge, A. G. & Lowy, D. R. (1984a). The p21 *ras* c-terminus is required for transformation and membrane association. *Nature* **310** 583-586.

Willumsen, B. M., Norris, K., Papageorge, A. G. Hubbert, N. L., & Lowy, D. R. (1984b). Harvey murine sarcoma virus p21 *ras* protein: biological and biochemical significance of the cysteine nearest the carboxy terminus. *EMBO Journal* **3** 2581-2585.

Wolf, E., Kramer, R., Polejaeva, I., Thoenen, H. & Brem, G. (1994). Efficient generation of chimeric mice using embryonic stem-cells after long-term culture in the presence of ciliary neurotrophic factor. *Transgenic Research* **3** 152-158.

Wong, E. A. & Capecchi, M. R. (1986). Analysis of homologous recombination in cultured mammalian cells in transient expression and stable transformation assays. *Somatic Cell and Molecular Genetics* **12** 63-72.

Wong, E. A. & Capecchi, M. R. (1987). Homologous recombination between coinjected DNA sequences peaks in early to mid-S phase. *Molecular and Cellular Biology* **7** 2294-2295.

Wu, H., Liu, X. & Jaenisch, R. (1994). Double replacement: strategy for efficient introduction of subtle mutations into the murine *Colla-1* gene by homologous recombination in embryonic stem cells. *Proceedings of the National Academy of Sciences of the USA* **91** 2819-2823.

Xu, G. F., O'Connell, P., Viskochil, D., Cawthon, R., Robertson, M., Culver, M., Dunn, D., Stevens, J., Gesteland, M., White, R. & Weiss, R. (1990). The neurofibromatosis type-1 gene encodes a protein related to GAP. *Cell* **62** 599-608.

Yamamoto, F. & Perucho, M. (1988). Characterization of the human c-K-ras promoter. *Oncogene Research* **3** 125-138.

Yamamoto, T., Kaibuchi, K., Mizuno, T., Hiroyoshi, M., Shirataki, H. & Takai, Y. (1990). Purification and characterization from bovine brain cytosol of proteins that regulate the GDP/GTP exchange reaction of *smg* p21s, *ras* p21-like GTP binding proteins. *Journal of Biological Chemistry* **265** 16626-16634.

Yamouchi, N., Kiessling, A. A. & Cooper, G. M. (1994). The ras/raf signalling pathway is required for progression of mouse embryos through the two-cell stage. *Molecular and Cellular Biology* **14** 6655-6662.

Yanisch-Perron, C., Vieira, J. & Messing, J. (1985). Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33** 103-119.

Zhang, H., Hasty, P. & Bradley, A. (1994). Targeting frequency for deletion vectors in embryonic stem cells. *Molecular and Cellular Biology* **14** 2404-2410.

Zheng, H. & Wilson, J. H. (1990). Gene targeting in normal and amplified cell lines. *Nature* **344** 170-173.

Zheng, H., Hasty, P., Brenneman, M. A., Grompe, M., Gibbs, R. A., Wilson, J. H. & Bradley, J. H. (1991). Fidelity of targeted recombination in human fibroblasts and murine embryonic stem cells. *Proceedings of the National Academy of Sciences of the USA* **88** 8067-8071.

Zijlstra, M., Li, E., Sajjadi, F., Subramani, S. & Jaenisch, R. (1989). Germline transmission of a disrupted β_2-microglobulin gene produced by homologous recombination in embryonic stem cells. *Nature* **342** 435-438.

Zimmer, A. & Gruss, P. (1989). Production of chimaeric mice containing embryonic stem (ES) cells carrying a homeobox *Hox 1.1* allele mutated by homologous recombination. *Nature* **338** 150-153.