GENE TRAP IDENTIFICATION OF *mbrg-1*: A NOVEL MAMMALIAN GENE ENCODING A BROMODOMAIN CONTAINING PROTEIN

by

Wuhua Jing

A thesis submitted to the

Faculty of Graduate Studies and Research

McGill University

In Partial Fulfillment of the Requirement for the Degree

of

Doctor of Philosophy

Department of Medicine

Division of Experimental Medicine

McGill University, Montreal, Quebec, Canada

June, 1995

Dedicated to Mom and Dad
for your selfless and unconditional love

獻給媽媽爸爸，感謝您們無私、無條件的愛

Abstract

A gene trap approach was used to screen for genes that, on the basis of their temporal and spatial expression patterns, may play important roles in nervous system development. In this strategy genes expressed in embryonic stem (ES) cells are tagged by insertion of a reporter construct. In one such clone the trapped gene, *mbrg-1*, was expressed throughout the developing and mature nervous system. In chimeras the expression pattern of the trapped gene revealed a spatially restricted pattern of cell mixing and migration in the cerebral cortex and spinal cord. Fusion protein encoded by this gene trapped locus was localized to the nucleus, suggesting that the endogenous gene may play a role in the function or organization of neuronal nuclei. The exon sequences of the endogenous gene present in the fusion transcript were cloned from the fusion transcript by 5' RACE, and overlapping cDNA clones were subsequently obtained from mouse embryonic (E12.5) and brain (P8) cDNA libraries. One cDNA clone of 1574 bp encodes a putative phosphorylated protein of 349 amino acids. In this sequence an 85 amino acid bromodomain, conserved across multiple species was identified. Most proteins with this motif have been shown to be either an essential part of transcription complexes or function as transcription activation adaptors. The transcript encoding the bromodomain containing protein was expressed ubiquitously, while only the alternatively spliced isoform disrupted by the gene trap insertion was abundant in differentiated neurons. Thus, this investigation also revealed the existence of unique splicing machinery apparently shared by all neurons. This novel murine bromodomain encoding gene is conserved among other mammals, and, by fluorescent in situ hybridization analysis was localized on mouse chromosome 2.

Résumé

La technique de "gene trap" a été utilisé pour selectionner des gènes qui, en fonction de leur expression spatio-temporale, pourraient jouer un rôle important dans le developpement du système nerveux.

Grâce à l'insertion du gène reporter LacZ, cette approche permet de marquer des gènes qui sont exprimés dans les cellules souches embryonnaires (ES). Les chimères générées à partir d'un clone LacZ positif ont présenté la particularité d' exprimer ce gène (*mborg-1*) dans le système nerveux embryonnaire et adulte. L'analyse histologique de cerveaux de ces chimères a permis de mettre en évidence que l'expression du gène ciblé était restreinte à des régions specifique du cortex cerebral et de la moelle épiniere.

La protéine de fusion codée par ce gène cible a été localisé dans le noyau, suggérant que le gène endogène pourrait être impliqué dans la fonction ou l'organisation du noyau neuronale. Les séquences d'exons du gène endogène ont été clonés à partir du transcrit de fusion par "5' RACE", par la suite des clones d' ADNc ayant des séquences homologues avec le gène endogène ont été obtenus à partir de banque d' ADNc d'embyon de souris (E12.5) et de cerveau (P8). Un de ces clones ADNc code pour une protéine phosphorylée putative de 349 acides aminés. L'analyse de cette dernière séquence a permis d'identifier une séquence bromodomaine de 85 acides aminés, bien conservée chez plusieurs espèces. Or, il a été bien établi que la plupart des protéines contenant une telle séquence representent un élément essentiel dans le complexe de transcription ou bien fonctionnent comme des adaptateurs activant la transcription. Dans le présent travail, on montre que le transcrit codant pour la protéine bromodomaine s'exprime de façon ubiquitaire, alors qu'une forme tronquée, interrompue par l'insertion de la construction LacZ, est abondante dans les neurones. De plus, cette étude met en évidence la présence d'un mécanisme d'épissage unique qui serait commun à tous les neurones. Enfin, ce gène codant pour un nouveau bromodomaine murin est conservé chez d'autre mammifères et l'analyse par hybridation fluorescente *in situ* a permis de localiser le gène *mborg-1* sur le chromosome 2.

Table of Contents

Index of Tables

Index of Figures

List of Abbreviations

ATP	adenosine 5' triphosphate
Bluo-gal	halogenated indolyl β-galactoside
bp	base pair
BSA	bovine serum albumin
DEPC	diethyl pyrocarbonate
ddH_2O	double distilled water
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EDTA	disodium ethylenediaminetetra-acetate
ES cells	embryonic stem cells
g	gram
GTP	guanidine 5' triphosphate
HCL	hydrochloric acid
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethansulfonic Acid
kb	thousand base pairs
M	molar
mA	milliamps
mg	milligram
MgCl	magnesium chloride
ml	milliliter
mM	millimolar
$MgSO_4$	magnesium sulfate
mRNA	messenger ribonucleic acid
NaCl	sodium chloride
NaOAC	sodium acetate
NaOH	sodium hydroxide
O.D.	optical density

PBS	phosphate buffered saline
RNA	ribonucleic acid
SDS	sodium dodecyl sulfate
SSC	sodium chloride/sodium citrate
ssDNA	salmon sperm DNA
µF	microfarad
µg	microgram
µl	microliter
wt	weight

Acknowledgments

An endeavor such as a Ph.D. is impossible to accomplish without the generous help and support of friends and colleagues. I would like to take this opportunity to thank those whom I was fortunate to know, work with and form friendships with over the years.

I would like to express my deep gratitude to Dr. Alan Peterson, my thesis supervisor, and also one of the most patient men that I have known. I thank him for having given me the opportunity to pursue my graduate studies in his laboratory, for his advice and encouragement, especially during the difficult times, and for his patience which I have perhaps abused to the ultimate limit. I especially appreciated the freedom he allowed which permitted me to develop scientific thoughts and experimental skills independently. I am particularly grateful for his unwavering support during the entire course of my graduate training and in the pursuit of my career. I thank him for teaching me that science is not a matter of consensus, and I leave his lab a much better and confident scientist.

I would like to thank the members of my thesis committee, Dr. Hans Zingg, Dr. Ken Hastings, Dr. Sam David and Dr. Bob Levine for their constructive criticism of my thesis project, and for their kind and generous support in the pursuit of my career.

Over the years I was extremely fortunate to work with a group of wonderful friends and colleagues in Dr. Alan Peterson's lab. I wish to thank them all for making the time I spent in the lab so enjoyful, and for bearing with me all the troubles and headaches that I have caused each one of them. I will miss them all!

I am indebted to Dr. Jack Jenuth, one of the kindest and most generous men I have ever worked with. I thank him for his friendship, and for his patience in enduring my fluctuating and often unbearable moods. He never said "no" when I asked for help, either with molecular biology techniques or with anything remotely related to the word "computer". Everyone has learned to turn to him when one feels totally abandoned by cyberspace. I am grateful for his careful proof reading of my thesis, and for saving my skin by sorting out the simplest and fastest way to organize references for this thesis. I also thank him and his wife, Karen, for their gracious hospitality and their kind support during the roughest times.

I have known Dr. Lorella Garofalo since my days in the department of Pharmacology. Her elegance and maturity impressed a then young and relatively "innocent" mind as being a model Canadian. She has set a standard

for many of us who strive to have a character like hers. So many times I gained my strength or composure after talking to Lorella. She has been a great source of moral support and a good friend that one can rely on. I thank her for her friendship, for her kindness and generosity, for her careful proof-reading of my thesis, and for her help with the French translation of the abstract.

I owe an enormous debt to Dr. Naima Buchnou for her generous help in constructing the day 12.5 chimeras during the last phase of my studies. I can not imagine anybody else who could smile at my often ridiculous requests and at times the impossible short notices, and be willing to offer their generous assistance. I greatly appreciated her kindness, determination and assertiveness. I also thank her for taking the time to initiate the French translation of the thesis abstract one day prior to catching a plane to Morocco.

Priscilla Valera is the one person I could never thank enough. She is the sweetheart and backbone of the lab, and I suspect that I have added to her migraines over the years. This thesis would not have been possible without her generous help and excellent technical assistance. It was a great pleasure to know and work with her. I thank her for her friendship and support during the time I spent in the lab. I know that I will always have someone to talk to when the going gets rough. I am most grateful to her for making time for me during her very busy schedule, and offering her generous help in histochemistry and immunocytochemistry.

My deepest thanks are extended to Irene Tretjakoff for introducing me to embryonic manipulation techniques. It was thanks to her that I saw for the first time the textbook maneuver of fertilized egg microinjection and embryo transplantation. I thank her for her flawless technical assistance in transplanting the ES cell-injected blastocysts, and for doing a superb job, along with Sue Albrechtson, Julie Tremblay, Nancy Woolmer and Jeanne Vestergaard in maintaining one of the finest animal facilities one can ever have.

I would not have survived the toughest times of my Ph.D. training without the help of a good friend, Dr. Ming Ru. She accepted a tired and worn-out student as her friend, and gave me strength when I most needed it. She walked me through the most frustrating part of molecular cloning, and never once hesitated to come to my rescue when nothing seemed to make sense to me. I will never forget her helping hands during the most difficult times. Thank you, Ming!

I wish to thank Drs. X. Zhang and P. Eydoux, The Montreal Children's Hospital, for identifying the chromosomal localization of *mborg-1* using FISH analysis.

I wish to thank Dr. Morag Park for always keeping her door open for discussions, and for her many valuable suggestions in molecular biology. I also would like to thank the gang in Morag's lab, especially Monica Naujokas, Becky Fixman, Xiuming Yang and Gerry Rodrigues for never letting me feel as a stranger in their lab.

I would like to reserve a special thanks for my best friend, Dr. Hiba El-Bizri, a kind, generous and gentle soul. Her incredible zest for life and success is an inspiration to all who know her. I thank her for seeing me through the good times and bad times throughout these years, for always being there when I needed someone, and most of all, for making me feel like a member of her family.

It was a great pleasure for me to know Dr. Soren Singel, in whom I see a rare combination of a balanced personality and the brilliance of a neurosurgeon. He has been a great friend that I can count on. I thank him for many of the joyful moments and interesting philosophical discussions we shared, and for doing almost an impossible job to relocate me along with the absurd amount of my collections down south of the border.

My special thanks to a good friend, Pierre Thèberge, for helping me to come to Canada and thus for making all this possible.

I would like to reserve a special line for Rosie for just being so wonderful. I thank her for making the last few weeks of my stay in Montreal most memorable and joyful, and for always making me feel like being home.

Peiqun has not only been my little brother, but also my closest friend and confidant. Over the years he has shared my happiness and sadness, my joy and frustration. I thank him for his fraternal love and care, and for making me feel that my family is never far.

Finally, to my parents, there are no words that can adequately express my gratitude for their unconditional love and support. I know that they have waited long for this labor to end, but they never suggested that I ought to do otherwise. I am blessed to have the most loving, caring and selfless Mom and Dad, and it is to them that I humbly dedicate this thesis.

This work has been supported by the Ludwig Institute for Cancer Research, Royal Victoria Research Institute Studentship, and a Studentship from The Canadian Network of Center of Excellence (NCE) for Neural Regeneration and Functional Recovery.

Introduction

1.1 Mammalian developmental gene identification

Mammalian development is a complex process involving spatial and temporal regulation of numerous gene products and complex cell-cell interactions. Previous genetic approaches used to further the understanding of mammalian development have largely been based on classical mutational analysis and molecular genetic analysis (Rossant and Joyner, 1989). In particular, efforts have been made to identify genes that play important roles in mammalian development, where many of the critical genes are yet to be found.

1.1.1 The classical genetic approach

Classical mutational analysis, in which spontaneous or experimentally induced mutations are phenotypically examined and the genetic alterations that cause the developmental defect are subsequently characterized, has been an extremely successful tool in studying invertebrate development. Notably for *Caenorhabditis elegans* and *Drosophila melanogaster*, the availability of numerous natural strains carrying chromosomal deletions, the ability to breed large numbers, the relatively small and accessible genome, and the fact that embryos develop *ex utero*, have made large-scale screening for relevant mutations a powerful means to identify genes that play a significant developmental role. A large number of developmentally important genes have been isolated on the basis of their mutational phenotypes in this manner. The most ideal mammalian candidate amenable to developmental analysis, the mouse, whose early embryology is fairly well understood (Rossant, 1986), has approximately twenty times the genome size of fruit flies,

a much longer gestation period, delivers fewer offspring, and its embryos develop *in utero*. Thus, initiating large-scale screening for mutations affecting embryonic development in the mouse is neither economically nor experimentally practical.

Investigations of some spontaneous or experimentally induced mouse mutants known to interfere with normal murine development has led to an understanding of the molecular basis of these mutations (Magnuson, 1986). Recent developments in gene cloning and mapping techniques have facilitated the identification of candidate genes for existing mouse mutants (Table 1). Many of these mouse mutants have served as models for human genetic diseases. Correlation of the phenotypic properties of the mutant locus and the known biochemical properties and/or expression pattern of a closely linked gene provides a basis to further investigate the relationship between the mutant locus and the candidate gene. The best examples, whereby mutant mouse genes were cloned based on mutational phenotypes are: *W*, *un*, *Splotch* (Sp^{2H}), *Sxr*, and *T* mutations. The molecular lesions have been localized within the *c-kit* (Geissler et al., 1988b), *Pax 1* (Balling et al., 1988), *Pax 3* (Epstein et al., 1991), *Sry* (Gubbay et al., 1990), and *Brachyury* genes (Herrmann et al., 1990), respectively. Molecular characterization of mutant genes is generally carried out either by a deterministic approach from the phenotype to the genotype via genetic mapping and positional cloning (Herrmann et al., 1990; Gubbay et al., 1990), or by molecular characterization of previously identified genes mapped in close proximity or allelic with the mouse mutants (Balling et al., 1988)

1.1.1.1 Positional cloning and genetic mapping

One of the most important tools in molecular cloning is the availability of genetic markers or cloning tags required to identify homologous sequences in or near the genomic loci. In the absence of these markers, molecular characterization can only be accomplished by genetic mapping and chromosome walking. Despite certain successes, difficulties in isolating genomic probes close to the mutant loci and unpredictable genomic rearrangements often present in mutant alleles have presented daunting challenges to this approach. In addition, the labor required and resources needed to carry out physical mapping and positional cloning also create barriers.

In a classical deterministic approach, from the phenotype to the gene, Herrmann and co-workers (1990) used a combination of molecular and genetic techniques to clone the gene for the mouse *T* mutation. Genomic probes isolated by microdissected spread metaphase chromosomes or by chromosome walking and jumping, were used to map the *T* locus through analysis of several alleles of the *T* mutation. Genomic and cDNA clones containing the *T* gene were then identified. Different mutant *T* alleles were found to be generated by either deletion or insertional mutation affecting functional splicing. The *T* gene was found to encode a novel protein which is highly conserved among vertebrates and is required during mesoderm formation and notochord morphogenesis. These processes are defective in the T mutation.

In the sex-reversed mutant, *Sxr*, a small part of the short arm of the Y chromosome is translocated onto the pseudoautosomal region at the distal end of the long arm of the Y chromosome (Vergnaud et al., 1986). This part of the short arm was found to be sufficient to confer maleness and can be transferred to the X chromosome during male meiosis by crossover of the

pseudoautosomal region shared by the X and Y chromosomes (Goodfellow et al., 1986). The resulting *Sxr* XX male offspring contains the *Sry* gene (a gene from the sex-determining region of the Y chromosome), within the translocated short arm chromosome. A mutated *Sry* gene has been seen in the XY female. Gubbay and co-workers (1990) isolated a 3.5 kb genomic fragment of the *Sry* gene by size-selected cloning using a human Y choromosome-specific probe (Sinclair et al., 1990). The mouse *Sry* gene was shown to share sequence homology with the human *SRY* gene. It was expressed at a developmental stage suggesting that it plays a role in Sertoli cell fate determination during male gonadal development. It was found to be a member of a family of at least five mouse genes, all of which encode putative DNA binding proteins.

1.1.1.2 Association of cloned genes with mutant loci

Most mouse mutants have been characterized by direct association of mutant loci with previously cloned genes. Advances in molecular cloning and mapping of genes have provided important genetic markers to access the mutant loci and target the molecular defects of mutant alleles.

The murine *dominant-white spotting* (*W*) mutation has pleiotropic effects on embryonic development of melanoblasts, primordial germ cells and hematopoietic stem cells. Homozygous *W* mutant mice are sterile, have severe macrocytic anemia and coat color abnormalities with extensive white-spotting (Geissler et al., 1981). Both the *W* locus (Geissler et al., 1988a) and the *c-kit* proto-oncogene (Yarden et al., 1987), a transmembrane tyrosine kinase receptor (RTK), map in close proximity on mouse chromosome 5. Geissler and coworkers (1988b) localized mutations in two spontaneous mutant *W* alleles, W^{44} and W^{x}, in regions of *c-kit* cDNA encoding amino acids 240 to

342 and 342 to 791, respectively. These results suggest that *c-kit* is the gene product of the *W* locus, consistent with the possibility that c-kit may play an important role in mediating signal transduction pathways in the development of hematopoietic, melanogenic and germ cell lineages. The *steel* (Sl) locus on mouse chromosome 10 shares a similar mutant phenotype with the *W* locus and its gene product has been identified as the ligand for *c-kit* RTK (Williams et al., 1990; Copeland et al., 1990; Flanagan & Leder., 1990).

The mouse *Pax 1* gene, a member of the mouse *Pax* gene family, was originally isolated on the basis of its sequence homology to the *Drosophila* genes containing the paired box (Deutsch et al., 1988). Using restriction fragment length polymorphism (RFLP) in recombinant inbred strains, Balling and co-workers (1988) mapped the mouse *Pax 1* gene close to a recessive mouse mutation, *undulated (un)* on chromosome 2. *Pax 1* was exclusively expressed along the entire sternum and vertebral column (Deutsch et al., 1988), structures distorted in mouse *un* homozygotes. The molecular defect of the *un* mutant was identified as a point mutation by the polymerase chain reaction (PCR) and sequencing, leading to a Gly-Ser replacement in a highly conserved part of the pair-box of the *Pax 1* gene (Balling et al. 1988). This single amino acid exchange dramatically decreases the DNA-binding affinity of the *un Pax 1* protein and alters its DNA binding specificity (Chalepakis et al., 1991). A point mutation was also found in the dystrophin gene in the mouse *mdx* mutant, an animal model of X-linked human Duchenne muscular dystrophy. The single base pair substitution within an exon causes premature termination of the polypeptide chain encoded by the dystrophin gene and renders the myopathic phenotype in the *mdx* mutant (Sicinski et al., 1989). Another mouse *Pax* gene, *Pax 3*, was identified as a candidate gene for the mouse mutant *splotch* (*Sp*), which exhibits spina bifida and encephaly. A

32 nucleotide deletion was found within the conserved paired homeodomain of the *Pax 3* gene in mouse mutant *splotch* (*Sp*) and the deletion generates a truncated protein as a result of a newly created termination codon Epstein et al., 1991).

These studies demonstrate that murine genes only known by the phenotype of a mutation are accessible to cloning. The list of candidate genes identified for mouse mutants continues to grow (Table 1). However, it is also obvious from these investigations that cloning genes responsible for the mouse mutant phenotypes by conventional genetic approaches remains a difficult and tedious task. Although each newly cloned gene can be mapped and tested for the availability of a mutant, the matching of the cloned genes and the existing mutants is likely to be fortuitous. Only a very small proportion of available mouse mutations have been characterized at the molecular level. It is also very difficult to determine the primary mutant defect by phenotypic examination alone. Most pre-existing mutants therefore have been identified when cloned genes were found to be allelic with developmental mutants. In addition, the number of spontaneous or radiation-induced mouse mutants affecting developmental processes (Lyon, 1989) is limited. The vast majority of the estimated 100,000 genes present in the mouse genome have not been identified. Nevertheless, the classical genetic approach from the phenotype to the gene has the obvious advantage in that phenotypic analysis of the mutant loci would yield clues to the biological properties of the cloned gene product. However, the development of gene targeting by homologous recombination has offered a direct means of generating null mutations by interrupting endogenous gene expression at will (Capecchi, 1989). Thus, genes isolated by various methods are amenable

Table 1 Mouse mutants and their characterized molecular defects

Mutant locus	Gene	Chromosome location	References
Shiverer (shi)	*MBP*	18	Roach et al., 1985 Molineaus et al., 1986
albino (c)	*tyrosinase*	7	Kwon et al., (1987)
Jimpy (jp)	*PLP*	X	Nave et al., 1987
sparse fur (spf)	*ornithine transcarbamylase*	X	Veres et al., (1987)
Whie spoting (W)	*c-kit*	5	Geissler et al., 1988
undulated (un)	*Pax 1*	2	Balling et al., 1988
mdx	*Dystrophin*	X	Sicinski et al., 1989
brown (b)	*tyrosinase-related protein 1*	4	Jackson et al., (1990) Zdarsky et al., (1990)
Tdy	*Sry*	Y	Gubbay et al., 1990
T	*Brachyury*	17	Herrmann et al., 1990
steel (Sl)	*Kit ligand*	10	Williams et al., 1990 Copeland et al., 1990 Flanagan & Leder., 1990
Snell Dwarf (dw)	*pit 1*	16	Li et al., 1990 Camper et al., (1990)
retinal degeneration (rd)	*cGMP phosphodiesterase*	5	Bowes et al., (1990)
osteopetrosis (op)	*CSF-1 (c-fms ligand)*	3	Yoshida et al., 1990
Splotch (Sp^{2H})	*Pax 3*	1	Epstein et al., 1991
Small eye (Sey)	*Pax 6*	11	Hill et al., 1991
patch (Ph)	*PDGF-aR*	5	Stephenson et al., 1991
Trembler	*PMP-22*	11	Suter et al., 1992
waved-1 (wa-1)	*TGF-a*	6	Luetteke et al., 1993 Mann et al., 1993
extra-toes (Xt)	*Gli 3*	13	Hui & Joyner., 1993
Bcg	Nramp	1	Vidal et al., 1993
waved-2 (wa-2)	*EDGF*	11	Luetteke et al., 1994
nude	*whn*	11	Nehls et al., 1994
Shaker-1 (sh 1)	*Type VII myosin*	7	Gibson et al., 1995
reeler (rl)	*reelin*	5	D'Arcangelo et al.,1995

to functional analysis using the knock-out technology, although a prerequisite for this complementary approach is the cloning of the endogenous genes to be targeted.

1.1.2 Molecular genetic approaches

1.1.2.1 Mammalian genes isolated based on conserved sequence homologies

One of the most common molecular genetic strategies used for identifying new genes that control mammalian development, is the analysis of mammalian genes isolated on the basis of their sequence homology to conserved sequence motifs within genes known to regulate developmental processes in other organisms, notably *Drosophila* and *C. elegans* (Kessel and Gruss, 1990). Three major murine gene families containing conserved sequence domains of the *Drosophila* homeobox, paired-box and the human and nematode POU-box containing genes have been isolated in this fashion.

1.1.2.1.1 Murine Hox genes

A homeobox is a 183 bp nucleotide sequence encoding a DNA binding domain of 61 amino acids (Gehring et al., 1994). A high degree of conservation of homeobox containing genes has been found between different organisms. In *Drosophila* the homeobox containing genes include three major classes of developmental gene families: the maternal effect genes, the segmentation genes and the homeotic genes. The mouse homeobox-containing genes, the *Hox* genes, were originally identified on the basis of their sequence homologies to the *Drosophila* homeotic gene *antennapedia* (*Antp*) and were found to be organized into four gene clusters believed to arise from duplications. They are located on mouse chromosome 6 (*Hox A*), 11 (*Hox B*), 15 (*Hox C*) and 2 (*Hox D*), respectively (Krumlauf, 1994).

Paralogous genes are defined as those encoding similar proteins from different clusters. These genes were identified, aligned and grouped into subfamilies. The physical location of each *Hox* gene cluster along the chromosome appears to correlate with the spatial order of its anterior border of expression with a few exceptions (*Hoxa-1* and *Hoxb-1*). During neurectoderm development, the anterior boundaries of expression of progressively 5' located *Hox B* genes reside successively posteriorly in the developing hindbrain (Wilkinson et al., 1989; Murphy et al., 1989) except for *Hoxb-1* that are expressed more posteriorly than other *Hox B* genes, even though it is the most 3' genes in the *Hox B* complex. In the developing limb bud, the 3' *Hox D* genes are expressed earlier and more proximally whereas the 5' *Hox 4* genes are expressed later and more distally (Dolle et al., 1989), thus raising speculation that the expression pattern is correlated with the existence of a morphogen released by the zone of polarizing activity (ZPA). The segmental pattern of expression of a significant number *Hox* genes indicate that they may be involved in pattern formation during mouse embryogenesis, analogous to the function of Drosophila homeotic genes.

1.1.2.1.2 Murine Pax genes

The murine *Pax* multigene family (Gruss and Walther, 1992) shares a conserved sequence domain, the paired-box, with the *Drosophila* paired-boxing containing genes that include three *Drosophila* segmentation genes (Bopp et al., 1986). In addition to the paired-box, some murine *Pax* genes also contain a second motif, the paired-type homeobox, both of which encode proteins that contain the helix-turn-helix structures. Interestingly, this protein domain has been shown to render proteins with DNA binding properties. Eight members of the mouse *Pax* gene family (*Pax 1* to *Pax 8*) have

been identified and are grouped into at least four classes. Each class is defined by genes within the class sharing identical intron/exon structure and extensive sequence homology of their pair-boxes, as well as the presence or absence of additional sequences outside of the paired box, notably the paired-type homeobox (Walther et al., 1991). Each class may therefore be composed of paralogous genes, similar to the murine *Hox* gene family. In contrast to the murine *Hox* genes, however, the mouse *Pax* genes have not been found to be clustered. Members of the *Pax* genes are expressed in a distinct spatiotemporal pattern during mouse embryogenesis, especially along the neuroaxis during development of the mouse nervous system. Notably, *Pax 3* and *Pax 6* are expressed differentially along the dorsal-ventral axis of the primitive neural tube. *Pax 3* is expressed in roof plate and neural crest cells while *Pax 6* is expressed in the basal and intermediate plates of the neural tube and later in the developing eye. Two mouse mutants, *splotch* (*sp*) and *small eye* (*Sey*), have been characterized by mutations in *Pax 3* (Epstein et al., 1991) and Pax 6 (Hill et al., 1991), respectively (see table 1). This is consistent with the embryonic expression pattern of *Pax 3* and *Pax 6*. Mutations in the *Pax 3* gene disrupt the formation of central nervous system structures and neural crest derivatives, and homozygous *sp* mutants therefore display spina bifida and encephaly as well as dysgenesis of spinal ganglia and Schwann cells, whereas mutations in the *Pax 6* gene interfere with eye formation in homozygous *Sey* mutants. These two mouse mutants are believed to be animal models representing human Waardenburg syndrome and Aniridia, respectively.

1.1.2.1.3 Murine *POU* box related genes

The *POU* box containing genes were first isolated by virtue of their ability to bind to octamer sites (ATGCAAAT) in enhancer and promoter

elements. They were identified as transcription factors in rat (*Pit-1*) (Ingraham et al., 1988), human (*Oct-1, Oct-2*) (Clerc et al., 1988) and *C. elegans* (*unc--86*) (Finney et al., 1988) and are defined as *POU* family by the sequence homology shared within the *POU* domain (Herr et al., 1988). The *POU* domain is a 150-160 amino acid region consisting of a *POU* type homeodomain of 60 amino acids linked to a *POU* specific domain of 75-82 amino acids located on the amino-terminal side. While both the homeodomain and the *POU* specific domain are involved in DNA binding, the *POU* specific domain seems to carry much of the binding site specificity (Ruvkun and Finney., 1991).

Several murine homologues of the *POU* family have subsequently been identified using cDNA probes derived from *POU* specific domains. The murine *POU* box containing genes are divided into six classes according to the sequence homology between the *POU* homeodomains and *POU* specific domains, as well as similarities of the linker sequence between them (Wegner et al., 1993). The expression pattern of the mouse *POU* gene family is rather interesting. Unlike the mouse *Hox* and *Pax* genes, there is no obvious regularity in the expression pattern of the *POU* gene family. However, most *POU* genes are differentially expressed during embryogenesis. Murine *Oct-4* expression is first detected in the inner cell mass of the pre-implantation embryos and in pluripotent embryonic stem cells, and is subsequently downregulated following stem cell differentiation and eventually confined to the germ cell lineage (Scholer et al., 1989a; 1989b; 1990a; 1990b). Thus, *Oct-4* may be involved in specifying the undifferentiated cell phenotype and act as a master gene in controlling the initial events governing a cascade of developmental processes. *Oct-6* expression is detected in specific neurons in the developing and adult brain as well as in early embryogenesis, and appears to be associated with rapid cell proliferation (Suzuki et al., 1990). *Oct-6* is

believed to play a role in the progressive determination of the premyelinating and myelinating phases of Schwann cell differentiation (Monuki et al., 1989; 1990). Interestingly, *Oct-2* is expressed in olfactory neuroepithelium (Latchman et al., 1992). The olfactory system, including the mitral cells of the olfactory bulb and olfactory neuroepithelium, is the only part of the mammalian central nervous system where growth and differentiation of sensory neurons continue throughout adult life (Graziadei and Monti-Graziadei, 1983). The olfactory neuroepithelium is constantly renewed by proliferating progenitor cells. Therefore *Oct-2* might play a role in the proliferation processes of this region of the nervous system.

The murine *Pit-1/GHF-1* gene also contains the *POU* box, and is first transiently expressed in the neural tube and later reactivated in the pituitary just before anterior pituitary differentiation (Dolle et al., 1990). The *Pit-1/GHF-1* protein has been shown to bind to octamer sites upstream of the growth hormone and prolactin genes (Ingraham et al., 1988., 1990) and appears to be involved in anterior pituitary development. Indeed, the mouse *Snell dwarf* (*dw*) mutants lacking growth hormone, prolactin and thyroid-stimulating hormone, contain a point mutation in the *Pit-1/GHF-1* gene. The point mutation results in a single amino acid exchange from Try to Cys in the *POU* domain, and the mutated protein is no longer able to bind to its target sequence (Li et al., 1990).

In summary, these studies have shown that isolating genes based on sequence homologies offers a valuable complementary approach to identify mammalian genes that play important roles during development. The expression pattern of these genes and correlation of mutations in some of these genes with existing mouse mutants have yielded important clues as to their developmental function. Continued efforts in identifying regulatory

elements and in generating gain-of-function mutations using transgenic mice, combined with gene targeting via homologous recombination to produce null mutations, should help further elucidate the regulation of the developmental pattern of expression, and the biological functions of these developmentally important genes.

1.1.3.1 Mammalian genes cloned by insertional mutagenic markers

The generation of mutations in genes that control developmental processes can be achieved by insertional mutagenesis (Gridley et al., 1990; Soriano et al., 1987). Mouse insertional mutants have resulted from natural proviral elements affecting the mouse genome. However, various experimental methods have been used to generate new insertional mutations in mice. These have included experimental retroviral infection of mouse embryos (Schnieke et al., 1983; Spence et al., 1989) and microinjection of DNA into fertilized eggs (Woychik et al., 1985; 1990; Jaenisch, 1988). Insertional mutations in the DNA of transgenic mice have provided a means to tag mutations of developmental genes in mammals, thus allowing molecular access to and analysis of such recovered mutations. A number of genes that control developmental processes have been cloned by fortuitous insertion into genes by either a retrovirus or a DNA marker in transgenic mice.

1.1.3.1.1 Endogenous proviral insertion

The mouse genome is estimated to contain between 500 to 1000 copies of endogenous retrovirus proviral elements, similar to the transposable P elements in *Drosophila*. (Stoye and Coffin, 1988). Some of the insertions produced by the proviral elements are accompanied by mutations. Thus, integrated retroviral elements provide a tag to identify and eventually clone

closely associated and mutated genes. A number of such mutant loci have been identified (Meisler, 1992) and a few have been characterized.

The recessive *dilute* (*d*) coat color mutation carried by many inbred mouse strains presents a light coat color due to a melanocyte abnormality and is usually associated with a neurological defect. One of the *dilute* alleles (d^{v}) has been identified as a result of the integration of an ectropic leukemia provirus (Jenkins et al., 1981), and the gene of the *dilute* locus has been characterized to encode a novel type of myosin heavy chain, which is differentially expressed in embryonic and adult tissues and is highly abundant in the nervous system (Mercer et al., 1991). The gene product of the mouse *agouti* locus, which controls eumelanin (brown-black) pigmentation in mice, has been found to encode a novel 131-amino-acid protein containing a signal sequence (Bultman et al., 1992; Miller et al., 1993). This novel protein was found to act as an antagonist of the melanocyte-stimulating hormone (MSH) receptor and blocks α-MSH induced eumelanin synthesis (Lu et al., 1994).

1.1.3.1.2 Retroviral Insertion

Jaenisch and coworkers have produced a series of transgenic mouse lines by retroviral infection of preimplantation mouse embryos with Moloney murine leukemia virus (Jaenisch 1976, 1977, 1980; Jaenisch et al., 1981). Of the 48 transgenic lines analyzed, two (Mov-13, Mov-34) were found to be homozygous embryonic lethal (Jaenisch et al., 1983; Soriano et al., 1987). Molecular characterization of the Mov-13 transgenic line showed that Mov-13 proviral sequence integrated within the first intron of the α1(I) collagen gene (Schnieke et al., 1983). This insertion resulted in a tissue specific transcriptional (Hartung et al., 1986) and translational block (Kratochwil et al., 1989) of the α1 (I) collagen gene.

Molecular analysis of Mov-34 integration indicated that the Mov-34 proviral sequence was inserted into the 5' side of an abundantly and ubiquitously transcribed gene and disrupted its expression. The endogenous gene was characterized (Gridley et al., 1991) and shown to contain seven exons and lack a TATA or CAAT box in its promoter sequence. It is predicted to encode a novel 321 amino acid protein with unknown function.

Zhou and coworkers (1993) isolated a candidate gene, nodal, for a retrovirally induced insertional mouse mutation. This recessive mutant is prenatal lethal and seems to be defective in the process of mesoderm induction. The candidate gene, nodal, was found to encode a new member of the transforming growth factor-beta (TGF-β) superfamily and is required in the formation and maintenance of the primitive streak in early embryogenesis (Conlon et al., 1994). Thus, nodal may function as an important signaling molecule essential for mesoderm induction and subsequent axial pattern formation during development.

Retroviral integration induced insertional mutations in mice can also be used as valuable human disease models. Adult transgenic homozygous mice for the Mpv17 integration, produced by exposing mouse embryos to a recombinant retrovirus, display progressive glomerular sclerosis and develop nephrotic syndrome and clinical renal failure. DNA sequences flanking the provirus were used as a probe to clone the cDNA of the mutated gene. The endogenous gene was found to encode a novel putative membrane associated protein of 176 amino acids and is expressed at high level in kidney, which is devoid of the endogenous transcript in homozygous animals (Weiher et al., 1990). Thus, studies of the gene disrupted in the Mpv17 mutant may provide an important tool to investigate mechanisms leading to clinical renal disorder in man.

1.1.3.1.3 Transgenic Insertion

A growing number of insertional mutations have been produced in the process of generating transgenic mice by microinjection of DNA constructs into zygote pronuclei. Transgenes are inserted into, and as a consequence, mutate the genomic loci in transgenic mice at the site of integration. The chromosome mapping of insertional mutants and cloning of insertionally mutated endogenous genes has been greatly facilitated by the ability to use transgenic sequences as linkage markers and cloning probes. Woychik and coworkers (1985; 1990) characterized and mapped one of the first insertional mutations produced by integration of the transgenic construct MMTV-*myc* into the *limb deformity* locus. The *limb deformity* gene was later cloned and identified as a large gene complex capable of generating a number of alternatively spliced messenger RNA transcripts encoding nuclear protein isoforms called *formins* (Woychik et al., 1990; Jackson-Grusby et al., 1992).

Two genes controlling the development of normal asymmetry of the viscera have been identified by transgenic insertional mutations. Yokoyama and coworkers (1993) discovered a recessive mutation caused by insertion of a tyrosinase transgene in chromosome 4 that resulted in a reversal of left-right polarity (situs inversus). Schreiner and coworkers (1993) identified an insertional mutation, *Legless*, in the pHT1 transgenic line (McNeish et al., 1988), which also disrupts the normal development of asymmetry of the viscera. The *legless* gene has been characterized and mapped to chromosome 12 (Singh et al., 1991). Sequences flanking both the tyrosinase and pHT1 transgenic insertion site were cloned, and both genes are predicted to play important roles in controlling embryonic turning and visceral left-right polarity (Klar 1994).

Certain disease genes were identified by transgene induced insertional mutations. The best example is the polycystic kidney disease gene (PKD) cloned by Moyer and coworkers (1994). The structure and expression of the PKD gene are directly associated with the transgene mutated locus. This gene was found to contain sequences motifs conserved in several cell cycle controlling genes. The mutant locus for microphthalmia (*mi*) was also cloned and characterized through use of the transgene as a molecular marker (Hodgkinson et al., 1993). The gene encodes a novel member of the basic-helix-loop-helix-leucine zipper (bHLH-ZIP) protein family of transcription factors, and is expressed in the developing eye, ear and skin, structures affected by *mi*.

These studies demonstrated that insertional mutagenesis is a valuable method for identifying previously unknown mouse loci. Although hundreds of transgenic mouse lines have been produced, relatively few transgene induced insertional mutations, however, have been characterized and mutant loci cloned. This is because many foreseeable problems can arise to complicate the process of molecular characterization. Transgene insertion often cause deletions or rearrangements at the site of insertion. Furthermore, the presence of repetitive sequences around the insertion site can make it difficult to isolate single-copy flanking sequences. Interestingly, close to 50% of visible insertional mutations generated by transgene integration were found to be allelic with many well categorized classical mutant loci (Meisler 1992). The reason for such a high frequencey of re-isolating classical mutations is unclear. The classical loci may provide preferential integration sites for transgene insertions. If this is true, the number of new insertional mutations generated by transgene integration may reach a point of saturation. This is unlikely because even though there are preferential sites for retroviral

integration within a single gene (Natsoulis et al., 1989; Robinson and Gagnon, 1986; Rohdewohld et al., 1987), there is no evidence that there are preferred sites for transgene insertion in the genome. Therefore, one possibility is the limitation of our ability to detect the visible abnormality of insertional mutations. That almost half of the visible insertional mutations are associated with classical mutant loci may simply be due to the fact that the phenotypes of these classical mutant loci have been well documented and well characterized. Thus, while the phenotypes of transgene-induced insertional mutations resembling the classical mutatant phenotypes are relatively easy to associate, many of the transgene generated mutations may be beyond our ability of detection. Furthermore, disruption of an endogenous gene by insertional mutation may not cause any observable phenotypes due to redundancy of analogous gene functions. Thus, the sensivity of detecting insertional mutations based only on mutant phenotypes may present a challenge to using this approach to access most of the developmentally important genes.

1.1.4.1 Trapping strategies

The first of the trapping strategies, the enhancer trap, was initiated in *Drosophila* as a screen method to identify genes on the basis of their expression pattern rather than mutant phenotype (O'Kane and Gerhing, 1987; Bier et al., 1989; Bellen et al., 1989). Analogous trapping strategies have since been applied to the mouse to identify enhancers, promoters, and genes that may play important roles during murine development.

1.1.4.1.1 Enhancer Trap

1.1.4.1.1.1 Eukaryotic gene regulation

In eukaryotic cells, temporal and tissue-specific gene expression is believed to be controlled by cis-acting regulatory elements which interact with specific transcriptional factors in response to extracellular inducers (Mitchell and Tjian, 1989; Dyan, 1989). Cis-acting regulatory elements can be identified either by analyzing the effects of mutations introduced into the cis-acting regulatory DNA sequences on gene expression, or by studying interactions between cis-acting elements and transcription factors revealed by footprinting and gel retardation assays (Maniatis et al., 1987; Mitchell and Tjian, 1989). The promoter element ensures accurate and efficient transcription upon interaction with trans-acting protein factors (Dyan and Tjian, 1985), whereas the enhancer acts to increase the rate of transcription (Serfling et al., 1985). In contrast to promoter elements, which often reside immediately 5' of the transcription initiation site and only function in correct orientation, enhancers can exert their effects many kilobases 5' upstream or 3' downstream away from the transcription start site, independent of orientation (Grosveld et al., 1987), and can also be located in introns. Although different promoters often show a common pattern of organization, composed of a TATA box that binds to RNA polymerase, and one or more upstream promoter elements (Maniatis et al., 1987) , enhancer sequences are not conserved and vary from one gene to another (Serfling et al., 1985). Both enhancer and promoter elements confer developmental and spatial regulation of gene expression during development.

1.1.4.1.1.2 Enhancer trap in transgenic mice

Since the success of enhancer trap strategy in *Drosophila*, efforts have been made to employ the same strategy in transgenic mice to identify genes that exhibit a temporal and tissue-specific pattern of expression during mouse

development. The enhancer trap construct has a weak promoter joined to a reporter gene (i.e LacZ). This weak promoter itself does not permit detectable expression of the reporter gene without the enhancement of transcription provided by an enhancer sequence (Kothary et al., 1988). Thus, the expression of the enhancer trap construct is subject to the chromosome position into which it is integrated (Butner and Lo, 1986) and is probably under the control of an adjacent enhancer-like element. Allen and co-workers (1988) used a construct of the *E. coli* LacZ reporter gene fused to a weak promoter from the herpes-simplex thymidine kinase gene. They showed that, in a number of transgenic mice produced, β-galactosidase staining was observed in two or more disparate tissues, including the nervous system. In adults of one transgenic line, expression of β-galactosidase was exclusively found in the brain. The expression of β-galactosidase was first detected at 11 days of gestation in cells in hind brain, neural tube and somites, and became gradually confined to only certain areas of the brain at postnatal day 5. Thus, the transgene was shown to display a spatially and developmentally regulated pattern of expression in the mouse developing nervous system by the enhancer trap strategy.

The same approach was used in other transgenic mouse experiments using similar constructs. One transgenic line carrying a construct consisting of the mouse heat shock promoter (hsp68) fused to LacZ was shown to express β-galactosidase in cells of the neural tube floor plate running the entire length of the medial ventral neural tube in mouse embryo (Kothary et al., 1988). One possibility here is that the pattern of β-galactosidase expression could be indicative of that of the endogenous gene, whose cis-acting elements might dictate the pattern of β-galactosidase expression in the developing mouse nervous system. It was also demonstrated that this transgene caused an

insertional mutation of the *dt* gene. The flanking region of the transgene insertion has been cloned and used as probes to characterize the wildtype *dt* locus (Brown et al., 1994).

Bonnerot et al (1990) also obtained seven transgenic mice using a construct expressing LacZ from the hypoxanthine guanine phosphoribosyl transferase (HPRT) promoter which normally directs ubiquitous expression of the HPRT gene. In their experiments, six of the seven transgenic mice expressed the LacZ gene, and each transgenic line was characterized by a specific, highly reproducible pattern of LacZ expression that was spatially and temporally regulated. Thus, the expression of the integrated construct must be complemented by cis-acting elements of the endogenous genome which exert dominant developmental control over the HPRT promoter. This study provided further proof that enhancer trap is a valid and useful approach to identify the regulatory elements of developmental controlling genes not only in invertebrates, but also in mice.

1.1.4.1.1.3 Enhancer trap in ES cells

There are limitations to the enhancer trap approach using transgenic mice by the conventional microinjection method. The major constraint is that the number of integration events required is limited by the number of transgenic mice obtained, as the yield of transgenic mice out of the injected fertilized zygotes is usually low (2-6%) (Palmiter and Brinster, 1986). The expression pattern of the trapped enhancer, as indicated by the β-galactosidase reporter gene activity, can only be fully assessed when the transgenic mouse lines are established. This may proof to be economically and experimentally unfeasible.

1.1.4.1.1.3.1 Embryonic Stem (ES) cells

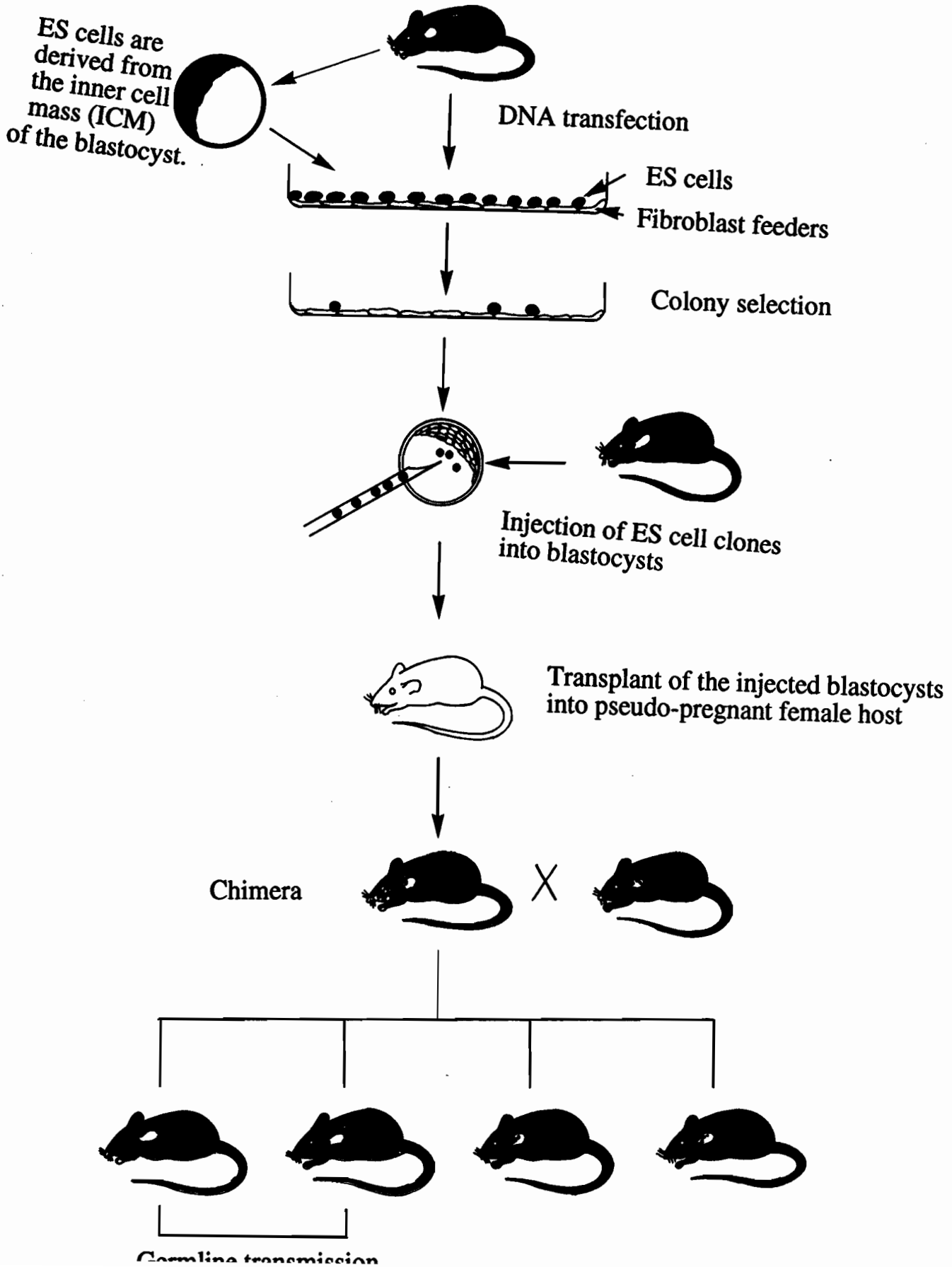
Using transgenic mice, produced by the conventional microinjection method, for enhancer trap analysis is limited due to the low frequency of insertional events generated which can be readily detected. Selection for insertional events in mouse embryonic stem (ES) cells provides one means of increasing the ease of detection of the rare integration events. ES cells are originally derived from the inner cell mass of mouse blastocysts and established in culture. The mouse blastocyst is composed of the inner cell mass (ICM) and the trophectoderm. The ICM is a clump of cells sequestered at one pole of the mouse blastocyst, which eventually give rise to the mouse embryo proper, whereas the trophectoderm, located at the periphery of the blastocyst, contributes mainly to the extraembryonic tissue (Hogan et al., 1986).

ES cells are pluripotent and can be maintained in an undifferentiated state when cultured on a mouse embryonic fibroblast feeder cell layer (Gossler et al., 1986; Robertson, 1987), or in the presence of a soluble polypeptide factor. This soluble polypeptide factor has been refered to, depending on its source of isolation, either as differentiation inhibitory activity (DIA) (Smith et al.,1988), myeloid growth factor human interleukin for DA cells (HILDA) (Moreau et al., 1988) or myeloid leukemia inhibitory factor (LIF) (Williams et al., 1988) (Figure 1).

Evans and Kaufman (1981) first successfully established ES cells from the epiblast of delayed-implantation blastocysts. Almost at the same time, Martin (1981) succeeded in culturing ES cells derived from immunosurgically isolated ICM in medium conditioned by teratocarcinoma-derived embryonic carcinoma (EC) cells. It was then established that ES cells can be manipulated *in vitro* without delayed-implantation or EC-cell conditioned medium (Robertson et al., 1983; Axelrod, 1984). However, in the absence of embryonic

Figure 1. Mouse embryonic stem (ES) cells and their experimental potential. The ES cells are derived from the inner cell mass (ICM) of the blastocyst. They can be maintained pluripotent *in vitro* and are amenable to a wide variety of genetic manipulations. Clones originating from rare events, such as a transfected DNA construct inserting into a specific gene, can be derived using both positive and negative selection systems. Injection of ES cells from such clones into normal blastocysts can give rise to chimeric mice in which the ES cell component has contributed to germ cells as well as to all somatic tissues. Some gametes produced by these chimeras can thus be derived from the ES cell lineage, and some of the F1 progeny from matings between a chimera and a normal mouse can carry exclusively the ES cell component, suggesting successful germline transmission. Therefore, any mutations introduced to ES cells and selected *in vitro* can be transmitted to offspring to evaluate their *in vivo* consequences.

Embryonic stem cells: In vitro and in vivo transition



fibroblast feeder cells (Doetschman et al., 1985; Gossler et al., 1986) or the inhibitory polypeptide factor (Williams et al., 1988; Smith et al., 1988; Moreau et al., 1988), ES cells spontaneously differentiate to form: 1) blood cell-containing cystic embryonic bodies similar to the visceral yolk sac of the embryo, when cultured in suspension, or 2) a variety of differentiated tissues, when cultured on substrate (Robertson, 1987).

ES cells transfected by infection with retroviral vectors (Doetschman et al., 1986; Robertson et al., 1986), or by calcium phosphate precipitation (Doetschman et al., 1986; Gossler et al., 1986), or by electroporation (Gossler et al., 1989) incorporate foreign DNA into their genome. These transfected ES cells display a normal karyotype, maintain pluripotentiality, and are capable of spontaneous differentiation *in vitro* indistinguishable from their untransfected counterparts. When injected into the developing mouse blastocyst, the transfected ES cells, like untransfected ES cells, can contribute to all the adult tissues as well as the germ line of the resulting chimeras (Gossler et al., 1986; Gossler et al., 1989).

Incorporated DNA constructs in ES cells are stably inherited through mitosis and found in cells of all chimera tissues. ES cells carrying a foreign DNA construct can also colonize germ lines and undergo successful meiotic division to generate viable gametes, thus making it possible to transmit the transgene to their progeny. For instance, cells obtained from tissues of the F1 progeny of ES cell derived chimeras transfected with a tk-*neo*^{r} construct, as well as from that of the F2 progeny of such F1 transgenic mice, remain G418 resistant in culture. By contrast, cells obtained from the F1 and F2 progeny of non-transfected-ES cell-derived chimeras do not survive in the presence of G418 (Gossler et al., 1986). Thus, the fact that foreign genes can be integrated into the genome of ES cells and that most of these cells continue to display a

high degree of *in vitro* and *in vivo* developmental potential, together with the feasibility of massive screening of ES cells before generating chimeras, provides a promising experimental system to identify potential regulatory genes and to investigate the temporal and spatial regulation of embryonic genes in mammalian development *in vivo*.

1.1.4.1.1.3.2 Enhancer trap in ES cells

The ES cell-chimera transgenic system has provided an important *in vitro* to *in vivo* transition allowing molecular access to gene regulation of complex mouse developmental events (Figure 1). It is possible to perform large scale screening for relevant mutations or integration events in cell culture, and to introduce genetically manipulated ES cell clones to chimera and transgenic mice in order to assess the developmental expression pattern of targeted genes *in vivo*. Gossler and co-workers (1989) have generated several enhancer trapped ES cell clones, each of which displays a unique pattern of expression in developing chimeras. From one of these enhancer trapped ES cell clones, Soininen and collaborators (1992) have isolated a novel mouse gene, *Enhancer trap locus 1* (*Etl-1*), located in close proximity to a LacZ enhancer trap integration site. This ES cell clone exhibits a spatially and temporally regulated pattern of β-galactosidase staining in the developing chimeras (Gossler et al., 1989; Korn et al., 1992), indicating that the endogenous gene might be expressed in the same manner. *Etl-1* was cloned using the enhancer trap vector as a cloning marker, and was found to be expressed prominently in the mouse central nervous system and epithelial cells. It encodes a novel protein that shares significant sequence homologies with several proteins from different species that are involved in *Drosophila* homeotic gene regulation (*Brahma*) and yeast transcriptional gene activation.

Etl-1 is the first mammalian member of this group of proteins identified and may play a role in gene regulatory pathways during mouse development. *Etl-1* is also the first mouse gene cloned by using a lacZ enhancer trap approach, thus validating the enhancer trap strategy as a means to identify genes that control developmental.

1.1.4.1.2 Promoter trap

In enhancer trap, the pattern of β–galactosidase expression is believed to have been directed by the cis-acting elements of an endogenous gene. Since enhancers can function over a long distance, this host gene can be located many kilobase pairs away from the inserted construct. Thus, the fact that β-galactosidase is expressed is not necessarily an indication that an insertional mutation has occurred. Even if an insertional mutation arises, the cis-acting element responsible for β-galactosidase expression may not be associated with that of the endogenous gene, which may have been disrupted by the transgene. Moreover, cloning transcriptionally active genes by this approach is difficult because enhancers can be located at considerable distance away from and on either side of the site of the enhancer trap vector integration.

It is possible that β-galactosidase expression may result not from the activation of a trapped enhancer but rather from a transcription read-through, in which the enhancer trap construct would be expressed from an endogenous promoter while the ATG codon of the reporter gene (β-galactosidase) happens to serve as the translation initiation codon. Although the latter case is expected to be rare, it deviates from the precise definition of enhancer trap and should instead be termed promoter trap.

Promoter trap vectors contain only the coding sequences of a reporter gene lacking its own promoter. Thus, the expression of a reporter gene can

initiate only from an endogenous promoter. von Melchner and Ruley (1989) designed a retroviral promoter trap construct in which the coding sequence for a selectable marker, the *Salmonella typhimurium* histidinol dehygrogenase (*his*D) gene, was inserted into the 3' long terminal repeat (LTR) of an enhancerless Moloney murine leukemia virus. Before integration into the mouse genome, the elongated viral LTRs are duplicated, placing the *his*D coding sequences in the 5' LTR just 30 bp from the flanking cellular DNA. The viral construct also contained a second selectable marker, neomycin phosphotransferase (*neo*^{r}) driven by the herpes simplex virus thymidine kinase (*tk*) promoter. NIH 3T3 cells were infected by the retroviral vector and cell clones resistant to L-histidinol or G418 were isolated. Histidinol is thought to kill mammalian cells by acting as a competitive inhibitor of histidine-tRNA aminoacyltransferase and, as a consequence, sufficient levels of histidinol dehydrogenase may be required to confer resistance (Hartman and Mulligan, 1988). While the efficiency of conferring *His*^{r} was found to be 2500 fold lower than the efficiency of conferring *neo*^{r}, in each of the *His*^{r} cell clones, the *his* sequences in the 5' LTR gene was expressed by transcripts initiated from an appending cellular promoter. The sizes of the transcripts expressed from the adjacent cellular promoters in different *His*^{r} clones varied by as much as 100 bp. This is expected if the proviruses are located at different distances from cellular promoters. The upstream cellular promoter sequences were cloned by an inverse PCR procedure and subsequently sequenced (von Melchner et al., 1990, von Melchner and Ruley, 1990). Promoter elements isolated from two *his*^{r} cell lines were shown to be transcriptionally active and capable of activating a linked reporter gene, chloramphenicol acetyltransferase (CAT). Subsequently, the same group of investigators (Reddy et al., 1991) constructed another

promoter-trap vector (U3LacZ) with a different reporter gene, *E. coli* LacZ. The LacZ coding sequences were inserted into the 3' LTR instead of the *hisD* gene, and cell clones expressing β-galactosidase activity were identified by histochemical staining and isolated by fluorescence-activated cell sorting. In all cell clones examined, β-galactosidase transcripts were found to be initiated from transcriptional promoters located in the flanking cellular DNA. In addition, cell clones in which LacZ expression was induced and repressed in growth-arrested and log-phase cells were isolated, thus suggesting that the identified promoter elements may participate in cell cycle gene regulation.

Similar to enhancer trap investigations, the promoter trap strategy was eventually tested in mouse embryonic stem cells since the ES cell-chimera-transgenic mice system provides a powerful *in vitro* and *in vivo* transition and allows a functional analysis of the isolated promoter elements during mouse development. The U3His retroviral promoter trap vector (von Melchner and Ruley, 1989) and a new U3Neo vector constructed by replacing the *His* coding sequences with neo^r in the U3His vector were used to infect ES cells (von Melchner et al., 1992). Histidinol or G418 resistant ES cell clones were isolated in which the coding sequences of the His^r or neo^r gene were found to be located in or near the 5' exons of transcriptionally active genes. Several of these ES cell clones carrying the promoter trap induced insertional mutations which were passed into the germ line and two of these were homozygous embryonic lethal. Interestingly, one of the genes disrupted by the promoter trap insertion was identified as a previously cloned gene, REX-1 (Hosler et al., 1989), which is highly expressed in embryonic carcinoma cell lines and is suppressed following cell differentiation. It contains four zinc finger motifs and is thought to function as a potential transcription factor.

Macleod and coworkers (1991) transfected the ES cells with a promoterless *neo*^{r} gene (NASTI) by electroporation in an attempt to identify and mutate genes that are actively transcribed in ES cells. Interestingly, they found that the promoter trap construct preferentially integrated close to or within a CpG island. Although the promoter trap approach selects for active promoters, the fact that a high frequency of CpG islands were identified suggests that there are a large number of CpG island containing genes active in ES cells.

These studies demonstrated that promoter trap provides a rapid and efficient means to isolate promoters active in different cell types and to clone the corresponding cellular genes. Furthermore, an insertional mutation is likely to have been generated by the integration of the promoter trap construct in close proximity to an active promoter element. Thus, promoter trap combines the advantage of enhancer trap in which the activity of the tagged gene can be evaluated in a chimera or a transgenic animal by the expression of the reporter gene, with the advantage of allowing a functional analysis in homozygous trangenic mice as the promoter trap insertion may more frequently be associated with an insertional mutation.

Peckham and coworkers (1989) discovered that in three embryonal carcinoma (EC) cell lines infected with a retroviral construct, which carries the selectable *neo*^{r} gene in place of *gag*, *pol*, and *env* genes of Molony leukemia virus but which retained the viral LTR and regulatory regions, the viral transcripts were initiated from the 5' flanking cellular genome. This was expected since retroviral genes are not normally expressed in EC cells due to inadequate function of the viral LTR promoter in this cell type. Thus, the transcriptional activation can either result from a cellular enhancer or a cellular promoter. However, RNase protection assays revealed that the

transcripts did not initiate from the normal viral initiation site, as might be expected if the cellular enhancer had activated the viral promoter, nor did it result from an obvious transcription read-through, as might be expected if an upstream cellular promoter had been utilized to transcribe the provirus. Since the *neo*^{r} coding sequence is located downstream of the 5' LTR with a inserted splice donor and a splice acceptor in between them, the expression of *neo*^{r} could arise from either the spliced or unspliced transcripts. Indeed, in two cell clones, comparison of genomic and cDNA clones corresponding to 5' cellular sequences, as well as the sizes of the *neo*^{r} messages indicated that the retrovrial construct was integrated into the first intron of the endogenous genes. The RNAs were generated by read-through from active cellular promoters, followed by splicing from cellular splice donor to the proviral splice acceptor. Genomic sequences isolated from one cell clone seemed to have identified a murine homologue of the yeast ribosomal protein gene L3, responsible for trichodermin drug resistance. Thus, trapping strategies not only can be used to identify cellular enhancer and promoter elements, but can also be designed to generate fusion messages between endogenous gene and the trapping construct, therefore isolating the endogenous gene coding sequence directly. This strategy is called gene trap (Gossler et al., 1989).

1.1.4.1.3 Gene trap

1.1.4.1.3.1 Eukaryotic RNA splicing

The eukaryotic genome is composed of protein coding sequences known as exons, which end up in mature mRNA, and the intervening sequences designated as introns, which are spliced out during RNA processing. Despite their complex organization and the significantly longer length of introns compared to exons, the 5' and 3' exon-intron boundaries are

highly conserved and usually named splice donor and splice acceptor sites, respectively. After the DNA sequence of a gene is faithfully transcribed to yield pre-mRNA, the pre-mRNA molecule undergoes a series of processing events, including 5' capping, splicing and polyadenylation, thus generating a mature mRNA molecule that is then transported to the cytoplasm. The precise mechanism of RNA splicing is very complex involving many cis and trans regulatory elements of the cellular splicing machinery. The general scheme of RNA splicing is thought to be a two step process (Aebi and Weissmann, 1987). The first step of RNA splicing involves the cleavage at the 5' exon-intron boundary in pre-RNA and the formation of a 5'-2' phosphodiester bond between the guanosine of the 5' end of the intron and the adenosine of the branch point. This branch point contains a highly conserved nucleotide sequence about 30 bp upstream of the 3' end of the intron. Thus, the first step RNA splicing event gives rise to a lariat intermediate. The second step involves cleavage at the 3' exon-intron boundary and joining of the two coding sequences, releasing the intron in a lariat structure (Padgett et al., 1986; Green, 1986). The production of mature mRNA thus requires the recognition of the splice donor site and splice acceptor site by the cellular splicing machinery, and subsequent cleavage and ligation to remove the intervening sequences and to join the protein coding sequences.

1.1.4.1.3.2 Mechanism of gene trapping

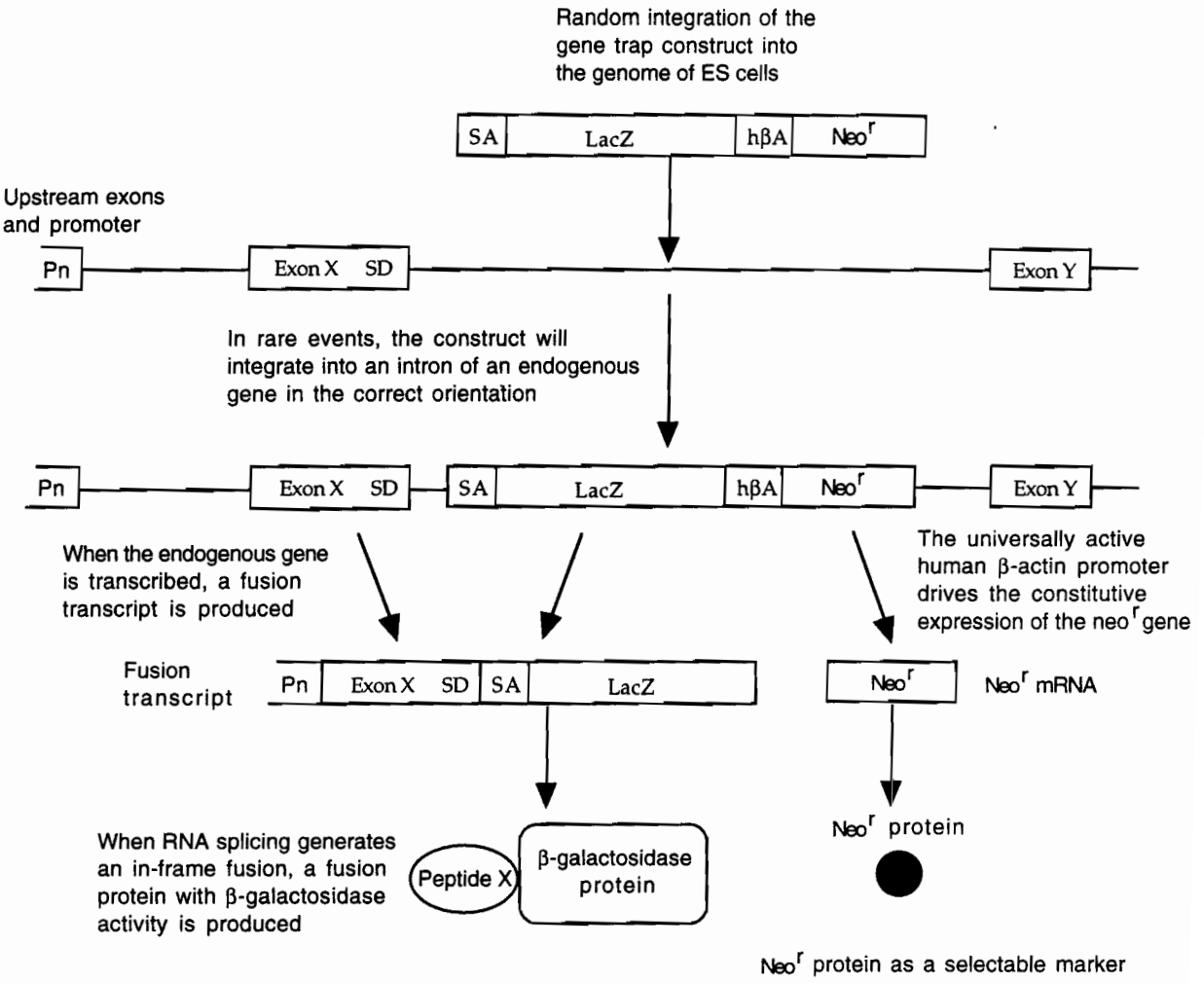
The essential feature and unique advantage of the gene trapping strategy is the production of a fusion transcript. The gene trap construct used involves a reporter structural gene (eg. LacZ) without either the promoter or the translation initiation codon but containing a splice acceptor site at its 5'

end. If this construct is integrated into intron sequences in the genome in the correct orientation, the splice donor site offered by the exon of an endogenous gene and the splice acceptor site donated by the construct can undergo mRNA splicing, resulting in a fusion transcript. If the fusion transcript is joined in frame, a fusion protein product will be produced in which β-galactosidase may retain its functional enzymatic activity (Silhavy et al., 1976; Casadaban et al., 1980). Thus, the developmental profile of β-galactosidase could be highly representative of that of the endogenous gene (Figure 2).

1.1.4.1.3.3. Gene trap as a screening procedure to identify new mammalian genes

Brenner and co-workers (1989) designed a retroviral construct in which a splice acceptor from the viral *env* gene was ligated to the 5' of the *E. coli* LacZ gene in all three reading frames, thus allowing the formation of translational fusions of the upstream endogenous exons with LacZ when the construct is inserted into introns. In one construct an ATG translation start codon was fused in frame with the LacZ gene to detect LacZ expression when the retroviral construct was integrated into introns 3' to a nontranslated first exon. The β-galactosidase expressing fibroblast cells were observed with all three reading frame constructs, although the frequency of identifying blue stained cells was 5-10 fold higher with the construct that contained an ATG codon in frame with LacZ. The increased frequency of detecting β-galactosidase expressing cells with the ATG containing construct reveals a tendency of the retroviral construct to integrate at sites near the beginning of genes and the existence of untranslated first exons in many genes (Crick, 1979). β-galactosidase expressing cell clones were isolated by fluorescent activated cell sorting, and each clone displayed a distinct pattern of

Figure 2. Schematic diagram illustrating the random integration of the gene trap construct into the genome of ES cells. In rare events in which this construct is inserted into an intron of an actively transcribed gene in the correct orientation, a fusion transcript is produced. If the fusion transcript is joined in frame, a fusion protein with β-galactosidase activity, readily detectable by a simple histochemical staining, can be generated.


Random integration of the gene trap construct into the genome of ES cells
SA
LacZ
hβA
Neo^r
Upstream exons and promoter
Pn
Exon X
SD
Exon Y
In rare events, the construct will integrate into an intron of an endogenous gene in the correct orientation
When the endogenous gene is transcribed, a fusion transcript is produced
The universally active human β-actin promoter drives the constitutive expression of the neo^r gene
Fusion transcript
Neo^r mRNA
When RNA splicing generates an in-frame fusion, a fusion protein with β-galactosidase activity is produced
Peptide X
β-galactosidase protein
Neo^r protein
Neo^r protein as a selectable marker

intracellular distribution of the β-galactosidase reaction product. This suggests that the normally cytoplasmically synthesized β-galactosidase is linked to a peptide sequence directing the subcelluar localization of the fusion protein. Interestingly, analysis of β-galactosidase expression following serum starvation has uncovered cell lines in which the LacZ gene was fused to genes induced by growth arrest in the G_0 phase of the cell cycle.

Gossler and co-workers (1989) transfected ES cells with a nonretroviral LacZ gene trap construct, which also contained *neo*^{r} under the control of the human β-actin promoter to screen ES cells that had incorporated the construct into their genome. β-galactosidase activity was detected either in the cytoplasm or the nuclei of ES cells, further indicating that a fusion protein was produced. The endogenous protein domain of the fusion product was likely to have provided the signal directing the subcellular localization of the fusion protein (Silhavy et al., 1976). Selected ES cells that expressed β-galactosidase were injected into mouse blastocysts to generate chimeras. Whereas in some chimeric mice β-galactosidase was constitutively expressed in all tissues, in a number of other chimeras, β-galactosidase showed a spatially restricted pattern of expression, on some occasions occurring exclusively in the developing mouse nervous system (Gossler et al., 1989). The transgenic mice obtained after breeding these chimeras showed that the expression pattern of β-galactosidase was inherited, indicating successful germ line transmission of the transgene (Skarns et al., 1992). Endogenous exon sequences upstream of LacZ in several ES cell lines were cloned by RACE-PCR, and the fusion message was found to be produced by correctly using the upstream splice donor site of the endogenous gene and the splice acceptor of the gene trap construct. Novel cDNAs corresponding to the trapped endogenous genes were isolated, each of which detected a unique-sized

endogenous transcript and exhibited a distinct pattern of developmental expression in chimeras. The expression pattern of the β-galactosidase protein, and that of the associated endogenous genes determined by in situ hybridization, were found to be similar. Several gene trap insertions have been transmitted into the germline and have resulted in recessive insertional mutations (Skarns et al., 1992).

Efforts have been made to improve the gene trap vector to increase the frequency of trap events over random integration. A novel gene trap vector was constructed (Friedrich and Soriano, 1991) in which the *neo*^{r} was placed in frame at the 3' end of a LacZ gene. The new construct, β-geo, is expected to yield G418 resistant cell clones also containing β-galactosidase activity. More than 95% of the ES cell gene trap transformants by either a plasmid or retroviral β-geo transfection were both G418 resistant and expressed β-galactosidase protein. This novel construct was equally capable of generating insertional mutations that could be passed into the germline. Niwa and co-workers (1993) designed a LacZ gene trap construct in which the polyadenylation signal of the *neo*^{r} was removed so that it needed to trap an endogenous polyA signal for expression of *neo*^{r} gene. The frequency of gene trap events were found to be five times higher with this construct. The 5' and 3' flanking regions of the integrated constructs in several ES cell lines were cloned. While all of the 5' flanking genomic sequences were discovered to be novel endogenous promoters, the 3' adjacent sequences were composed of the 3' end of the *neo*^{r} gene and the polyA tail, separated by unknown sequences in between. Although the intervening unknown sequences do not contain the consensus sequence of the mammalian polyadenylation signal, the polyA tracts for the *neo*^{r} gene are clearly derived from the endogenous gene. Interestingly, a large deletion or rearrangement spanning over more

than 10 kb around the 3' integration sites were detected in two gene trap cell lines. Since the expression of the gene trapped locus only requires that the 5' flanking region remain intact, genomic organization in the 3' flanking region was never previously examined (Skarns et al., 1992). Such large deletions and rearrangements of genomic sequences occurring 3' to the integration are significant because the insertional mutant phenotypes observed may not only be due to the disruption of the trapped gene, but could also be attributed to disruption of a gene adjacent to the trapped gene locus.

Thus, the gene trap approach provides a powerful means to isolate numerous unknown genes that function in mouse development. Genes actively transcribed in early mouse embryogenesis are randomly pre-screened at the ES cell stage and genes of interest selected on the basis of when and where they are expressed in chimeric and transgenic mice. The introduction of the LacZ reporter gene has greatly simplified the analysis of the expression profile of the trapped endogenous genes, since the expression pattern of the fusion message accurately reflects the expression profile of the endogenous gene (Skarns et al., 1992). Gene trap vectors are expected to generate spliced fusion transcripts between the reporter gene and the endogenous gene present at the site of integration, thus allowing such tagged genes to be readily analyzed at the molecular level. All insertions of the gene trap construct should result in a mutation in the host gene, therefore permitting the assessment of the function of endogenous genes. In that sense, the gene trap method in ES cells can be regarded as gene-targeting to unknown genes, as opposed to gene targeting of known genes.

1.2 The bromodomain containing genes

The bromodomain is a 85 to 87 amino acid sequence motif found in proteins encoded by genes present in species as diverse as human, mouse, hamster, *Drosophila*, *C. elegans* and yeast (Lygerou et al., 1994, Figure 5B). Proteins containing bromodomains have either one or two copies of the motif. When two bromodomains are present in the same protein, they are either in tandem succession or separated by sequences unrelated to the conserved motif. The sequence homologies among different bromodomains vary between 25-80%, with an average of 50-60% identity between two bromodomains within the same protein. Five amino acids are invariant and others are highly conserved among different bromodomains. The location of the bromodomain(s) within a protein is variable, and the secondary structure prediction indicates that the bromodomain may fold into two α helices followed by reverse turns (Haynes et al., 1992). Although the functional significance of the bromodomain is not clear, studies of individual members of bromodomain containing genes in different species have suggested that it may be involved in protein-protein or protein-DNA interactions during the process of transcription activation.

1.2.1 Yeast bromodomain containing genes

1.2.1.1 Yeast genes with both a bromodomain and a helicase/ATPase domain

The first group of genes found to possess bromodomain(s) was the yeast *SWI* gene family. *SWI* genes were originally identified as regulators of *HO*, a gene involved in mating type switches (Peterson and Herskowitz, 1992). One of the *SWI* genes, *SWI2*, was shown to be involved in transcriptional activation of *HO*, and later found to be identical to *SNF2*, a nuclear protein required for the transcriptional activation of the *SUC2* gene in response to decreased glucose levels (Laurent et al., 1991). Sequence analysis of the

SWI2/SNF2 has revealed the presence of a bromodomain, as well as a domain conserved among a variety of proteins with determined or presumed helicase activity (Laurent et al., 1992; Peterson and Herskowitz, 1992). Helicase activity has been implicated in chromatin decondensation during DNA replication (Gruss and Sogo, 1992). The helicase domain in *SWI2/SNF2* contains seven subregions, which are highly conserved (Henikoff, 1993) among a variety of known or putative helicase from different organisms including *E. Coli*, viruses, yeast, *Drosophila*, mice and humans (Gorbalenya et al., 1988, 1989; Hodgman, 1988; Linder et al., 1989; Okabe et al., 1992; Soininen et al., 1992; Troelstra et al., 1992; Delmas et al., 1993; Schaeffer et al., 1993). Two of these seven regions match the consensus sequence of the bipartite NTP binding motif common to ATP or GTP dependent enzymes. In *SWI2/SNF2*, these two regions have been shown to hydrolyze ATP actively and are required for transcriptional activation (Laurent et al., 1993). *SWI2/SNF2* functions as a complex with other *SWI* and *SNF* proteins to facilitate activation by gene-specific regulatory proteins. (Laurent et al., 1991; Laurent and Carlson, 1992). Although the bromodomain is conserved in many transcriptional regulators, including *SWI2/SNF2*, and may mediate protein-protein interactions among transcriptional activators (Haynes et al., 1992), it did not appear to be essential in conferring the transcriptional regulation function of *SWI2/SNF2* (Laurent et al., 1993). Thus, the *SWI* and *SNF* proteins seem to form a multimeric complex that mediates transcriptional regulation, and each *SWI* or *SNF* protein constitutes an indispensable subunit of the complexes. A direct association of the *SWI/SNF* complex and the nuclear localized glucocorticoid receptor (GR) has been demonstrated, and *SWI/SNF* factors have been shown to be required for the *in vitro* and *in vivo* positive transcriptional regulation of GR (Yoshinaga et al., 1992). The purified

SWI/SNF complex was found to consist of 10 subunits and activate the ATP-dependent binding of GAL4 to nucleosomal DNA (Cote et al., 1994).

Several yeast *SWI2/SNF2* homologues have been identified, one of which, *STH1*, was cloned by screening a yeast genomic library with a *SWI2/SNF2* probe. *STH1* was shown to be highly homologous, but distinct both in structure and function to *SWI2/SNF2*. The *STH1* nuclear protein contains a helicase domain and a bromodomain that are 72% and 46% identical to that of the *SWI2/SNF2* protein, respectively. It exerts an essential function for yeast mitotic growth.

1.2.1.2 Yeast genes with bromodomain(s) only

Besides the *SWI/SNF* trasncription complex, The *SPT* genes, originally identified as suppressors of insertion mutations caused by *Ty* elements, also form another transcription complex and closely interact with TATA binding proteins (Eisenmann et al., 1992). One member of the family, *SPT7*, was shown to contain a highly conserved bromodomain. The recently identified *BDF1* gene, was originally cloned from a temperature sensitive mutant that fails to activate a reporter construct containing the small nuclear RNA (snRNA) gene at non-permissive temperatures. The *BDF1* protein contains two copies of bromodomains without the helicase motif and regulates the snRNA gene by transcriptional activation (Lygerou et al., 1994). Another yeast protein that contains the bromodomain without the helicase motif is GCN5, which was first identified as a positive transcriptional regulator of amino acid biosynthetic genes (Georgakopoulos and Thireos, 1992). It was later discovered that GCN5 forms part of the transcriptional complex and is essential in promoting the activation domain of other transcription activators. Deletion of the bromodomain reduces the activity of the activation domain of GCN5

(Elfring et al., 1994). Thus, *brm* appears to be the closest functional homologue of *SWI2*.

1.2.2.2 *Drosophila* genes with bromodomain(s) only

The *Drosophila* female sterile homeotic (*fsh*) gene is a maternal effect gene involved in establishing embryonic segments and specifying their identities. The *fsh* gene interacts synergistically with the homeotic gene *Ultrabithorax* (*Ubx*) and with the positive homeotic gene regulator ***trithorax*** (*trx*) in giving rise to homeotic transformations. The cloned *fsh* cDNA is expressed in ovaries and early embryos and encodes a protein that contains two bromodomains (Haynes et al., 1989). Interestingly, The bromodomian within the *fsh* protein shares unusually high sequence homology (80%) to that of the human *Ring3* protein, which is encoded by the *Ring3* gene in the human major histocompatibity complex II (MHC II) locus (Beck et al., 1992). The *Ring3* protein also contains two bromodomains. Although the function of the *Ring3* gene has yet been determined, the structural similarity between *fsh* and *Ring3* implies a potential functional correlation.

1.2.3 Human bromodomain containing genes

1.2.3.1 Human genes with both bromodomain and helicase/ATPase domains

The human homologue of yeast *SWI2/SNF2* and *Drosophila brm*, *BRG1* was cloned by Khavari et al. (1993) using a *brm* cDNA probe to screen a human Hela cell cDNA library. *BRG1* turns out to be a nuclear protein that shares extensive sequence homology within its DNA-dependent ATPase domain and the bromodomain to that of the yeast *SWI2/SNF2* and the *Drosophila brm*. The DNA-dependent ATPase domain seems to be important for the function of transcription activation. Point mutation of the conserved

ATP binding site in *BRG1* abolished its ability to complement transcriptional activation in *swi2*$^-$ yeast cells, and also caused a dominant negative effect in transcription activation in human cells. In addition, *BRG1* can form a complex with the retinoblastoma protein to induce cell cycle arrest (Dunaief et al., 1994). However, no mutation analysis was done on the bromodomain to assess its functional significance in mediating the transcriptional regulation of *BRG1*. Another human counterpart of the *Drosophila brm* gene was cloned by Muchardt and Yaniv (1993) and named *hbrm*. The *hbrm* protein is distinct from *BRG1*, and consists of a proline/glutamine-rich region followed by a charged amino acid region (P/Q-charged domain), a conserved helicase domain and a bromodomain. Both the P/Q-charged domain and the helicase domain are essential for *hbrm*'s function as a strong transcriptional activator. They are required for *hbrm*'s co-operativity with nuclear receptors, such as the glucocorticoid receptor (GR), and to a less extent, the retinoic acid receptor α (RARα), in trancriptional activation of their respective target promoters. In contrast, the bromodomain seems to be dispensable in the co-operation between *hbrm* and nuclear receptors. However, the bromodomain deletion mutant appeared to have a higher steady state protein level in transfected cells compared to the wild type *hbrm*, and the mutant protein is no longer restricted to the nucleus. Therefore, the bromodomain may not be involved in transcriptional activation, but could regulate the turnover and the nuclear transport of the protein.

1.2.3.2 Human genes with bromodomain(s) only

Several human genes that contain only a bromodomain but not the helicase and ATPase domains have been discovered, most of which have been implicated in transcriptional regulation of gene expression. The human

CCG1 gene was identified as essential in regulating the progression from G1 into the S phase of the cell cycle by modulating the activity of a subset of genes required for entry to S phase (Sekiguchi et al., 1991; Wang and Tjian, 1994). A temperature sensitive mutant hamster cell line (*ts-13*) that causes cell cycle arrest at the G1 phase was used as a recipient of the DNA-mediated using human gene transfer. The human *CCG1* gene was cloned from the rescued *ts-13* transformants using human Alu sequence as a probe. The human *CCG1* gene encodes a DNA binding protein that contains a bromodomain. In addition, the putative protein has four potential phosphorylation sites for casein kinase II and was phosphorylated by this enzyme *in vitro* (Sekiguchi et al., 1991). It also contains two consensus sequences for cAMP-dependent protein kinase and four for Ca^{++}-calmodulin-dependent protein kinase phosporylation sites. Interestingly, the *CCG1* protein was later found to be identical to the 250K subunit of the transcription factor IID (TFIID) complex (Ruppert et al., 1993). This subunit is a potential binding site for activators, and is essential in inducing activator-dependent transcription (Hisatake et al., 1993). Furthermore, TFIID 250k was shown to be able to rescue the *ts-13* cell line and overcome the G1 arrest (Wang and Tjian, 1994). Intriguingly, a newly cloned and mapped human retinoblastoma protein, p300, also contains a bromodomain (Eckner et al., 1994). p300 was shown to mediate adenovirus E1A oncoprotein induced transcriptional silencing effect on the SV40 enhancer. Binding of E1A to p300 allows S-phase entry of quiescent cells, suggesting that p300 may be a negative regulator of cell growth (Moran, 1993). The retinoblastoma protein (RB) has also been shown to potentiate glucocorticoid-redeptor-activated transcription in collaboration with *hBrm* (Singh et al., 1995). Thus, that a bromodomain containing protein can act as either a cell cycle activator (CCG1) or inhibitor (p300) indicates that the

bromodomain motif may function as an adaptor molecule that mediates the activational or inhibitional regulation of transcription factors.

1.2.4 Mouse bromodomain containing genes

One of the mouse bromodomain containing genes, *brg1*, was cloned by screening mouse cDNA libraries with a cDNA probe derived from the human bromodomain containing gene *BRG1*. A partial mouse *brg1* cDNA was obtained and found to be highly homologous to the *Drosophila brahma* (*brm*) gene in regions of the bromodomain, the putative helicase domain and the DNA dependent ATPase domain (Randazzo et al., 1994). The sequence conservation in the bromodomain signifies that *brg1* may be involved in the transcription activation of its downstream target genes. Interestingly, *brg1* is widely expressed during early mouse embryogenesis and later differentially expressed in different embryonic tissues. *brg1* RNA is especially abundant in the mouse developing nervous system and vertebral column in areas where developmental patterns have been well defined by murine Hox gene expression. In light of *brm* playing a significant role in regulating the *Drosophila* homeotic genes, *brg1* may function as a murine *brm* homologue in regulating the murine homeobox containing gene during mouse pattern formation. The expression pattern of the mouse Hox gene family is extremely complex. Thus, it would not be surprising that Hox gene regulators are composed of sophisticated multigene families that coordinate their functions. Given the existence of multiple bromodomain genes in other organisms, if their mouse homologues have any functional role in regulating Hox gene expression during mouse morphogenesis and organogenesis, more murine bromodomain containing genes are yet to be discovered to fulfill this daunting task.

CREB is a transcription factor that binds to a DNA element known as cAMP regulated enhancer (CRE). Once phosphorylated by protein kinase A, CREB was shown to interact with a co-activator CBP, a binding protein of CREB. The cDNA of CBP was cloned from an expression library by virtue of its ability to bind CREB, and was found to encode a nuclear protein containing a bromodomain (Chrivia et al., 1993). A 133 amino acid domain (amino acids 1680-1812) in the C-terminus of CBP, which contains the zinc-finger structure and is conserved in the yeast coactivator ADA-2, was shown to be essential in CBP's function in transcriptional activation via interaction with TFIIB. Interestingly, the bromodomain (amino acids 1108-1194) resides outside of this transcriptional activation domain in the C-terminus. No deletion studies were done on the bromodomain to asses its function in transcriptional activation (Kwok et al., 1994). Lundblad and coworkers (1995) showed recently that p300 was able to substitue for CBP in potentiating CREB-activated gene expression, and this function was suppressed by adenoviral E1A. Since the bromodomain was identified in both CBP and p300, this motif thus appears to be a common feature of these transcription activators, although no data has directly pinpointed a role of the bromodomain in transcriptional activation.

In summary, although many genes of different organisms encode proteins that contain bromodomains, they seem to, structurally and functionally, belong to different gene families. While a large number of bromodomain containing proteins also possess helicase and ATPase motifs (*SWI2/SNF2, STH1, brm, BRG1, hbrm, brg1*), there are a few proteins in different species having only the bromodomain (*SPT7*, GCN5, *fsh*, *Ring3*, p300, *CCG1*, CBP). Both types of proteins seem to be involved in regulating transcriptional activation. But in proteins with both the helicase domain and the bromodomain(s), the bromodomain is usually functionally dispensable in

transcription regulation, even though it has been implicated in regulating protein stability and nuclear transport. Since helicase and ATPase domains may function to unwind DNA helices and disrupt chromatin structure, these proteins may be directly involved in transcriptional activation. The bromodomain may be important for those proteins with only this conserved motif to function as transcription regulators. In the case of *CCG1* and CBP, the bromodomain containing protein either forms a subunit of a large basal transcriptional complex TFIID (*CCG1*), or it interacts directly with a basal transcriptional factor TFIIB to facilitate transcriptional activation (CBP). Thus, these proteins appear to mediate protein-protein interaction and modify transcritptional regulation by acting as an adaptor for other transcription activators.

1.3 The pattern of mammalian central nervous system development

The mammalian central nervous system is composed of an enormous number of highly differentiated neurons and supporting glia cells. A remarkable feature of neural development is the manner in which so many different types of neurons are generated by following a strict spatial and temporal sequence. Structurally diverse regions of the central nervous system arise from an apparently homogenous neuroepithelium. Neuronal precursors originate within the ventricular zone (VZ) near the lateral ventricles (brain) or the central canal (spinal cord). One important step in neurogenesis is the migration of young neurons to their final sites of differentiation and the formation of specific axonal connections with other neurons. Cell migration and distribution is thus important in neural development because cells eventually reside away from where they originate.

1.3.1 The pattern of cortical development

The mammalian cerebral cortex is a highly laminated structure consisting of multiple layers of distinctive cytoarchitectonic areas, and is organized into various functional domains. The neocortex develops as a series of waves of progenitor cells that migrate outward from the proliferative VZ lining the cerebral ventricles. Each successive wave bypasses its predecessors and finally resides external to them, producing the typical six-layered mammalian cortex in an "inside-out" manner. Therefore, migration seems to be an obligatory step to move young neurons from their birthplace to their appropriate sites of differentiation. What migratory path these neural progenitors follow to generate the laminated layers with highly differentiated functional units is a matter of controversy. Two major patterns of neuronal migration have emerged, one proceeding radially away from the VZ (Rakic, 1988) and the other tangentially parallel to the brain surface (Price, 1993). Radial glial processes appear to guide radial migration (Rakic, 1971; 1972) whereas axons from preexisting neurons seems to direct tangential migration (Rakic, 1990).

The radial unit hypothesis put forward by Rakic (1988) derives from numerous autoradiographic studies combined with structural analyses of migratory behavior of cerebral neurons in a series of images. It postulates that positional information for prospective neurons is already determined at the level of VZ, the layer of precursor cells that generates cortical neurons and glia. Neurons in the deeper layers are born and migrate before neurons in more superficial layers. This hypothesis thus implies that very early in neurogenesis, a small group of cells would become committed to their specific positional fates. The final topographic locations of cortical neurons depends on the spatial distribution of their ancestors in the ventricular zone, where

positions of progenitors have been assigned on the "protomap". There is a point-to-point correlation between the eventual cytoarchitectonic positions of mature cortical neurons and the topographic representation of their ancestors in the ventricular zone. This hypothesis thus predicts that postmitotic neurons migrate in a strict radial fashion along the pre-formed radial glial fibers to reach their final assigned destination, and each radial unit represents a compartment consisting of clonally related progeny of one or more neuronal precursors. These cells within the compartment migrate radially and respect the boundaries of the unit. An alternative theory argues against the idea of predetermined positional map (O'Leary, 1989). It postulates that the architectural organization of the developing cortex possess a great deal of plasticity rather than being rigidly decided in early neurogenesis. The pattern of cortical development is greatly influenced by extrinsic factors such as the innervating fibers from other regions of the brain near the end of neurogenesis. Thus, the migration pattern of cortical neurons during neurogenesis becomes a central issues, since the radial unit hypothesis predicts a strict radial migration, whereas the latter theory foresees a widespread tangential dispersion that disregards the radial boundaries.

The radial unit hypothesis is originally based on ^{3}H-thymidine labeling experiments in which cells were labeled at their last division and their migrating behavior analyzed by a series of autoradiographic images. This approach, however, only provides information on the pattern of a large group of cells born at the same time, but not about the specific movements of individual cells. Studies using mouse aggregation chimeras (Crandall and Herrup, 1990; Fishell et al., 1990) provided support to the radial unit hypothesis. Mouse chimeras are composed of two types of cells with distinguishable genotypes. The ratio of the two different genotypes of cells in

various regions of a chimeric tissue is expected to be the same, if there is an extensive and random cell mixing during the development of that tissue. This seems to be the case in the developing somatosensory barrel fields (Goldowitz, 1987). The development of the chimeric cerebral cortex revealed a different pattern (Crandall and Herrup, 1990; Fishell et al., 1990). A high degree of discrepancy in neuronal genotype ratios was observed in the chimeric cerebral cortex (Fiehell et al., 1990), in particular, in spatial distribution in the anterior-posterior, but not the medial-lateral dimension (Crandall and Herrup, 1990). Clonally related neurons are thus organized into radial patches perpendicular to the neuraxis, a finding consistent with the radial unit hypothesis. Studies using prospective tracing methods such as fluorescent dye injections or recombinant retroviral labeling that mark cell lineages provided alternative supporting evidence. Grove and coworkers (1993) labeled progenitor cells in the embryonic rat cortex using a retroviral vector that expressed the *E. coli* β-galactosidase gene. They discovered that the labeled progeny of cortical precursors form clusters. These clusters were composed of a single cell type, either neurons or astrocytes in grey matter, and oligodendrocytes or astrocytes in white matter, as revealed by immunocytochemistry and intracellular dye labeling. Similar results were obtained by Luskin and his colleagues (1988, 1993). When E12-E14 mouse embryonic cortices were infected with a retroviral label and the spatial distribution of their progeny examined at E18-PN0 and PN7-PN23, respectively, clonally related daughter cells (either neurons or glia) were found to be arranged in a radial array (Luskin et al., 1988). Injecting the rat cerebral cortex at the onset of neurogenesis also produced neuronal- and glial-specific clones in the developing cortex and striatum (Krushel et al., 1993). Furthermore, cells that made up the clusters were found to be of a particular

type of neuron or glia (Luskin et al., 1993). A prominent postnatal clustering of clonally related neurons were also found in the anterior part of the subventricular zone (Luskin, 1993). In keeping with the radial unit hypothesis, these results suggested that separate progenitors may exist for different types of neurons or glia very early in neurogenesis, and lineage determination may have occurred among heterogeneous neuronal progenitors within the proliferative ventricular zone.

The strongest evidence for the presence of an alternative migration pattern during corticalgenesis also came from studies employing retroviral cell lineage marking (Austin and Cepko, 1990; Price and Thurlow, 1988; Walsh and Cepko 1988). Using a retroviral vector carrying the *E. Coli* β-galactosidase gene to label the precursors and subsequently visualize the migration pattern of their progeny, Walsh and Cepko (1988) discovered that clonally related cortical neurons did not rigidly follow a radial path. Rather there is a substantial tangential migration crossing the boundaries of different radial columns. The error that could be introduced in interpreting these retroviral labeling experiments using a simple *E. Coli* β-galactosidase gene indicator is two fold. On the one hand, two daughter cells derived from the same progenitor can be mistakenly perceived as two different clones if they migrate far apart (splitting error). In addition, daughter cells of two different clones may cluster together and be wrongly viewed as members of the same clone, when in reality they are derived from two different infected precursors (lumping error). Therefore, a technique that would allow differential labeling of each individual clone was required to address such a question. This much needed innovation was indeed designed by Walsh and Cepko (1992). In their modified retroviral labeling strategy, cells were marked with a library of retroviral vectors containing distinct DNA inserts. Since different DNA

inserts could be identified by polymerase chain reaction, progeny of each labeled precursor could thus be traced by virtue of the distinct retroviral DNA insert they carry. This method was thus able to identify clonally related daughter cells whether they were clustered together or separated far from each other. Walsh and Cepko (1993) infected the ventricular zone with a library of 100 retroviruses in low titre so that only a few cell clones were infected. The low viral titre infection was intended to reduce the probability that cell clones carrying the same DNA insert actually result from multiple infections of more than one progenitor with the same viral vector (Kirkwood et al., 1992). They observed that three days following infection, the majority of labeled cells remained radially aligned with little dispersion. After six days, however, about half the cells underwent widespread tangential dispersion, some migrated as far as 2 mm and some dispersed beyond the neocortex into the neighboring olfactory bulb. These results thus suggested that there are two modes of cell migration during cortical development. While radial movement is prominent at the start of cortical development, tangential dispersion appears to be the principle mode of cell allocation during late neurogenesis. Thus, in contrast to the radial unit hypothesis, these investigations revealed that clonally related neurons migrate rather freely across functional boundaries, and specification of cortical domains occurs well after neurogenesis.

The precise time sequence of cell migration, which could not be addressed by the retroviral studies, was investigated by Fishell and coworkers (1993). Cells in the VZ in whole-mount preparations were fluorescently marked, and cell movement within the VZ of E15 murine telencephalon was visualized by time-lapse video fluorescent microscopy. This method, however, only allowed the migrating behavior of individual cells to be

followed over much shorter periods than that permitted by the retroviral technique. Nevertheless, they observed extensive horizontal cell movements within the ventricular zone (VZ), although there were clear boundaries in the VZ that these progenitors did not cross. Thus, despite the existing restrictions in the pattern of cell migration within the proliferative zone, the broad horizontal cell mixing could account for the widespread cell distribution seen by Walsh and Cepko (1992; 1993). This is because horizontal cell movement in the VZ could allow daughter cells from the same progenitor to separate from each other on a horizontal plane and end up in various final destinies even if a strict radial migratory path is followed by each sibling. The majority of migrating cells in the overlying intermediate zone was indeed observed to follow the radial path by time-lapse confocal microscopy, although 20% of them were found to disperse tangentially (O'Rourke et al., 1992).

Several inherent problems exist for the retroviral and fluorescent cell labeling technique. Both methods only label a small number of cortical progenitors at one time, thus allowing the migration trajectory of only a fraction of cortical cells to be followed. Although dismissed as statistically unlikely (Walsh et al., 1992), the posibility that spatially divergent cell clones possessing the same viral DNA tag may arise from multiple infections with the same viral vector can not be totally discounted (Kirdwood et al., 1992). Furthermore, Retroviral labeling are most amenable for marking cells late in neurogenesis (post E12.5 in the mouse and post E14.5 in the rat), the precise migratory behavoir of precursor cells in early neurogenesis could therefore not be revealed. Finally, although retroviral labeling could identify the final divergent positions of sibling cells of a seemingly common progenitor, it could not determine whether siblings follow different migratory paths, one

radial and one tangential. Thus, to further our understanding of the cell migration pattern during corticalgenesis, an alternative approach was needed.

A novel method of labeling cortical cells was devised by Tan and Breen (1993), who took advantage of a transgenic mouse line carrying a LacZ transgene on the X chromosome. Due to X chromosome inactivation in female transgenic mice, approximately half of the cells express β-galactosidase, thus appearing blue; and half do not, thus appearing white. Since X chromosome inactivation occurs before corticalgenesis, the developing cortex should have been a fine grained mosaic of blue and white cells. However, Tan and Breen (1993) observed that in the developing cortex of female transgenic mice, the blue cells were organized into radial stripes. These stripes varied in position and width, suggesting a random mosaicism rather than functional domains in the developing cortex. The presence of such multiple radial stripes in the organization of the developing cortex indicates that cells within the radial stripes must have followed a strict radial migration pattern and were unable to cross radial boundaries freely. In addition to the striking radial columns observed, a component of tangential dispersion of cortical cells was also revealed by their study. About one third of cells in either the blue or the white stripes are of the other color. These cells of the opposite color must therefore have migrated from either the neighboring radial columns or radial columns located some distance away. Thus, consistent with the retroviral experiments by Walsh and Cepko (1993), there appears to be two population of cells during corticalgenesis: one which migrates strictly radially, thus generating radial stripes; the other that disperses tangentially and can cross radial boundaries.

A question which remains unanswered by the retroviral and transgenic investigations is the identity of the migrating cells. Since

oligodendrocyte progenitors are known to distribute widely and are not essential in cortical patterning, it is thus important to determine whether the radially or tangentially dispersing cells are neurons or glia. In a more comprehensive follow-up study, Tan and coworkers (1995) identified that the radially and tangentially migrating cells are either neurons or glia based on morphological criteria and immunocytochemical double labeling. The radial stripes seen in the transgenic cortex were not composed of descendants from one single progenitor since both neurons and glia were found to make up a single radial column. In addition, among the tangentially dispersing cells, the ratio of neurons and glia was 2 to 1. In their original report (Tan and Breen, 1993), the extent of contribution of radial and tangential migration pattern to the cell mixing in the radial stripes was found to be 2 to 1, as reflected by the ratio of predominant colored cells to less prominent colored cells in a given radial column. The same ratio was identified for neurons and glia (Tan et al., 1995). Moreover, this ratio remained constant along the longitudinal axis and in the medial and dorsal-lateral regions of the developing cortex, whereas no obvious radial stripes were observed in the lateral most one third of the cerebral cortex. These results indicate that there are clearly two distinct patterns of cell migration during corticalgenesis. The radial dispersion seems to be the prominent mode, since in most regions of the developing cortex, two thirds of the cells within a given stripe are of the same color. Furthermore, the same ratio was found for neurons and glia when they are separately analyzed in a single radial column. That is, one third of all neurons or glia in a given stripe are of opposite color. Nevertheless, there also exists an indisputable tangential migration that accounts for one third of cells of opposite colored cells found in a given radial column. Alternatively, tangential dispersion could be considered as a prime mode of cell allocation.

This is because only half of the two third majority-colored cells in a given stripe may represent the radially dispersing clones. The other half (accounting for another 1/3) has an equal chance as the one third of cells of the opposite color to migrate tangentially and enter the seemingly radial stripes. Taken together, during cortical patterning, there is a subpopulation of cortical neurons and glia that migrate radially following the protomap as proposed by Rakic (1988). Those cells may be important in imprinting the positional information and laying down the foundation for different functional domains, whereas the tangential migrating cells add additional complexity and flexibility during the construction of the cerebral cortex.

X-chromosome inactivation in transgenic mice provides an attractive experiment model to study cell mixing and mosaicism in various tissue genesis. This system is similar to the use of mouse chimeras to study cell lineage, with the advantage that X-chromosome inactivation is a naturally occurring event within a normally developed mouse. More importantly, the labeled and unlabeled cells in the transgenic mice are genetically identical and therefore are equipotent in every biological aspect. However, a transgene carried on the X-chromosome has been shown capable of partially escaping from X inactivation and to be reactivated in a small subset of cells (Wu et al., 1992). This phenomenon makes the interpretation of cell lineage analysis in X inactivated transgenic mice more complicated. Therefore, construction of chimeras for the study of cell lineage and pattern formation of different tissues might possess an ultimate advantage over the X inactivated transgenic system.

1.3.2 Pattern of spinal cord development

The spinal cord arises from the neural tube, which in turn develops from the neuroplate with an apparently homogenous neuroepithelium. The neuroplate folds up to form a hollow neural tube that surrounds the central canal. The rostral part of the neural tube gives rise to the brain, while the caudal part develops into the spinal cord. The ventricular zone that encloses the central canal is the origin of the central nervous system.

Like their cortical counterparts, spinal neurons are generated in the VZ and eventually migrate to their final sites of differentiation. The migratory pattern of spinal neurons and glia have been extensively studied in the embryonic chick spinal cord. Leber and coworkers (1990) introduced recombinant retroviral markers that contained the LacZ reporter gene into the genome of dividing progenitor cells and followed the distribution of single clone progeny both within the VZ and during migration. They showed that while progenitor cells in the VZ are pluripotent and are capable of giving rise to cells with multiple phenotypes, including neurons and glia, there was a clear spatially restricted pattern of migration when precursor cells move out of the VZ to their final sites of differentiation. The dispersion of clonally related neurons was very restricted along the rostral-caudal axis, spanning less than a quarter of a segment in length, although there was evidence of wider longitudinal migration of glia in late spinal cord genesis (Leber and Sanes, 1995). There was also a moderate degree of restriction of cell movement along the dorsal-ventral axis, and clonally related cells were confined to planar arrays (Leber et al., 1990). Most clonally related cells were arranged in radial arrays when they started to move out of the VZ, indicating that they initially followed a strict radial path. These clonal progeny then turn sharply and migrate dorsal-ventrally or ventral-dorsally in parallel to the VZ in a tangential pattern. Furthermore, there is a progressive restriction in the

extent of cell dispersion within the VZ (Leber and Sanes, 1995). Spatially restricted distributions of spinal neurons and glia indicate that during development, progenitor cells in the spinal cord may not mix randomly. The spatially confined cell clusters or longitudinal segments represent clonal progeny of a single progenitor that occupy a spatially distinct compartment with specific borders.

In contrast, Musci and Mullen (1992) reported an extensive cell mixing during spinal cord development using aggregation mouse chimeras. In their investigation, mouse chimeras were constructed from two distinct genotypic strains, one with high and the other with low levels of β-glucuronidase. The enzymatic reaction can be detected histochemically and the differential level of enzymatic activity is thus determined to represent the genotype ratio. When the genotype ratio of different regions of the spinal cord in adult chimeras are compared, they discovered an average of only 4 to 5 percent difference in left-right, cranial-caudal, and dorsal-ventral regions of the spinal cord. These results suggest that during spinal cord neurogenesis there is a significant degree of cell mixing, as reflected by the fine-grained mosaicism of the spinal cord cells indicated by the genotype ratio in the chimeras. Several inherent problems, however, exist in this study. The cell marker β-glucuronidase used to distinguish between cells of different genotypes can only be applied with relative confidence to large cells with diameters above 10 μm. Histochemical staining becomes ambiguous with smaller cells. Thus, quantitation of β-glucuronidase levels was limited to cells greater than 10 μm in the above study. Furthermore, although within a region the variation of left-right, dorsal-ventral was only about 5%, the whole left-right, dorsal-ventral differences within the cervical, thoracic, and lumbar regions were statistically significant, with a maximum difference of 18% observed for both

orientations. There were also significant rostral-caudal gradients observed in five of the eight chimeras analyzed, with one reaching as much as 10% in longitudinal genotype variations. Moreover, there were substantial regional variations (up to 30%) when subregions of different spinal segments were compared. Finally, as the authors pointed out, the genotype ratios were calculated as a whole for a designated area. Thus, the fact that there were no significant differences between regions did not necessarily mean that there was a homogenous and random cell mixing, but rather that the variations between subregions had averaged out to roughly the same proportions in the regions.

The study conducted in this thesis utilized chimeras produced by blastocysts injected with a gene trapped ES cell clone. The pan-neuronal marker expressed by this ES cell clone provided a useful means to study the pattern of generation of the nervous system. The question of the pattern of cell migration during cortical genesis and early cell mixing during spinal cord genesis was addressed in this investigation.

1.4 Statement of the problem

The classic genetic approach to development used so successfully in invertebrates, especially *Drosophila*, has been possible mainly because a large number of genes could be identified on the basis of their mutational phenotypes. This in turn was possible because the *Drosophila* genome could be readily mutated and large scale screening for relevant mutations could be performed. Similar tools have not been available for those interested in mammalian development where most of the critical genes are yet to be found. Numerous strategies have been employed in an attempt to identify genes that play important roles during mammalian development. However,

most of the methods utilized so far have various limitations, largely due to the enormously complicated mammalian genome and inaccessibility of the post-implantation embryos in mammals. In this thesis the gene trap approach (Gossler et al., 1989) was used and appears to overcome part of these long standing problems. In this approach, genes of interest are selected on the basis of their temporal and spatial expression pattern, rather than mutational phenotypes. Since such genes are tagged by the expression of a fusion transcript, this approach also allows them to be readily analyzed at the molecular level. A successful trapping of a actively transcribed gene also indicates a simultaneous creation of an insertional mutation in the form of a fusion transcript. Combined with the ES cell-chimera-transgenic technology, the expression pattern of the trapped genes can be readily determined *in vivo*. Therefore, the gene trap method can be applied to any system to identify genes that may play significant roles during the development of that system. This strategy was used in this thesis in an attempt to screen for genes which, on the basis of their expression pattern, play important roles during nervous system development. A novel mammalian gene was characterized in this thesis using the gene trap approach. This new gene appears to bear many of the characteristics of a mammalian transcription regulator with a specific expression pattern in the nervous system.

In addition to the primary objective of novel mammalian gene identification, chimeras generated by injection of the ES cell clone that expresses a pan-neuronal fusion protein, were used in this investigation to address the question of the pattern of the nervous system development. The controversy surrounding how neuronal precursor cells migrate to their final destination during the development of cerebral cortex has been the focus of intensive investigations. The current view suggests two modes of cell

dispersion: migration following a strict radial path, or widespread tangential dispersion. Retroviral labeling studies (Walsh and Cepko, 1992; 1993) have shown quite convincingly that there exists a widespread tangential cell dispersion during cortical genesis, whereas the radial columns seen in the X-inactivated transgenic mice (Tan and Breen, 1993; Tan et al., 1995) have provided evidence of radial cell migration associated with tangential cell dispersion.

The pattern of spinal cord development has also been a subject of intensive investigations with seemingly contradictory conclusions. Musci and Mullen (1992) used aggregation chimeras to address the question of early cell mixing during spinal cord genesis. They showed that there was a fine-grained mosacism in the developing spinal cord of the chimeras, indicating a extensive random cell mixing during spinal cord development. The results of the retroviral labeling experiments conducted in the chick embryonic spinal cord, however, arrived at a different conclusion. Leber and Sanes (1995) demonstrated that there were apparent longitudinal segments and dorsal-ventral patches of the retroviral labeled cells in the developing chick spinal cord. These segments and cohort patches represent clonally related progeny of a single progenitor cell. Thus, on the basis of this data, cell mixing during early spinal cord development is not at all random. Clonally related cells are distributed in a spatially restricted fashion and there is a great deal of variation in cell distribution during the development of the spinal cord.

The gene trap derived chimeras used in this study provided an alternative approach to test the current model of cortical and spinal cord genesis. Mouse chimeras have provided a useful means to study cell lineage and embryonic patterning during development. This is because each tissue of a chimera, including the nervous system, is composed of two distinguishable

cell populations. The pattern of distribution of the two cell populations allows a retrospective analysis of the extent of early cell mixing and the subsequent pattern of cell migration during organogenesis. Chimeras produced by the gene trapped GT9.6 ES cell clone, in which the fusion transcript encoded by the trapped gene locus is expressed pan-neuronally in the developing and mature nervous system, provided a unique opportunity to investigate the pattern of organization during the development of the cerebral cortex and the spinal cord.

2.1 The Gene trap construct

The gene trap construct pGT4.5 was originally built by Gossler et al. (1989). In brief, a mouse *En*-2 (Joyner and Martin., 1987) genomic DNA clone consisting of a 5' 1.8 kb EcoR I-Sst I fragment, which includes a 150 bp homeobox-containing exon and upstream intron sequences containing the splice acceptor site of the homeobox, was used. This was joined in frame to the 5' end of a LacZ gene deleted of the 24 nucleotide sequence encoding the first 8 N-terminal amino acids. The construct also contains neo^r driven by a human β-actin promoter. The polyadenylation signals for LacZ were derived from SV40, and a 3' 2.7 kb Bgl II-Sst I *En*-2 genomic fragment containing the 3' untranslated region provided the polyadenylation signal for neo^r. The construct has a unique Hind III site in the vector sequence which was used to linearize the construct (Figure 3).

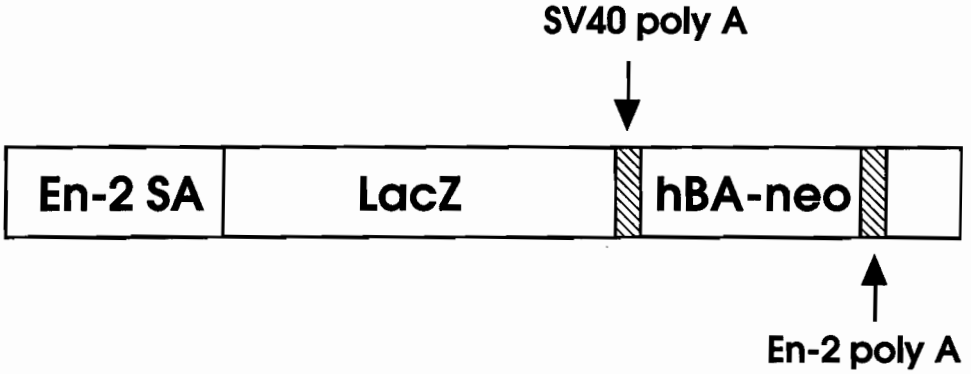
2.2 Embryonic stem (ES) cell culture

The CCE embryonic stem cells (ES cells, passage 9) were derived from XY blastocyst-stage embryos of the 129/Sv/Ev strain. This strain is homozygous for both the black and the agouti alleles (*BB AA*), used as coat color markers of chimerism, and for the $Gpi\text{-}1^c$ allele, used for tissue genotype analysis (Robertson et al., 1986). The wild type agouti allele (*AA*) is dominant and the hair from an agouti and black (*AA BB*) mouse is black with a subapical band of yellow. The non-agouti allele (*aa*) is recessive, and a black and non-agouti (*BB aa*) mouse (e.g. C57Bl/6 inbred strain) is therefore plain black.

Figure 3. Schematic illustration of the PGT4.5 gene trap construct. The PGT4.5 gene trap construct (Gossler et al., 1989) contains a promoterless LacZ gene without its translation initiation codon (ATG). It is ligated at its 5' end to a splice acceptor (SA) sequence derived from the mouse *En-2* gene. This construct also contains a *neo*^{r} gene as a selectable marker. The *neo*^{r} gene is driven by a ubiquitously active human β-actin promoter. The polyadenylation signals for LacZ were derived from SV40, while a 3' 2.7 kb Bgl II-SstI *En-2* genomic fragment containing the 3' untranslated region provided the polyadenylation signal for *neo*^{r}. The construct has a unique Hind III site in the vector sequence and this was used to linearize the construct. In those rare cases in which this gene trap construct inserts in the correct orientation and frame into an intron of a gene actively transcribed in ES cells, a fusion protein with β-galactosidase activity can be expressed.

Gene Trap Construct

pGT4.5



The male ES cell line was used because the XY composition is stable in culture compared with XX lines which tend to suffer deletion or loss of one X chromosome (Robertson 1983). Furthermore, injection of the XY CCE ES cells into blastocysts results in either an appropriate (XY<->XY) or an inappropriate (XY<->XX) combination of genotypes. In the XY<->XX combination the pattern of sexual development of the chimera may be affected. A sex distortion phenomenon that produces more male than female chimeras has been observed for XY<->XX chimeras (McLaren 1976). Where sex conversion has occurred and the resulting chimera is a phenotypically fertile male, it will transmit its ES cell-derived genome only. In embryos with XY<->XX combinations the fertile males show no transmission of genetic markers carried by the XX component, providing strong evidence that XX cells are incapable of forming functional sperm (McLaren, 1976). Males have the added advantage of taking much less time to sire the same number of offspring as females. This is important when a low percentage of germ cells arising from the ES cell line is expected. The sex conversion effect will not occur if the cell line used is XX.

The ES cells were cultured on STO fibroblast feeder layers (Evans and Kaufman, 1981) which were stabely transfected with *pSV2-neo*^{r} (Southern and Berg, 1982) by Ca^{++} phosphate precipitation or lipofection, and mitotically inactivated by γ-irradiation (4000 rads). The ES cell cultures were re-fed with fresh ES cell culture media either daily or according to the acidity of the medium, as indicated by the change of the medium color from purple to yellow. Change in pH is due to the accumulation of the acidic metabolites from the fast growing ES cells. The tissue culture media for ES cells contained DMEM, 10% fetal bovine serum (Gibco), 10% calf serum (Hyclone), 10% LIF supernatant (Genetics Institute), a nucleoside cocktail (Sigma), 0.1 mM β-

mercaptoethanol (Gibco), 0.01 mM non-essential amino acids (Gibco), 2 mM glutamine (Gibco), 50 units/ml of penicillin and 100 µg/ml of streptomycin (Gibco), and the pH for this media is alkaline (Robertson, 1987).

2.3 ES cell transfection and cloning

The CCE ES cells, at passage 13 to 15, were electroporated with the pGT4.5 construct according to Schwartzberg et al. (1989) with some modifications. Cells were plated at a density of $3x10^6$ cells per 10 cm plate, fed every two days, and trypsinized with 0.05 % porcine derived trypsin (Gibco) 2 hours after the second feeding. Cells were counted using an *Improved Neubauer* hemocytometer (Hausser). Briefly, 50 µl of a thoroughly trypsinized cell suspension were collected using a micropipette and mixed well with 50 µl of 0.4% trypan blue dye solution (Sigma). Dye exclusion indicated cell viability. An aliquot of the well mixed cell suspension was transferred immediately to the edge of the hemocytometer chamber covered with a coverslip, and the cell suspension was let run out of the pipette and drawn under the coverslip by capillary action. Cells lying within a square of 1 mm were counted under an inverted light microscope (Leitz Laborlux D) at a magnification of 200x . The cells in four such squares were counted and the average of the four counts were calculated. The concentration of the sample is derived as follows: $c = 2n/v$ where c = cell concentration (cells/ml), n = number of cells counted within a 1 mm^2 field, and v = volume (ml). For the *Improved Neubauer* slide, the depth of the chamber is 0.1 mm, the volume of the cell suspension contained within a 1 mm^2 (v) is thus 0.1 mm^3 or 10^{-4} ml. The number 2 is derived from the dilution with an equal volume of trypan blue dye. The formula is thus: $c = 2n \times 10^4$.

Trypsinized cells were washed twice with phosphate-buffered saline (PBS) which was free of Ca^{++} and Mg^{++}, and resuspended in PBS at a concentration of $4x10^7$ cells per milliliter. Cell suspensions containing $2x10^7$ cells (0.5 ml) were mixed in an electroporation cuvette with 15-25 μg of HindIII linearized pGT4.5 DNA and electroporated at 220 v and 960 μF using a Biorad Gene Pulser. After the pulse, cells were incubated at room temperature for 10 minutes. Cells were then plated at a density of $5x10^6$ onto 10 cm dishes containing STO feeder layers. Cells were fed with ES cell culture media containing 400 μg/ml of G418 the next day, and re-fed by changing the media containing G418 every two days thereafter due to the slower doubling rate of ES cells under drug selection.

A range of 40 to 60 sizable G418 resistant colonies appeared per 10 cm dish 9 to 12 days after G418 selection. Cloning was then carried out in either of the two ways. (1) Colonies were picked and placed into 24 well cloning trays containing 0.05% trypsin. Each of the trypsinized colonies was split into two wells in two different 24 well plates. Duplicate plates representing 24 colonies were thus produced. One day after plating, one of the two duplicate 24 well plates was tested for β-galactosidase activity using a histochemical reaction (Sanes et al., 1986) in which Bluo-gal (BRL) is used as a substrate. In brief, cultured cells were rinsed with PBS (pH 7.3) and fixed for 30 min. at 4°C with 0.2% glutaraldehyde in PBS. The cells were then washed with PBS and overlaid with a histochemical reaction mixture containing 3.1 mM potassium ferrocyanide, 3.1 mM potassium ferricyanide, 1 mM $MgCl_2$ and 0.5 mg/ml Bluo-gal in PBS. The Bluo-gal was dissolved in N,N-Dimethylformamide (DMF) at 20 mg/ml and diluted into the reaction mixture. The cells were incubated in the reaction mixture overnight and the β-galactosidase positive cells were identified as blue staining cells at 200x magnification under an

inverted microscope. The β-galactosidase positive clones were then plated from the replica 24 well cloning tray onto 6 cm plates, and subsequently expanded and frozen to permit further molecular and developmental analysis of the disrupted gene. (2) Colonies in each 10 cm dish were pooled and split into two 10 cm plates, one of the which was tested for β-galactosidase activity. Plates negative for β-galactosidase activity were discarded along with their duplicates. Plates with duplicates containing β-galactosidase positive cells were subjected to subcloning using 96 well cloning plates. Single cell clones were passaged into two wells in duplicate 24 well plates and were subsequently cloned as described above.

2.4 Blastocyst injection and generation of chimeric mice

C57Bl/6, MF1 and CD1 (Charles River) mice were used as the source of blastocysts. An important consideration in the choice of the mouse strain for blastocysts is that the ES cell line being used to construct the chimeras must be easily distinguishable, either visually or biochemically, from the host blastocyst cell population. The MF1 and CD-1 are albino (homozygous for the *c* allele; *cc*) and homozygous for the $Gpi\text{-}1^b$ allele, whereas the C57Bl/6 inbred strain is black, non-agouti (BB, aa) and homozygous for the $Gpi\text{-}1^a$ allele. Thus, The presence of an ES cell derived contribution can be identified by the presence of pigmented coat hair in the MF1 and CD-1 genetic background, and by the presence of agouti hair in those mice derived from injected C57Bl/6 blastocysts.

B6C3 F1 mice were used for foster mothers. F1 hybrids are produced by crosses of two inbred strains of mice (C57Bl/6 X C3H in this case) and are therefore heterozygous at all loci for which the inbred strains are homozygous for different alleles. Thus, the F1 hybrids are genetically as

uniform as the mice of the inbred strains. Furthermore, the heterozygosity allows the F1 hybrids to maintain favorable alleles not held in common by the two parent population, a phenomenon known as heterosis or hybrid vigor. Therefore, the F1 hybrids are more tolerant to environmental changes than inbred strains. They grow faster, survive to maturity in greater proportions, live longer, and in turn reproduce earlier and more abundantly. Hybrid litters are usually larger than inbred litters (Green, 1966). F1 hybrids have proved to be better foster mothers than either inbred strains or outbred stock. Eighty to ninety percent of operated F1 mothers maintained pregnancy compared to the less than 50% pregnancy rate obtained in other strains. Moreover, F1 females generally make better mothers after delivery of the pups and the majority of operated embryos that come to term survive to weaning age and give healthy, fertile animals (Robertson, 1987).

Blastocysts were obtained from the pregnant females (C57Bl/6, MF1, CD1) by timed matings. The time of natural mating was controlled by the light cycle of the mouse room. The controlled regime of the light cycle is 12 hour light to 12 hour dark, and females pass through the oestrous cycle every 4-5 days. Ovulation of oestrous females occurs 3-5 hours after the onset of the dark period and the males will copulate with oestrus females at about the middle of the dark cycle. The oestrus host females over six weeks old were selected by examining the color, moistness and degree of swelling of the vagina shortly before the onset of the dark cycle. The selected females were caged singly with one male. The following morning each female was checked for the evidence of mating as indicated by the presence of a solid seminal copulation plug in the vagina (including vasectomized matings for foster mothers as described below). The plugged females were designated as day 0 post coitus. The day 3.5 blastocysts were obtained from the plugged females

according to the following procedure. (1) Day 3.5 or day 2.5 pregnant females were sacrificed by cervical dislocation, and uterine horns and oviducts on both sides were removed and placed on a depression dish. (2) Uterine horn and oviduct on either side were flushed with Brinster's BMOC medium (Gibco). The day 3.5 blastocysts were recovered from the uterine horns, and the day 2.5 morulae from the fallopian tubes. (3) The blastocysts and morulae were located in a depression dish using a stereo dissecting microscope (Wild M5A) and removed with a mouth controlled drawn-out pasteur pipette and placed in micro-drops of Brinster's BMOC medium under light weight paraffin oil (Fisher). Day 2.5 morulae were incubated at 37°C and 5% CO_2 in a micro-drop culture of Brinster's BMOC medium overnight during which time they develop into fully expanded blastocysts.

ES cells at exponential growth phase were used for blastocyst injection. All available cultures were inspected and a cell suspension from a culture which displays a healthy cellular morphology was prepared. Cells were trypsinized, and although every effort was made to obtain a single cell suspension, doublets and triplets were commonly seen in the cell suspension. Using a mouth controlled drawn-out pasteur pipette a small volume of the freshly prepared cell suspension was transferred into the injection chamber. The injection chamber contained blastocyst injection medium which is DMEM supplemented with 20 mM HEPES and 10% FBS. The medium was added to a depth of 2-3 mm. A layer of paraffin oil was overlaid on the surface of the medium and the chamber was placed onto the microscope (Leitz Labovert FS) stage. The fully expanded blastocysts (8 to 10 each time) were collected into the same chamber which is connected to a cooling device (Technology Inc., NJ) to maintain the chamber temperature at 10°C. Cooling gives the cell membranes a degree of rigidity, providing appropriate resistance

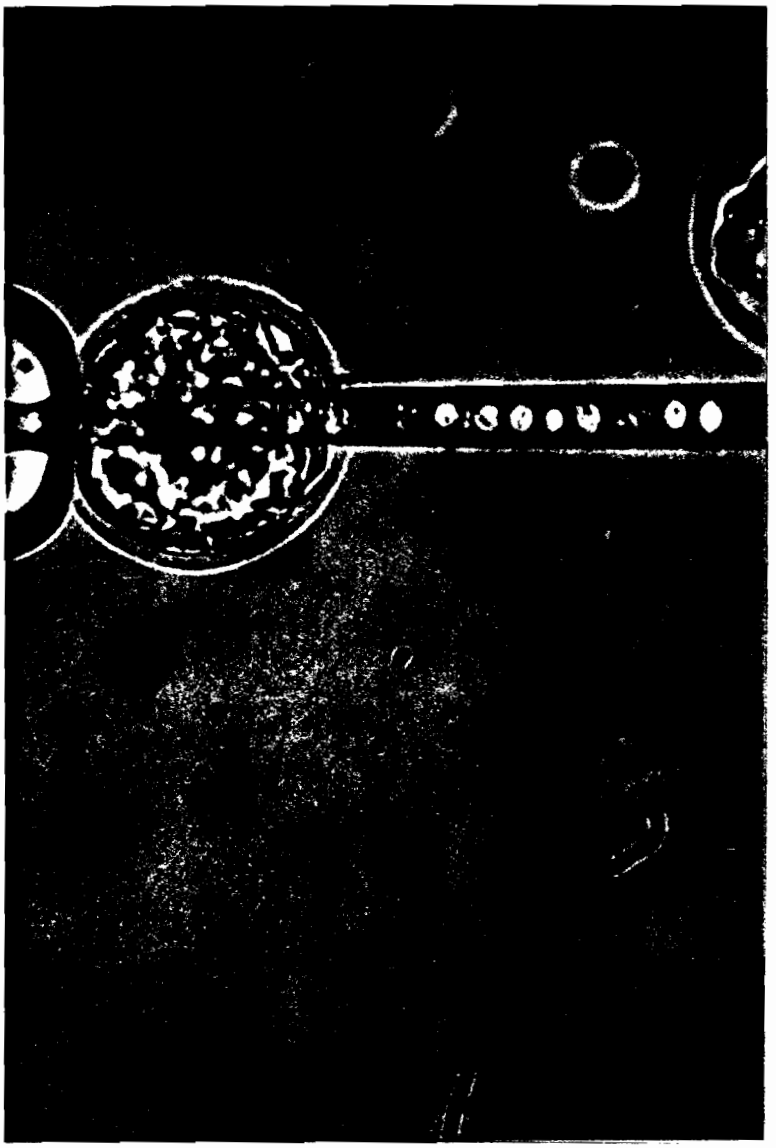
that allows easier penetration of the injection pipette. It also helps to prevent stickiness associated with cell lysis which necessitates the frequent changing of the injection instruments.

Between 8-12 cells were injected into the blastocoel cavity of each day 3.5 blastocyst. The cells (8-12) were selected individually and were collected into the injection pipette. This was then pushed through the zona pellucida and trophoblast layers into the blastocoel cavity of a blastocyst immobilized by suction on the end of the holding pipette. The cells were expelled into the blastocoel, and pipette was withdrawn and the blastocyst released from the holding pipette.

The glass capillary tubing used to construct the pipettes is thin walled (1.0 mm external diameter, 0.8 mm internal diameter) borosilicate glass without a fibre (Clark Electromedical). They were pulled using an automatic electro-magnetic electrode puller. The holding pipettes were constructed by blunt flame-polishing using an Beaudouin microforge. The injection pipettes were built by snapping the glass capillary by hand under a stereo dissecting microscope, using a sharp scalpel blade and a spongy transparent silicon rubber sheet. A 30° bend was introduced into both the holding and injection pipettes by heating the bending point on the pipettes with a flame-polishing filament using KOPF microforge. The bend enables the injection and holding pipettes to be moved freely in the injection chamber in the plane of focus

After injection, the blastocoel cavity of the blastocyst collapses. The microinjected blastocysts were cultured in Brinster's BMOC medium for 30 minutes and were transplanted immediately. In our experience, the re-expansion of the injected blastocysts before transfer was not necessary and often takes hours. The injected blastocysts that displayed lysis were discarded, and the remaining injected blastocysts were surgically transferred to the

Figure 4. Photograph illustrating mouse blastocyst injection with ES cells. A mouse blastocyst, immobilized by a holding pipet, is perforated by a microinjection pipette containing ES cells. Such manipulations were done in a microinjection chamber connected to a fix-staged microscope (Leitz). Eight to 12 ES cells were collected into the injection pipet and were expelled into the blastocoel cavity. The injected blastocysts were then transplanted into the uterus of a pseudopregant female and allowed to develop to term.



uterine horns of day 2.5 pseudopregnant females. Pseudo-pregnancy was induced by mating B6C3 F1 females to vasectomized B6C3 F1 males. Some litters were delivered naturally 17 days after injection, and others were delivered by caesarian section after 17 days. Only MF1 blastocysts were used to obtain day 12.5 chimeric embryos from pregnant recipients sacrificed at day 12.5. Chimerism was scored by eye pigmentation in MF1 albino background at embryonic day 12.5, or by coat pigmentation in MF1 and CD1 albino background 4-5 days after birth, and by the presence of agouti coat color in the C57Bl/6 background on postnatal day 10.

2.5 Germline transmission of chimeras

Mouse germ cells differentiate from cells of the primitive ectoderm and migrate from their origin to the genital ridges. The migration involves some active movement of the germ cells, either along tracts of extracellular matrix material or in response to chemotactic substances released by the cells of the genital ridge. The genital ridge of the mouse embryo is also derived from the primitive ectoderm, and in XX/XY chimeras, formed either as a result of aggregating morulae or by blastocyst injection, it has a profound influence on germ cell differentiation. XX germ cells in a predominantly XY local environment begin to develop in the male direction but do not form sperm, whereas XY cells in an XX environment can form oocytes (Hogan, 1986).

Morphological differentiation of the sex organ is caused by a localized effect and is dependent on the relative contributions of XX and XY cells to the fetal genital ridge. The pattern of integration of XY ES cells in a female embryo affects sexual differentiation and determines the fate of the two cell populations in terms of their ability to form functional germ cells. The ability

of the injected ES cells to form sperm is governed by the sex of the host embryo, and the number and position of the colonizing cells in the early embryo. In the appropriate XY<->XY combination, cells from one of the two strains that make up the chimera contribute to, at most, 3% of the sperm (Mclaren, 1976). This low level may reflect the ratios of cultured cells to host cells within the gonad or might result from the fact that in some embryos with XY:XY strain combinations, cells of one genotype may predominate within the sperm population. This has been reported for the dominance of C3H germ cells in C3H<->C57Bl/6 chimeras (McLaren, 1976). It seems therefore, at least in some strain combinations, that one or the other cell line is often excluded from the germ cell population. What is more striking is that in some male C3H<->C57Bl chimeras, no relation is found between the strain composition of the germ cells and that of the somatic tissues: in some male chimeras the germ cells were all from one component, and the somatic tissue largely from the other (McLaren, 1976). In the inappropriate XY<->XX combinations three possibilities exist: (1) Low incorporation in which insufficient Y-bearing cells are present to influence the sex phenotype. The resulting chimeras are fertile females with XY cells able to form functional ova, but, in the predominantly XX environment, incapable of producing functional sperm. (2) A higher level of incorporation in which the Y chromosome can exert a masculinizing influence on the differentiation of the gonad. However, the XX cells from the host are unable to form functional sperm. Therefore, a small percentage of the resulting animals have been observed to be sterile hermaphrodites with the appearance of both sex organs (McLaren 1976). (3) A high contribution of XY cells to the gonad above the threshold causes the differentiation of a functional male reproductive system.

These animals are phenotypically fertile males and produce a single class of spermatozoa derived from the XY ES cells (Robertson, 1986).

To test the ability of the unmanipulated ES cells to colonize the germ cells of the chimeras, chimeric males from C57B1/6 and CD1 host background were set up to breed at 6 weeks of age with C57B1/6 and CD1 females, respectively. For gene trapped ES cell clones, chimeras from the C57B1/6 and MF1 host background were set up with their respective sex partners. Germline transmission was identified by the presence of agouti pigmentation in the coat of F1 animals 10 days after birth.

2.6 β-Galactosidase histochemistry

2.6.1 Mouse perfusion and tissue preparation

Adult chimeric mice were anaesthetized with Avertin (8 mg/kg) and perfused transcardially with 30 ml cold 0.5% paraformaldehyde, 2.5% glutaraldehyde in 0.1 M PBS pH 7.2, using a 25-gauge butterfly needle. Tissues were subsequently dissected out and immersed in the same fixative at 4^0C for one hour. Day 12.5 embryos were obtained by caesarean section at 12.5 days post-coitum. Embryos were removed from the uterus and dissected free of fetal membranes under 0.1 M PBS pH7.2. Embryos were fixed by immersion in a mixture of cold 0.5% paraformaldehyde, 2.5% glutaraldehyde in 0.1 M PBS pH 7.2 for 2 hours at 4°C. Following fixation, incisions were made along the sagittal plane of the embryonic brain to increase penetration of stain.

2.6.2 Bluo-gal histochemistry

Bluo-gal (Bethesda Research Laboratories, BRL) was dissolved in N,N-Dimethylformamide (DMF) at 20 mg/ml and diluted into the reaction mixture. Following a brief rinse in PBS, adult or embryonic tissues were

stained by incubating at 37^{0}C overnight with Bluo-gal, 0.4 mg/ml in PBS containing 10 mM potassium ferricyanide, 10 mM potassium ferrocyanide and 2 mM magnesium chloride. Stained embryonic brain and spinal cord were dissected out from the embryos. The β-galactosidase positive cells in both the embryonic and adult tissues were identified as blue staining cells at 25x magnification under a dissecting microscope (Leitz, Wild M5A).

To obtain 12 μm thick cryostat sections, mouse adult or embryonic brain and spinal cord were removed and transferred to 30% sucrose in 0.1 M phosphate buffer pH 7.3 for an extra 2 to 3 days. Tissue samples were embedded in 15% gelatin and the blocks were trimmed, frozen in isopentane in liquid nitrogen and stored under isopentane at -80^{0}C until sectioning. Twelve μm thick sections were cut in a Leitz cryostat (MGW Lauda) at -20^{o}C and each section was mounted onto a TES (3-aminopropyltriethoxy-silane, Sigma) coated slide.

2.7 β-Galactosidase immunocytochemisty

2.7.1 Mouse perfusion and tissue preparation

Adult chimeric mice were anaesthetized with Avertin (8 mg/kg) and perfused transcardially with 10 ml of cold (4^{o}C) PBS pH 7.2, followed by 100 ml periodate-lysine-paraformaldehyde (2% paraformaldehyde, 0.1 M lysine and 10 mM $NaIO_4$ in 0.1 M phosphate buffer, pH 7.2) (Mclean and Nakane, 1974). Mouse tissues were removed and immersed in the same fixative at 4^{o}C for five hours, and were then transferred to 30% sucrose in 0.1 M phosphate buffer pH 7.3 for an extra 2 to 3 days. Tissue samples were embedded in 15% gelatin and the blocks were trimmed, frozen in isopentane in liquid nitrogen and stored under isopentane at -80^{0}C until sectioning.

2.7.2 Mouse β-galactosidase antibody

The mouse β-galactosidase antiserum directed against the β-galactosidase protein was kindly supplied from the Hybridoma Laboratory of the Canadian Network of Center of Excellence (NCE) Neural Regeneration and Functional Recovery Core Facilities by the late Dr. A. Côté. Both the primary antiserum and secondary antibodies were diluted in 0.5 M Tris-HCl buffer, pH 7.6 to which 1% normal horse serum (NHS) was added. The primary mouse anti-β-galactosidase antiserum was diluted at 1:1000, while the secondary biotinylated horse anti-mouse IgG (Vector) was diluted at 1:200.

2.7.3 Immunocytochemistry

Twelve µm thick sections were cut in a Leitz cryostat (MGW Lauda) at -20°C and each section was mounted onto a gelatin coated slide. A drop of cold formal sucrose fixative (7.5% sucrose and 4% formaldehyde in 0.07 M phosphate buffer, pH 7.4) was deposited onto each slide and sections were incubated in the fixative for 30 minutes. Following a 30 minute rinse in 0.5 M Tris-HCl (pH 7.6), non-specific protein binding sites were blocked by incubating the sections in 7% normal horse serum (NHS) at room temperature for 30 minutes. The sections were incubated overnight at room temperature in the primary mouse anti-β-galactosidase antibody. After rinsing with 0.5 M Tris HCl for 30 minutes, sections were incubated in the secondary biotinylated horse anti-mouse IgG (1:200) for an hour. Following another 30 minute rinse with 0.5 M Tris-HCl, sections were incubated in Avidin-Biotinylated HRP Complex (ABC) (Vector) for 1 hour. HRP activity was detected using 1 mg/ml diaminobenzidine (DAB) with 0.03% H_2O_2 in 0.1 M Tris-HCl, pH 7.6, for 10 minutes. The sections were then rinsed in distilled water and mounted in Farrant's medium (BDH).

2.8 Molecular biology methods

2.8.1 DNA Probes

The LacZ probe, a 119 bp RACE-produced probe from endogenous mouse genomic sequence and two other cDNA probes were used for Southern and Northern blot analysis, as well as for cDNA library screening. The LacZ probe was a 3.1 kb full length *E coli.* LacZ gene sequence, whereas the 119 bp endogenous probe contains the sequences 5′ of the lacZ splice acceptor site produced from the 5′ RACE-PCR performed for the ES cell clone GT9.6. The RACE-PCR product was cloned into a TA PCR cloning vector (Invitrogen) and sequenced. A 5′ primer (5′ AGGAATTCTAAGTATTGACATTACACAG 3′) and a 3′ primer (5′ ACCTGGAATTCGGAGCGAGTAGTAGA 3′) both containing a newly created EcoR I site (underlined) with the 119 bp RACE-PCR product, were used to PCR amplify the 119 bp endogenous sequence. This 119 bp PCR product , along with the 3.1 kb LacZ gene sequence and the two cDNA probes, were used as templates for random priming (Prime-it Kit, Strategene) to produce radioactive probes.

The 735 bp cDNA probe was obtained from cDNAs cloned from the screening of a mouse embryonic day 12.5 cDNA library, whereas the 1574 bp cDNA probe was cloned from cDNAs recovered from the screening of a mouse postnatal day 8 brain cDNA library. The nucleotide sequences of the 119 bp RACE-PCR product, the 735 cDNA clone and the 1574 bp cDNA clone are shown in Figures 23, 27, and 32, respectively.

2.8.2 Northern blot

2.8.2.1 Isolation of total RNA from ES cells and mouse tissues

Total RNA was isolated by the method of Chomczynski and Sacchi (1987) with few modifications. Mice were killed by cervical dislocation, tissues removed and frozen immediately in liquid nitrogen. ES cells or tissues were lysed in solution H. Solution H was prepared by sequentially adding the solutions listed below to a stock solution consisting of 25 mM sodium citrate, pH 7.0, 0.5% w/v N-Lauryl-sarcosine, 4 M guanidium thiocyanate, and mixing after each solution was added (volumes are per 100 ml stock solution): 0.72 ml β-mercaptoethanol, 10 ml 2 M sodium acetate pH 4.0 and 100 ml diethyl pyrocarbonate (DEPC) H_2O saturated phenol.

Tissue culture media was removed from 150 mm ES cell plates and the cells were rinsed twice with PBS pH 7.0. For each 150 mm plate, 5 ml of solution H was added and cells were scraped from the plate using a disposable plastic scraper (Nunc) and transferred to a 50 ml polypropylene tube. The frozen mouse tissue was weighed and added to 1 ml of solution H per 100 mg tissue. ES cells and mouse tissues in solution H were homogenized with a polytron (Brinkmann Instruments) until well mixed. The following steps were performed with the samples on ice: To every 5 ml of homogenized solution, 1 ml chloroform was added and the mixture was shaken vigorously for 1 minute. Following a 15 minute incubation on ice, the mixture was centrifuged at 12,000*g* for 20 minutes at 4^0C and the aqueous phase was collected. One volume of isopropanol was subsequently added. The solution was mixed by inversion and placed at -20^0C for one hour, and the RNA was collected by centrifugation at 12,000*g* for 20 minutes. The RNA pellet was washed with 70% ethanol and dissolved in DEPC treated H_2O and frozen at -80^0C.

2.8.2.2 Poly A$^+$ RNA isolation

Poly A^+ RNA was isolated by the method of Aviv and Leder (1972) with some modifications. Total RNA was diluted to 0.4 µg/µl and incubated at 65^0C for 10 minutes, cooled on ice, and brought to a final concentration of 0.5 M $LiCl_2$. Fifty mg of oligo-dT cellulose (Collaborative Research Inc.) per mg of total RNA were resuspended in elution buffer (10 mM Tris-HCl pH 7.5 at 20^0C, 1 mM EDTA, 0.05% SDS). This suspension was poured into a 1 ml disposable syringe loaded with glass fiber and the RNA was applied. The column was washed with more than 10 volumes of binding buffer (20 mM Tris-HCl pH 7.6, 1 mM EDTA, 0.1% SDS, 0.5 M $LiCl_2$) until the eluent A_{260} was below 0.1. Elution buffer was heated to 65^0C and 4 column volumes were used to elute the Poly A^+ RNA. The eluted RNA was precipitated by adding 1/10 volume of 3 M NaOAc and 2 volumes of 100% ethanol. The precipitated RNA was collected by centrifugation at 10,000*g* for 15 minutes at 4^0C, washed with 70% ethanol, air dried, and dissolved in a minimum of DEPC treated H_2O.

2.8.2.3 Formaldehyde Gel Electrophoresis of RNA

All RNA gels were run in an IBI Medium Gel apparatus suitable for up to 10 samples. Typically, a 75 ml formaldehyde agarose gel was loaded onto each apparatus. To prepare a 100 ml 1.2% formaldehyde agarose gel, 1.2 g of agarose was boiled in 74 ml ddH_2O and cooled to 65^0C. Sixteen ml of formaldehyde and 10 ml of 10X running buffer (2 M 3-[N-morpholino]propanesulfonic acid (MOPS) pH 7.0, 5 mM sodium acetate (NaOAc), 1 mM EDTA), each prewarmed to 65^0C, were added. This solution was stir-mixed and poured into the gel tray. RNA samples (up to 20 µg total RNA) in 4.5 µl DEPC H_2O were diluted with 2.0 µl 10X running buffer, 3.5 µl of formaldehyde and 10 µl of formamide. The samples were incubated at 65^0C

for 15 minutes and 2 µl of loading buffer (50% glycerol, 1 mM EDTA, 0.4% bromophenol blue) were added to each sample prior to their application onto the gel. The samples were electrophoresed overnight in 1X running buffer at 30V. The RNA was stained with 0.5 µg/ml ethidium bromide in TAE (40 mM Tris-acetate, 1 mM EDTA) for 2 hours followed by destaining in ddH_2O for 2 hours. The RNA was visualized under short wave ultraviolet light and photographed.

2.8.2.4 RNA blotting

The formaldehyde agarose gel was blotted onto a positively charged nylon membrane (Boehringer Mannheim) in 5X SSC solution for 2 hours using a vacuum blotting apparatus (Tyler Research Instruments). Successful blotting was checked by visualizing the RNA on the filter using a portable UV light source. The filter was then UV cross-linked in a UV Stratalinker 2400 and stained with 0.02% methylene blue. To carry out RNA staining, the filter was rinsed briefly in 1X SSC, and stained in 0.3 M NaOAc and 0.02% methylene blue for 3 minutes without shaking. The filter was then dipped in 1X SSC very briefly, and ribosomal RNAs were visualized, marked and photographed under white light. To remove the stain, the filter was washed in 2X SSC containing 10% SDS at room temperature for 15 minutes with shaking. RNA blot of P19 and RAC65 cells was generously provided by Stacy Costa in Dr. M. McBurney's laboratory at University of Ottawa.

2.8.2.5 Nucleic Acid Hybridization

The filter was sealed in a bag and all buffers were introduced and removed through a corner of the bag which was cut when necessary and re-sealed. The filter was prehybridized in 15 to 25 ml of prehybridization buffer

containing 0.5 M sodium phosphate buffer pH 7.2, 1 mM EDTA, 1% Bovine serum albumin (BSA), 7% SDS and 200 µg/ml denatured salmon sperm DNA (ssDNA) at 65^0C for a minimum of one hour. The prehybridization buffer was removed and replaced with an equal volume of hybridization buffer containing 30% formamide, 0.2 M sodium phosphate buffer pH 7.2, 1 mM EDTA, 1% BSA, 7 % SDS, 200 µg/ml of denatured ssDNA and $1.5x10^7$ to $9.0x10^7$ CPM ^{32}P-labeled denatured DNA probe. The filter was hybridized at 65^0C for 12 to 16 hours. The filter was removed from the bag and washed once with 2X SSC, 0.1% SDS for 30 minutes at room temperature, and with 0.1X SSC, 0.1% SDS for 30 minutes at 65^0C. The filter was wrapped in plastic Saran wrap and autoradiographed at -80^0C for 24 to 48 hours. Autoradiography was conducted in light-proof cassettes by exposure to X-Omat AR film (Kodak) with the use of Cronex Quanta III intensifying screens.

2.8.3 Southern blot

2.8.3.1 Isolation of genomic DNA from ES cells and tissues

To isolate genomic DNA from ES cells, trypsinized ES cells from each 150 mm plate were washed with 5 ml PBS pH 7.3 three times and resuspended in 1 ml PBS. Rat and quail genomic DNA, which were used for cross-species Southern blot hybridization, were extracted from rat tail and quail muscle. Samples of rat tail and quail muscle (kindly provided by Dr. K. Hastings, Department of Molecular Genetics, Montreal Neurological Institute) were ground manually to small pieces by mortar and pestle. Cell suspensions and ground tissue pieces were then transferred to a 50 ml falcon tube and digested with 200 µg/ml Proteinase K dissolved in solution containing 0.1 M EDTA, 10 mM Tris-HCl pH 8.0, 0.1 M NaCl, 0.5% SDS at 55^0C for 3 hours with gentle agitation. An equal volume of neutralized phenol was

added to the solution, and the mixture was shaken at room temperature for 30 minutes. Ten volumes of ethanol were added and the solution was mixed by gentle inversion for 10 minutes at room temperature until a visible DNA precipitate appeared. The DNA filamentous precipitate was removed using a bent Pasteur pipette and was transfered into 20 ml 100% ethanol twice, 5 minutes each time. Following a brief air-dry, the DNA pellet was resuspended in 200-400 µl ddH_2O.

Genomic DNA of fruitfly, zebrafish, and snake were kindly supplied by Dr. R. McGowen (Department of Zoology, University of Manitoba), and genomic DNA of human, yeast and nematode were generously provied by Dr. D. Rosenblatt (Department of Medical Genetics, McGill University), Dr. H. Bussey (Department of Biology, McGill University), and Dr. Martin Chalfie (Columbia University), respectively.

2.8.3.2 Southern blot analysis

Ten µg of cellular DNA was digested to completion with an excess of Kpn I restriction enzyme in the supplied buffer (Boehringer Mannheim). The digests were precipitated with ethanol, resuspended in water and electrophoresed in a 1% agarose gel at 30V overnight. The gel was blotted using a vacuum blotting apparatus (Tyler Research Instruments) onto a positively charged nylon membrane (Amersham) at a negative water pressure of 20 to 25 inches. The surface of the gel was flooded with 25 ml of 0.25 M HCl and held under vacuum for 5 minutes to nick the high molecular weight DNA. This solution was removed and 25 ml of 0.4 M NaOH was applied in the same manner and the gel incubated for 5 minutes to achieve neutralization. This solution was removed and the gel was submerged in 0.4 M NaOH to twice the height of the gel and transferred for 1 hour. The

membrane was removed and briefly rinsed in 2X SSC and UV cross-linked as described in the RNA protocol.

The membrane was prehybridized in 15 to 25 ml of hybridization buffer (50% formamide, 10% dextran sulfate, 1% SDS, 1 M NaCl, 200 µg/ml ssDNA) at 42^{0}C for a minimum of one hour. The hybridization buffer was removed and replaced with an equal volume of fresh hybridization buffer containing 1.5x10^7 to 8.5x10^7 CPM ^{32}P dCTP-labeled denatured DNA probe. The filter was hybridized at 42^{0}C for 12 to 16 hours. The filter was removed from the bag and washed once with 2X SSC, 0.1% SDS for 30 minutes at room temperature, and with 0.1X SSC (0.5X SSC for cross species Southern blot), 0.1% SDS for 30 minutes at 65^{0}C. Subsequently, the filter was wrapped in plastic Saran wrap and autoradiography was conducted in light-proof cassettes by exposure of the wrapped blots to X-Omat AR film (Kodak) with the use of Cronex Quanta III intensifying screens at -80^{0}C for 24 to 48 hours.

2.8.4 RACE-PCR

Rapid amplification of cDNA ends (RACE) was performed by the method of Frohman (1990) with some modifications (Figure 23). All oligodeoxyribonucleotide primers used were chemically synthesized using the "Applied Biosystems 380B DNA synthesizer" according to the manufacturer's instructions.

2.8.4.1 Reverse transcription

Poly A$^+$ RNA was isolated from ES cells and 1 µg of the poly A$^+$ RNA was used to synthesize cDNA in a 20 µl reaction. Reverse transcription mixture, containing 2 µl of 10X RTC (reverse transcription component) buffer [500 mM Tris-HCl pH 8.15 at 41^{0}C, 60 mM MgCl, 400 mM KCl, 10 mM

Dithiothreitol, dNTPs (Pharmacia) 10 mM each], 0.25 µl (10 units) of RNasin (Promega Biotech), and 0.5 µl (1 pmol) gene specific primer, was first mixed on ice. One µg of poly A^+ RNA in 16.5 µl of DEPC H_2O was heated to 65^0C for 3 minutes, cooled rapidly on ice and applied to the reverse transcription mixture, to which was added 10 Units of avian myeloblastosis virus reverse transcriptase (AMV-RT, Life Sciences). The reaction was incubated at 42^0C for 2 hours and subsequently diluted to 1 ml with TE (10 mM Tris-HCl, pH 7.5, 1 mM EDTA) and then stored at 4^0C.

2.8.4.2. 5'-end tailing

The excess primer from 5' end reverse transcription was removed from the cDNA product by spin filtration using Centricon 100 (Amicon Corp.). The filtration step was repeated once with 0.2X TE, and the retentate was collected and concentrated to 10 µl using speedvac centrifigation.

The cDNA was tailed by adding, to 10 µl of Centricon-purified cDNA, 4 µl 1 mM dATP, 4 µl 5X tailing buffer (BRL) (1X tailing buffer = 100 mM potassium cacodylate pH 7.2, 2 mM $CoCl_2$, 0.2 mM DDT), and 10 Units terminal deoxynucleotide transferase (TDT) (BRL). The reaction was incubated at 37^0C for 5 minutes, followed by heat treatment at 65^0C for 5 minutes. The reaction mixture was then diluted to 500 µl with ddH_2O.

2.8.4.3 First round of amplification

The tailed cDNA was amplified as follows: A 50 µl reaction containing RACE-PCR cocktail (50 mM KCl, 10 mM Tris-HCl, pH 8.3, 1.5 mM $MgCl_2$, 0.001% gelatin), dNTPs each at 0.2 mM, 2 pmols (dT)17-R_0 primer, 25 pmols gene-specific primer 1 (GSP_1), and 10 µl tailed cDNA was placed at 95^0C for 7 minutes and cooled to 72^0C. Taq polymerase (2.5 units, BRL) was added and

overlaid with 30 µl of light mineral oil. The reaction was annealed at 50^0C for 2 minutes and the cDNA extended for 40 minutes at 72^0C. The mixture was amplified using a Perkin-Elmer-Cetus DNA Thermal Cycler with the following parameters: 94^0C for 45 seconds, 50^0C for 25 seconds, and 72^0C for 3 minutes for 30 cycles, followed by a 25 minute final extension at 72^0C.

2.8.4.4 Second round of amplification

The first round amplification products were diluted 1:20 with ddH_2O, and 1 µl of diluted material was amplified using primers GSP_2 and R_1 as described above, with the exception that the two minute annealing and 40-minute extension period were eliminated. The PCR amplified fragments were subcloned into a TA cloning vector using a TA cloning Kit (Invitrogen).

2.8.4.5 Detection of the specificity of the RACE-PCR reaction

The PCR products were subjected to agarose gel electrophoresis. The gel was stained with ethidium bromide and subjected to Southern blot, without acid nicking, onto a positively charged nylon membrane (Boehringer Mannheim). The membrane was hybridized with a [γ-^{32}P]ATP end labeled oligo probe (GSP3, Figure 23, 24), synthesized from sequences internal to the primers used in the RACE-PCR reaction. To end-label the oligo probe, oligonucleotides were phosphorylated in a reaction mixture containing 50 mM Tris-HCl pH 7.6, 10 mM $MgCl_2$, 5 mM dithiothreitol (DTT), 1 mM spermidine and oligonucleotides at a concentration of 250 µg/ml. Fifty µCi of [γ-^{32}P]ATP (5000 Ci/mmol, Amersham) and unlabeled ATP (final concentration 66 mM), 7.5 units of T4 polynucleotide kinase (Pharmacia) were added to the reaction mixture and the reactions incubated at 37^0C for 30

minutes. The end-labeled oligo probe was added to the hybridization mixture as described for Southern blot analysis.

2.8.5 cDNA library screening and cloning

2.8.5.1 cDNA libraries

A λgt11 random-primed cDNA library, kindly supplied by Dr. Mark Hanks (Mount Sinai Hospital, Toronto), was constructed using an Amersham kit from poly(A)$^+$-selected day 12.5 embryonic CD-1 mouse RNA. A λgt11 oligo-dT primed cDNA library constructed from postnatal day 18 NIH Swiss mouse brain was obtained from American Type Culture Collection (ATCC 37431). Both cDNA libraries have been amplified.

2.8.5.2 Propagation of Virus

E. *Coli* Y1090 or Y1088 cells were the host for λgt11. A single bacterial colony was used to inoculate 50 ml LB broth supplemented with 0.2% maltose and 10 mM $MgSO_4$ and grown at 37^{0}C until the culture reached an A_{600} of 0.5. Bacterial cells were spun down and resuspended in 10 mM $MgSO_4$ in the same volume (A_{600}=0.5), or in half of its original volume (A_{600}=1.0) for the purpose of phage DNA isolation. A volume of 1.8 ml and 600 µl of cells (A_{600}=0.5) were needed for each 20 cm x 20 cm and 150 mm plate, respectively. 20 cm x 20 cm square plates (Biomed) were used for primary screening, while 150 mm circular plates (Fisher) were used for secondary and tertiary screening. The phage titer was determined by inoculating each 200 µl of Y1090 cells (A_{600}=0.5) with 1 µl of phage solution serially diluted with the SM buffer (1 to 10^{12}). The mixture was incubated at 37^{0}C for 15 minutes, and then added to 3 ml NZY (Gibco) top agarose at 45 to 50^{0}C and plated on a 100 mm NZY plate. The plates were incubated at 37^{0}C overnight and the number of phage

plaques were subsequently counted. The plaque forming units (PFU) were determined as per milliliter of concentration of the cDNA libraries based on the dilutions. For primary screening, aliquots of the library suspension containing 500,000 PFU were mixed with 1.8 ml of cells (A_{600}=0.5) and incubated at 37^0C for 15 minutes. This was added to 50 ml NZY (Gibco) top agarose at 45 to 50^0C, mixed by briefly vortexing, and immediately poured and spread evenly onto a 20 cm x 20 cm square NZY agar plate. The poured plates were left at room temperature for at least 30 minutes to allow the top agarose to solidify, and incubated at 37^0C overnight upside down to avoid cross contamination of phage plaques that could result from water smearing across the surface of the NZY plates due to the condensation effect. For secondary and tertiary screening, appropriate phage titres were used to infect bacterial cells (A_{600}=0.5) so that the phage plaques formed were well separated from each other.

2.8.5.3 Overlay and Hybridization

Bacteriophage plates containing phage plaques were overlaid with appropriately sized Hybond-N (Amersham) square or circular nylon hybridization filters and holes were asymmetrically pricked in them with a 20 gauge needle containing China Ink so that orientation could be subsequently determined. Filters were carefully removed and submerged for 2 minutes in 1.5 M NaCl and 0.5 M NaOH for denaturation, and for 5 minutes in 1.5 M NaCl and 0.5 M Tris-HCl (pH 8.0) for neutralization. After a brief rinse with 2X SSC and blotting on Whatman 3MM paper, the filters were UV cross-linked with a Stratalinker 2400 and stored in 5X SSC at 4^0C. Replica filters of each plate were made in the same manner.

Hybridization for 20 cm x 20 cm square filters and for 150 mm circular filters was carried out in 20 cm x 20 cm square plates (Biomed) and glass dishes (Fisher), respectively. The prehybridization and hybridization conditions were identical to that described for Southern blot analysis, except that 100 ml of the hybridization solutions were used for up to twenty 20 cm x 20 cm square filters or twenty 150 mm circular filters. Following hybridization, the filters were washed once with 2X SSC, 0.1% SDS for 15 minutes at room temperature, and with 0.1X SSC, 0.1% SDS for 30 minutes at 65^0C. The filters were briefly dried with paper towel and radioactive ink was placed asymmetrically on the filters for subsequent orientation. They were then wrapped in plastic wrap and autoradiography was conducted in light-proof cassettes by exposure to X-Omat AR film (Kodak) at -80^0C for 48 hours to 5 days with the use of a single Cronex Quanta III intensifying screens.

2.8.5.4 Isolation of positive clones

The autoradiogram was aligned with the plate, positive clones were identified and plaques removed using the end of a sterilized Pasteur pipette. Each plaque was placed in 1 ml sterile SM (Maniatis, appendix A) and vortexed. For each positive clone, new 150 mm plates containing fewer than 200 PFU were produced and screened (secondary screening) as above. Single positive plaques were isolated and the same procedure described above was followed until all plaques hybridized with the probe indicating that they contained a homologous insert.

2.8.5.5 Isolation of DNA from recombinant phage

Recombinant phage DNA was isolated by the method of Bellomy and Record Jr. (1989) with some modifications. Phage plaques were isolated as a

plug with a Pasteur pipette and placed in 1 ml of SM. The phage was allowed to elute from the plug for at least 2 hours at room temperature. The starting ratio of phage to bacteria is the most critical variable affecting the ultimate yield of phage. Ideally, the ratio of phage to bacteria is such as to allow the phage one final round of replication just before the carrying capacity of the media is reached, as evidenced by the abundance of the lysed bacterial debris. Ten to 100 µl of eluted phage were added to 500 µl of prepared host bacteria (A_{600}=1.0) and allowed to preabsorb at 38^{0}C for 30 minutes. To ensure that one obtained an appropriate ratio of phage to cells to yield good lysis, two, or more, parallel cultures were usually set up, one with 10 ml and one with 100 ml.

The preabsorbed phage were added to 37 ml NYZ broth (Gibco) supplemented with 0.2% maltose and 10 mM $MgSO_4$ in a 250 ml Erlenmeyer flask and grown at 38^{0}C with vigorous shaking for 12 to 15 hours. When complete lysis was apparent, 1 ml of chloroform was added to the culture to kill the remaining bacteria, and the incubation was continued for another 30 minutes. The mixture was spun down, the supernatant was transferred to 50 ml Oakridge tubes and 370 µl of nuclease solution [50 mg DNase 1 (Sigma 4638), 50 mg RNase A (Boehringer Mannheim) , in 10 ml of 50% glycerol, 30 mmol NaOAc, pH 6.8] was added. The mixture was incubated at 37^{0}C for 30 minutes with periodic inversion of the tubes, and 2.1 g of NaCl was then added and mixed gently until dissolved.

After spinning the samples for 20 minutes at 7000 rpm in a JA-17 rotor at 4^{0}C, the supernatant was transferred to new tubes containing 3.7 g polyethylene glycol [PEG, Mol. Wt. 6,000-8,000 (Fisher)]. The PEG was dissolved by gentle shaking on a rotating shaker platform at 4^{0}C, then spun at 700 rpm for 20 minutes at 4^{0}C in a JA-17 rotor. The phage was suspended in

500 µl SM and transferred to an eppendof tube. Chloroform extraction of phage DNA from PEG was carried out by mixing the suspension with chloroform (500 µl) gently but thoroughly. The mixture was then spun for 5 minutes at room temperature in a microfuge. The supernatant was transferred to new tubes and 20 µl of 0.5 M EDTA, 5 µl of 20% SDS, and 10 µl of proteinase K (20 mg/ml) were added. The mixture was incubated at 65^{0}C for 30 minutes followed by extraction with phenol and chloroform. One hundred and seventy µl of 6 M ammonium acetate was added to the supernatant and precipitated with 700 µl of isopropanol. The mixture was spun down for 15 minutes in a microfuge at 4^{0}C, and the DNA pellets were rinsed with 70% ethanol and air-dried. The DNA pellets were dissolved in 500 µl TE and 5 to 10 µl were used for analysis. The yield ranged between 50-100 µg DNA, sufficient for restriction analysis and subcloning into plasmid.

DNA from positive clones was digested with EcoR I and electrophoresed on a 1% TAE agarose gel. DNA was heated to 65^{0}C for 15 minutes to separate the two arms of the phage DNA before loading onto the gel. The electrophoresed DNA was transferred to a positively charged nylon membrane (Amersham), and the filter was probed using identical DNA probes and hybridization conditions as described for the screening procedure. The cDNA inserts from the phage clones that hybridized to the DNA probe were gel purified and subcloned into the Bluscript KS$^+$ vector. Both strands of each clone were sequenced as described in section 2.8.6.

2.8.5.6 Subcloning of recombinant phage DNA insert into Bluescript KS$^+$

Isolated recombinant phage DNA was digested with EcoR I and ligated into dephosphorylated Bluscript KS$^+$ vector cut with the same enzyme. Phage DNA was cut with EcoR I and the insert was gel purified from a 1% agarose

gel using a QIAEX DNA Gel Extraction Kit (Qiagen). A minimum of 200 ng of Bluescript KS+ plasmid DNA, cut with EcoR I and treated with calf intestinal alkaline phosphatase (Boehringer Mannheim) at 37^{0}C for 30 minutes. The linerized and dephosphorylated plasmid DNA was mixed with the insert DNA at a molar ratio of 3:1 in 10 µl ligation buffer (NEB). The reaction was started by the addition of 40 Units of T4 ligase (NEB) and incubated at 12^{0}C for 12 hours.

Fifty to 100 µl of DH5a cells were electroporated (25mF, 2.5KV, and 200Ω) with 1 µl of the ligation mixture, and 500 µl of SOC medium (Maniatis, 1989) was added and incubated at 37^{0}C for 1 hour with gentle shaking. A 25 to 100 µl aliquot was spread evenly on the surface of an LB plate supplemented with 50 µg/ml ampicillin and covered with 25 µl of 40 mg/ml X-gal (BRL). The plates were incubated at 37^{0}C overnight. White bacterial clones were picked, and mini-scale plasmid DNA preparations were obtained using a QIAprep-spin plasmid Kit (Qiagen). The purified mini-prep DNA was then subjected to restriction analysis and positive clones were identified. Large scale plasmid DNA preparations of positive clones were prepared using the QIAGEN Plasmid Maxi Kit.

2.8.6 DNA sequencing of plasmid DNA

The 300 bp product from the RACE-PCR was cloned into a TA cloning vector (Invitrogen) and sequenced. Both strands of two separate clones were sequenced using T3/T7 and Reverse/Universal primers by the chain termination method of Sanger et al. (1977) using T7 DNA polymerase (Pharmacia).

The DNA sequencing service of the Montreal Cancer Institute at Notre Dame Hospital was used to sequence the two cDNA clones (735 bp insert and

1574 bp insert) in Bluescript KS$^+$. Dideoxy sequencing of DNA with fluorescently 5' end-labeled primers was performed on the A.L.F. Automatic DNA Sequencer^{TM} (Pharmacia). Sequencing reactions were done with the AutoRead^{TM} sequencing kit (Pharmacia) using the T7 DNA polymerase and the 7-deaza-dGTP to avoid compressions. The sequencing results were provided as a four color plot; each color representing a base: A, C, G or T. The data was stored on computer disks and converted readily to PCgene files for sequence analysis.

All sequence data were analyzed by PCgene. Searches for sequence homologies to all known sequences was carried out using the Blast algorithim at the NCBI genome center via E-mail (blast@ncbi.nlm.nih.gov).

Chromosome mapping of the *mborg-1* gene in human and mouse was done by Dr. X. Zhang at the Montreal Children's Hospital Cytogenetics Laboratory (Dr. P. Eydoux). Flourescent in situ hybridization (FISH) was performed using the 1574 bp cDNA clone as a probe. The FISH protocol was based on a technique derived from Lemieux et al. (1992).

Results

3.1 Gene trap in ES cells

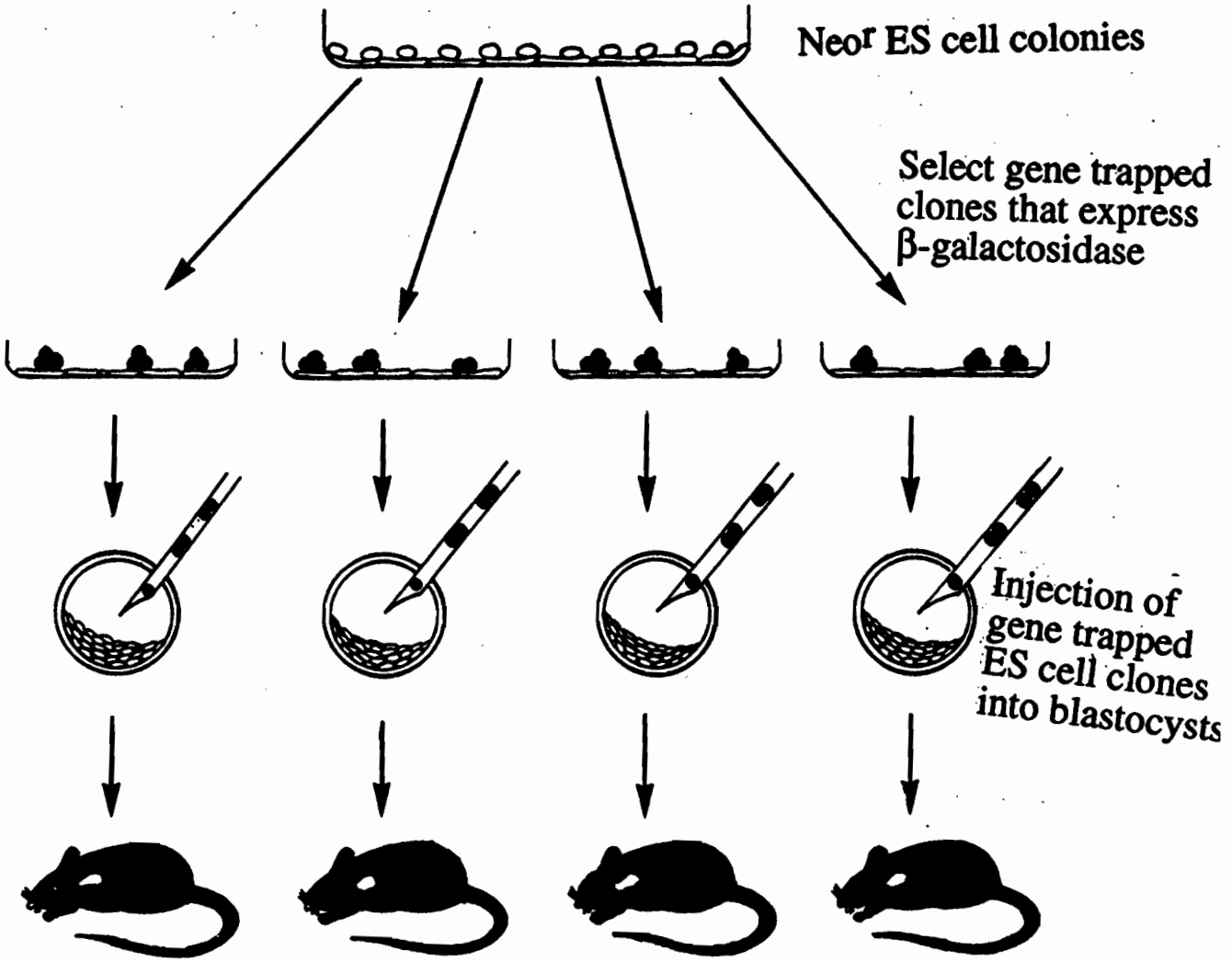
The male CCE ES cells were cultured on mitotically inactivated STO fibroblast feeder layers and transfected with the pGT 4.5 gene trap construct by electroporation. The transfected cells were then subjected to G418 selection. A total of 862 *neo*^{r} colonies were obtained. Of these ES cell colonies, four were LacZ positive and three were established as clones. The frequency of *neo*^{r} colonies expressing lacZ is thus about 0.5% (Figure 5). This frequency is somewhat lower than what was originally reported (Gossler et al., 1989).

3.2 Chimera formation

Non-transfected CCE ES cells (*BB AA; Gpi-1*^{c} /^{c}) were tested for their ability to form chimeras. The CCE ES cell line has been shown reproducibly to colonize the germ line of male chimeras with high efficiency (Robertson et al., 1986; Kuehn et al., 1987). Eight to twelve cells were microinjected into 3.5 day blastocysts obtained either from day 3.5 pregnant females or from an overnight culture of day 2.5 morulae (Figure 5). The mouse strains used for blastocyst generation were C57Bl/6 (*BB aa; Gpi-1*$^{a/a}$) and CD1 (*cc; Gpi-1*$^{b/b}$). The injected blastocysts were then transferred into the uterine horns of pseudopregnant females and allowed to develop to term (Robertson, 1987). The chimera offspring were identified by the presence of pigmented coat-hair and iris on the albino CD1 background, and by the presence of agouti hair on the black C57Bl/6 background (Figure 6).

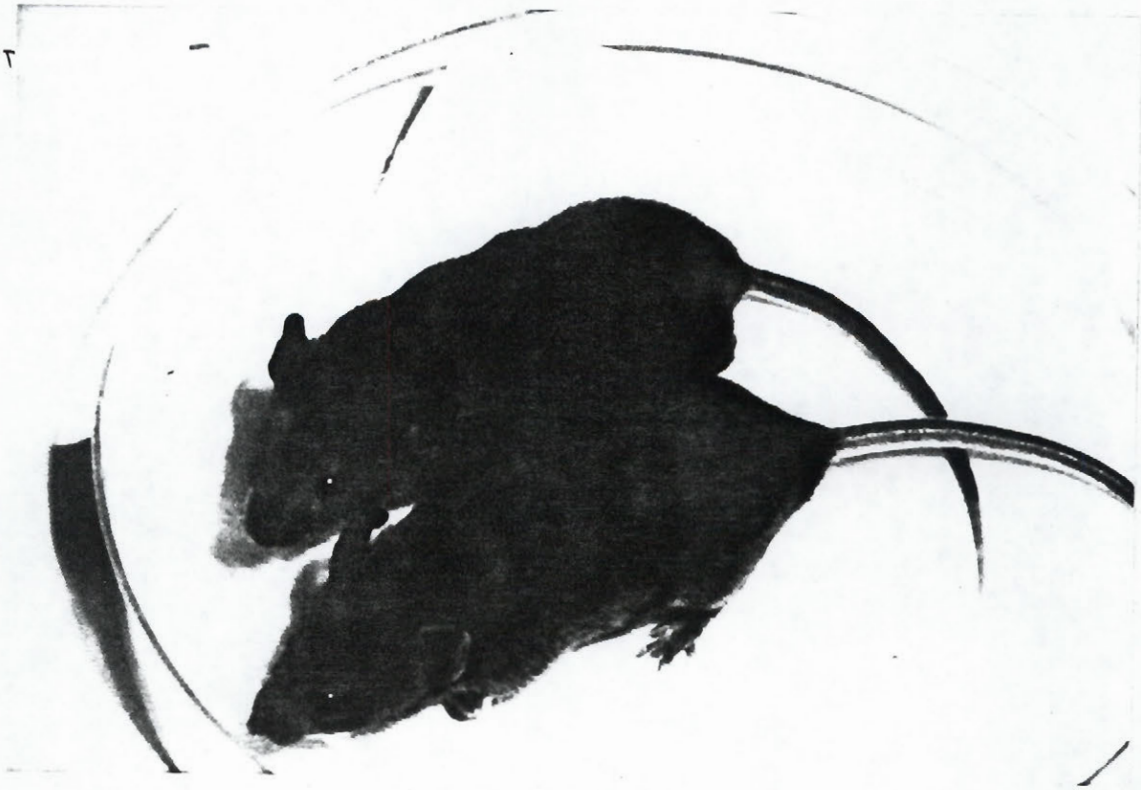
Figure 5. Schematic illustration of the gene trap strategy. The gene trap construct is introduced into ES cells by electroporation. Successfully transfected cells give rise to clones that are resistant to G418, and amongst these, a small proportion (1/215) have trapped genes and express β-galactosidase activity. The subsequent expression profiles of such trapped genes is then determined by evaluating chimeras containing these cells. When the developmental and tissue specific profiles of such disrupted genes are of sufficient interest, locus related sequences can be cloned by virtue of the known LacZ sequences present in the fusion transcript. Moreover, the *in vivo* consequences of the mutant can be evaluated after germ line passage.

Electroporate gene trap construct
Select in G418 for neor ES cell colonies



Investigate chimeras from different ES clones
to determined which cell types express β-galactosidase.

Figure 6. Photograph of chimeras derived from blastocysts of different strain backgrounds. The ES cells used in the experiments were derived from the 129 strain that is agouti. Therefore, ES cells were injected into blastocysts from a black or an albino strain so that the resulting chimeras could be readily identified their coat color. Upper panel shows a CD-1 blastocyst-derived chimera in which a small portion of the coat was pigmented by ES cell derived melanocytes. In the lower panel are two chimeras derived by injecting ES cells into the C57Bl/6 blastocysts. The predominant agouti coat indicates extensive ES cell contribution to these two chimeras. Both were also capable of germline transmission (Table 2 and 3).



3.3 Germline transmission of chimeras derived from the unmanipulated ES cells

An important issue of the gene trap experiment is the ability of the CCE ES cells to contribute to the germ cells of derived chimeras. To test the status of the CCE ES cells used in my experiments, phenotypically male C57Bl/6 and CD1 chimeras derived from unmanipulated ES cells were mated with C57Bl/6 and CD1 females, respectively. The resulting litters were scored for progeny expressing the black or agouti phenotype. Because the CCE ES cell line was derived from an *AA* bearing strain, the presence of an agouti phenotype indicated transmission of the ES cell derived genome.

The contribution of ES cells to the coat was found to be markedly influenced by the genetic background of the host blastocyst, and was proportional to the ability of the resulting chimeras to undergo germ line transmission. It was consistently observed that the ES cells contributed extensively when C57Bl/6 blastocysts were used. In the CD-1 background, however, coat pigmentation was less pronounced (Figure 6). Of the five fertile male C57Bl/6 chimeras produced, two chimeras showed by visual estimates a greater than 90% contribution to the coat from ES cells. These two chimeras also transmitted the ES cell genome to their F1 progeny (Figure 7). One of the two chimeras sired in 5 litters 22 agouti offspring out of a total of 26, while the other sired one litter of 12 offspring, all of which were agouti (Table 2). In contrast, the three male chimeras generated in the CD1 background had a much lower level of ES cell contribution to their coat, and none of the 17 to 55 offspring sired by each chimera was pigmented (Table 2 and 3).

These results confirm a previous report that the use of recipient blastocysts from different mouse strains significantly affects the degree to which injected ES cells contribute to chimeras, and the ability of such cells to

Table 2 and 3. Germline transmission of chimeras derived from the unmanipulated ES cells. Phenotypically male chimeras derived from C57Bl/6 and CD1 blastocysts injected with the unmanipulated ES cells were mated with C57Bl/6 and CD1 females, respectively. The resulting litters were scored for progeny carrying agouti phenotype carried by the CCE cell line. ES cells were found to contribute extensively when C57Bl/6 blastocysts were used with chimerism ranging from 40 to 95% as judged from the agouti coat color contribution. In the CD-1 background, however, coat pigmentation was less pronounced (5%). Of the five fertile male C57Bl/6 chimeras produced, two chimeras (C57Bl 1* and C57Bl 5*) showed, by visual estimates, a greater than 90% contribution from ES cells to the coat and these mice also transmitted the ES cell genome to F1 progeny. One of the two chimeras (C57Bl 1*) sired 29 agouti offspring out of a total of 33 in 6 litters, while the other (C57Bl 5*) delivered one litter of 12 offspring, all of which were agouti. In contrast, the three male chimeras generated in the CD1 background had a much lower level of ES cell contribution to their coat, and none among the 17 to 55 offspring sired by each chimera were pigmented. The numbers in brackets represent the ratio of female vs male pups.

Table 2

Breeding data to test for germ-line transmission of the ES cell genome

Chimera	No. of litters	No. of progeny Agouti	No. of progeny Black or Albino	Percentage of ES cell derived gametes	Females versus Males in offspring	Chimerism
C57Bl 1*	6	29(18:11)	4(1:3)	29/33 (88%)	19:14	95%
C57Bl 2	4	0	18	0	9:9	40%
C57Bl 3	3	0	22	0	13:9	50%
C57Bl 4	1	dead	dead	0	ND	50%
C57Bl 5*	1	12	0	12/12 (100%)	5:7	90%
CD-1 1	3	0	55		27:28	5%
CD-1 2	3	0	53		26:27	5%
CD-1 3	6	0	36		ND	5%

* Chimeras that have undergone germline transmission
ND=not determined

Table 3

Construction of chimeras from CCE (XY, 129/Sv/Ev) ES cells

Type of chimeras	No. of chimeras	Males	Females	Set-up	Males	
					Bred	Germline
C57Bl	7	5	2	5	5	2
CD-1	5	3	2	3	3	0

Figure 7. Photograph showing germline transmission of the ES cell genome from a C57Bl/6 derived chimera. On the left is a male, predominantly agouti chimera indicative of extensive ES cell contribution. It was mated with a C57Bl/6 female (middle) and all of their progeny (right) expressed the agouti coat color, indicating successful germline passage of the ES cell genome.

colonize the germ line (Schwartzberg et al., 1989). Our findings also suggested that the CCE ES cells used to construct the C57Bl/6<->129 chimeras are capable of germ line passage, and that the ES cell culture and the microinjection conditions employed in our laboratory were reliable in maintaining the pluripotency of the ES cells.

3.4 Failure of the GT9.6 ES cell clone to colonize the germ cells of the resulting chimeras

The gene trap ES cell clone GT9.6 was injected into blastocysts derived from C57Bl/6 or MF1 mouse strains and chimeras were identified as previously described. These two host mouse strains were used because they have both been shown to produce chimeras capable of germline transmission. Moreover, MF1 females have an added advantage over C57Bl/6 females in that a greater number of of their embryos develop to the blastocyst stage (DeChiara et al., 1990).

Eleven C57Bl/6 and 30 MF1 chimeras derived from the GT9.6 ES cell clone were produced, many of which had more than 95% ES cell coat color contribution. When all of these chimeras were crossed with their respective sex partners, none of their progeny carried the agouti coat color, typical of the 129 inbred strain from which the CCE ES cells were derived. Each of these chimeras sired more than 10 litters (12 to 20 pups per litter) and the coat color of the offspring inherited was exclusively that of their host blastocysts, plain black for C57Bl/6 and albino for MF1 (Table 4). These results indicate an apparent failure of the GT9.6 ES cell clone to colonize the germ cells of the chimeras. Consequently, a direct analysis of the developmental regulation of the gene trapped in the GT9.6 ES cell clone in transgenic mice was not possible. Similarly, in the absence of germ line passage of the insertional

Table 4

Construction of chimeras from Clone GT-9.6

Type of blastocysts	No. of blastocysts injected	No. of litters born	No. of chimeras Male	Female	No. of chimeras Bred	Germline
C57Bl	515	69	5	6	11	0
MF-1	459	152	17	13	30	0

mutation, it was not possible to assess the phenotype of mice homozygous for the interrupted allele.

3.5 Expression of clone GT 9.6 in the nervous system of mouse chimeras

Following transfection of the gene trap construct by electroporation, four ES cell clones that express β-galactosidase activity were obtained (Figure 8). In each one of these clones the construct must have been inserted into an actively transcribed gene, simultaneously trapping and mutating it. One of these ES cell clones (clone GT 9.6), when injected into mouse blastocysts, produced chimeras in which the trapped gene was expressed. It was expressed predominantly in all neuronal populations of the mature mouse nervous system. β-galactosidase activity was observed widely in almost all neuronal cell types in the brain of adult chimeras. β-galactosidase labeled cells were found in regions of cerebral cortex, hippocampus, caudate putamen, brain stem, thalamus, and other areas (Figure 9). In the cerebellum of the chimeras, the molecular, Purkinje cell layer and the granular cell layer contained both labeled and unlabeled cells, although the granular cell layer was disproportionally populated with a higher number of β-galactosidase expressing cells (Figure 10).

In other parts of the mouse nervous system, β-galactosidase expressing cells were found in areas of spinal cord known to contain interneurons and motor neurons (Figure 11), in layers of the retina (Figure 12), and in the olfactory epithelium (Figure 14). Thus, the expression pattern of this gene did not appear to be restricted to any group of functionally related neurons nor to neurons in a given tissue, nor a given organ. Rather the gene trap fusion protein was detected in a wide variety of neurons. This preliminary histochemical analysis suggested that the trapped gene may be expressed in a

Figure 8. Photograph showing ES cells in culture. One of the four ES cell clones, bearing different trapped genes, is shown. All of these clones expressed readily detectable β-galactosidase activity. The β-galactosidase reaction product is blue. ES cells (arrow) were grown on fibroblast feeder layers (arrowhead).

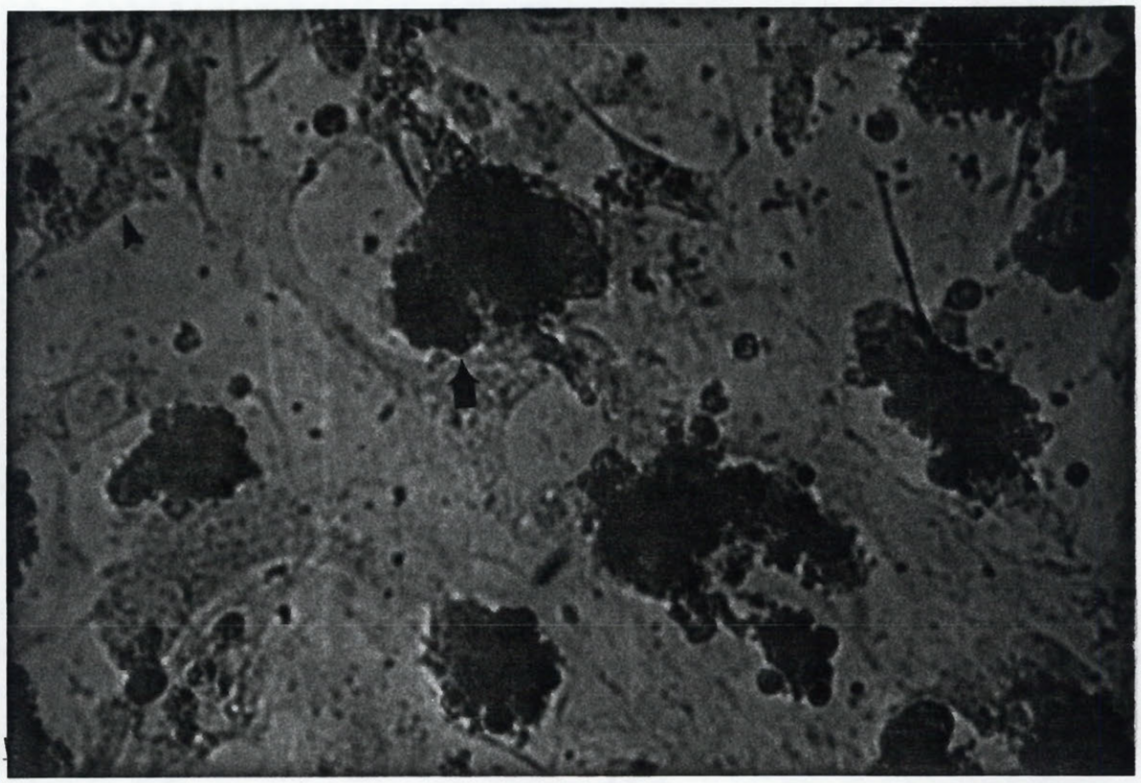
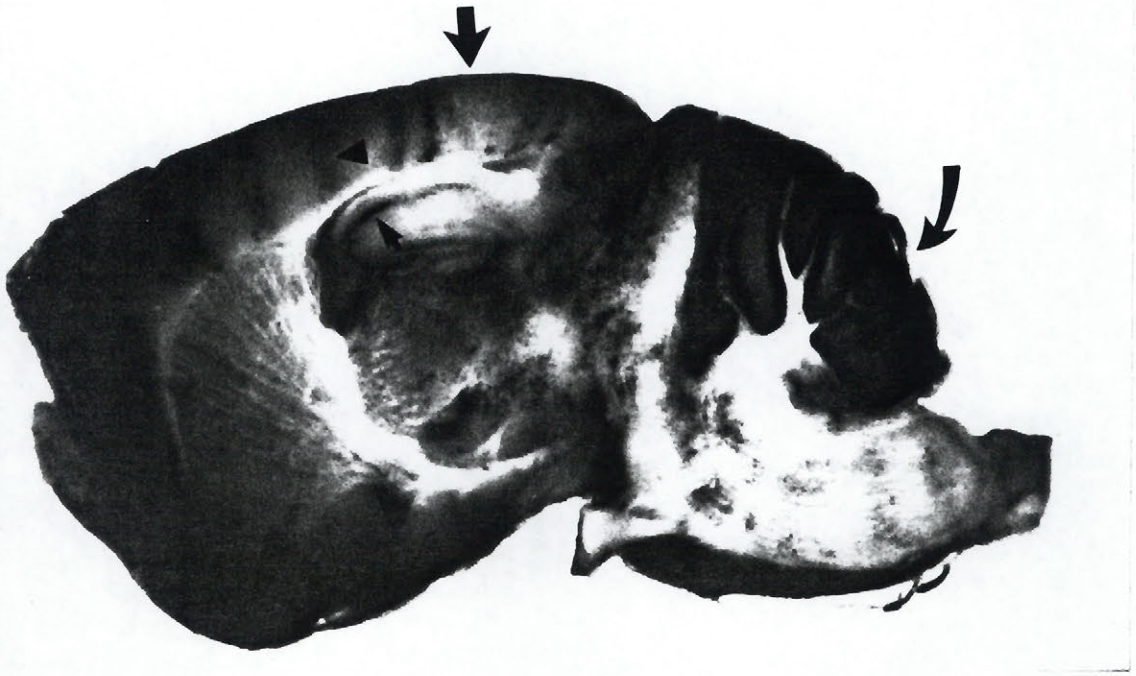


Figure 9. Whole mount sagital brain section from two adult GT9.6 chimeras histochemically stained for β-galactosidase activity. (A). Brain section from a GT9.6 chimera with less than 50% ES cell contribution (B). Brain section from a GT9.6 chimera with more than 95% ES cell contribution. The blue stained cells represent neurons in which the trapped gene was expressed. These β-galactosidase expressing cells are presumably derived from the ES cell component. Cerebral cortex, cerebellum and hippocampus are indicated by arrow, curved arrow and small arrow, respectively. As in all other chimeras analyzed, β-galactosidase expressing cells found in cerebral cortex were organized into radial columns (arrowhead) in both chimeras shown here.

A



B

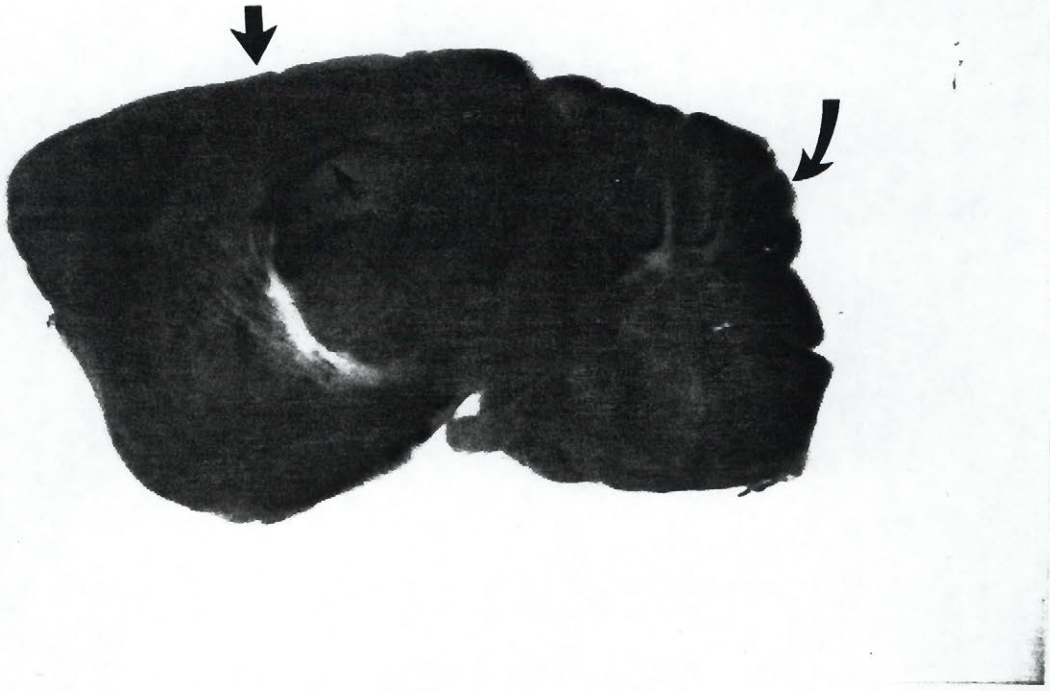


Figure 10. Photograph illustrating β-galactosidase expressing cells in the cerebellum of a GT9.6 chimera. Upper panel show a whole mount sagital section of the cerebellum from a GT9.6 chimera. Lower panel illustrates a 12 μm thick cryostat section of the cerebellar cortex. The cerebellar cortex is composed of a molecular cell layer (M), a Purkinje cell layer (arrows) and a granular cell layer (G). β-galactosidase labeled cells are observed in both the molecular and the granular cell layer of the cerebellum. The whole mount section of the cerebellum (upper panel) shows a high proportion of β-galactosidase expressing cells in the granular cell layer. Labeled Purkinje cells were seen in other sections (Figure 15, 16).

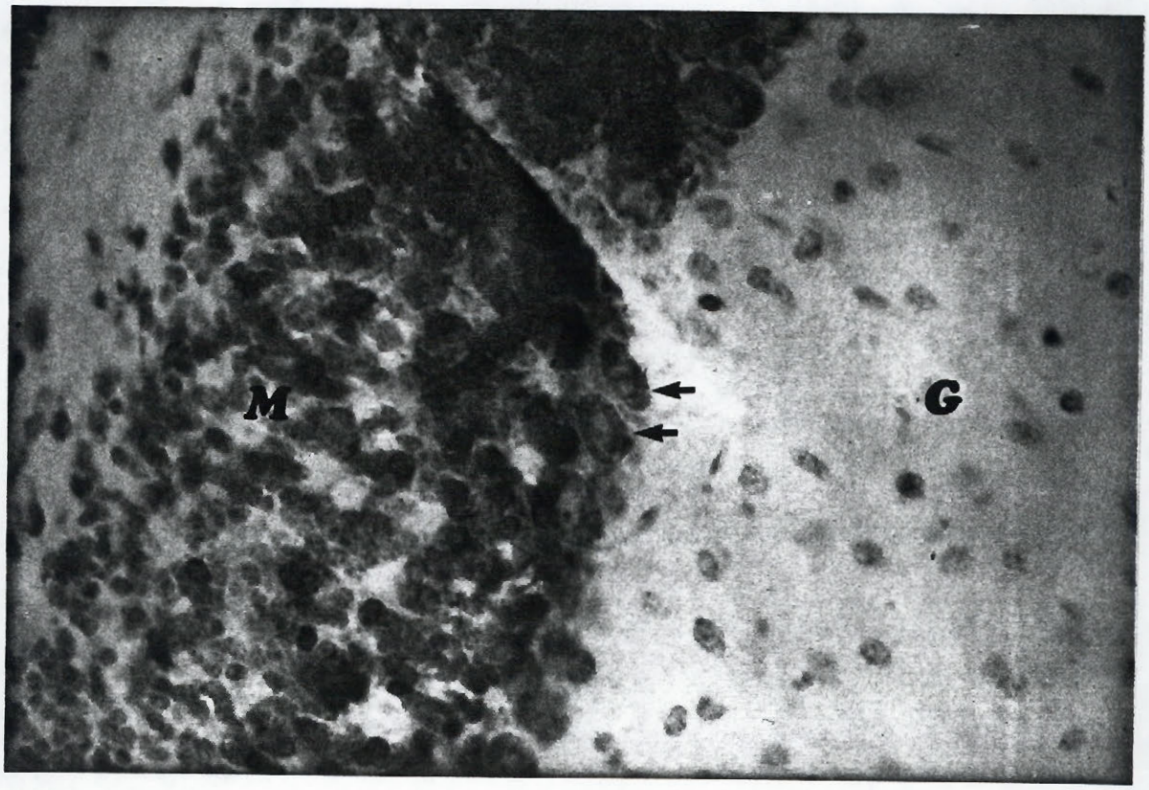

M
G

Figure 11. Photograph illustrating β-galactosidase expressing cells in the spinal cord of a GT9.6 chimera. Twelve μm thick cryostat cross section from the thoracic level of the spinal cord from a GT 9.6 MF1 mouse chimera. Ventral (V) and dorsal horn (D) of the gray matter of the spinal cord containing mostly the large sized motor neurons (arrows) and the interneurons (arrowhead), respectively, are indicated. β-galactosidase labeled neurons were found predominantly on one side of the cord (right). The contralateral side contained fewer labeled cells (left). The spinal cord section shown here is from a GT 9.6 MF1 mouse chimera with less than 50% of ES cell contribution to its coat.

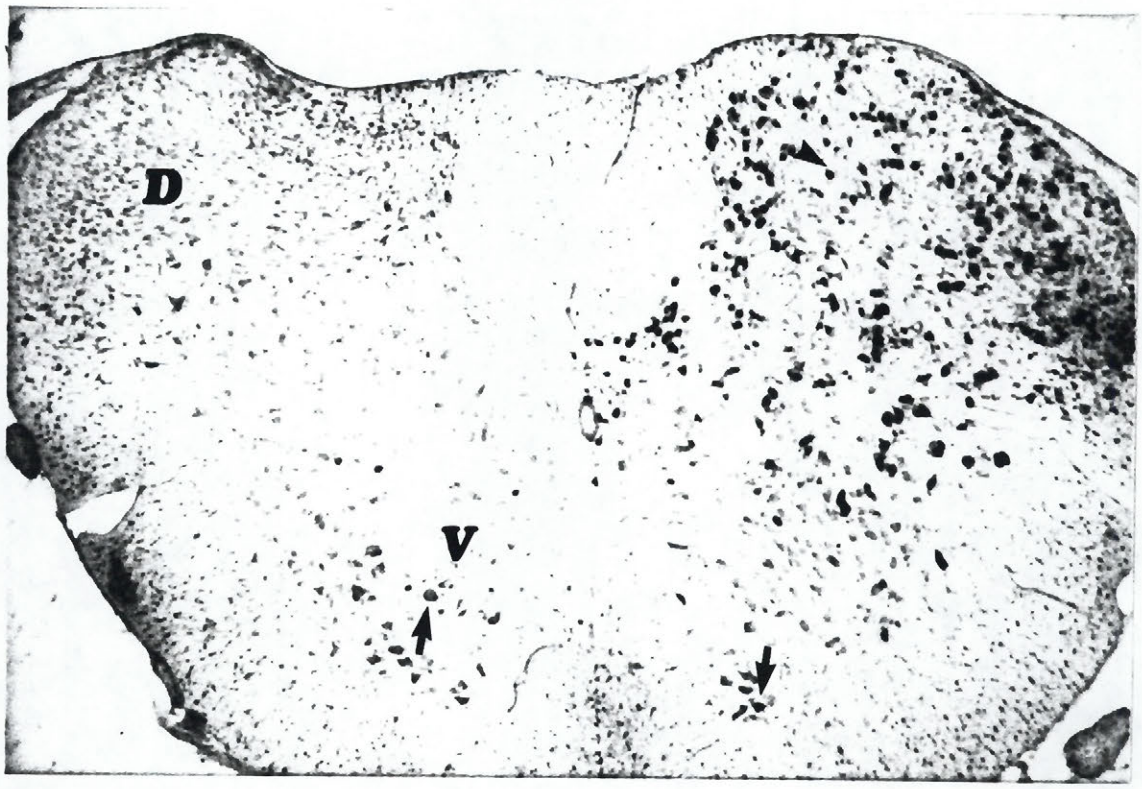
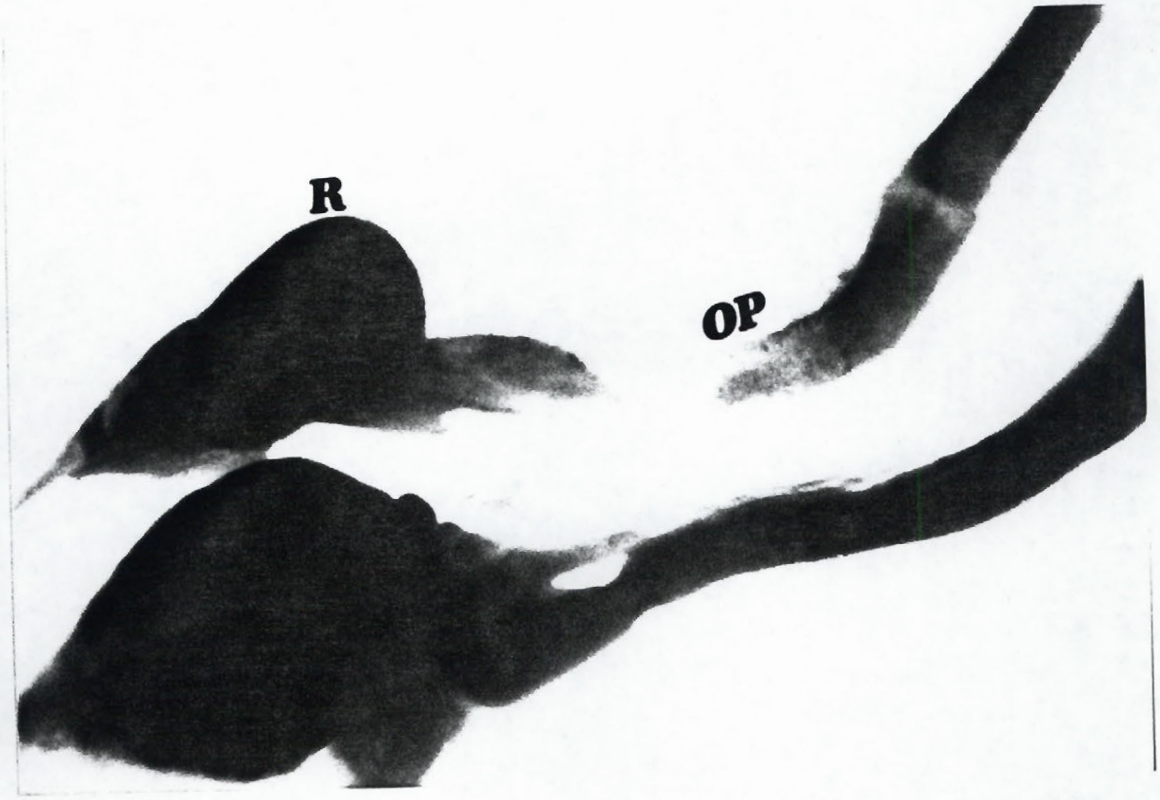
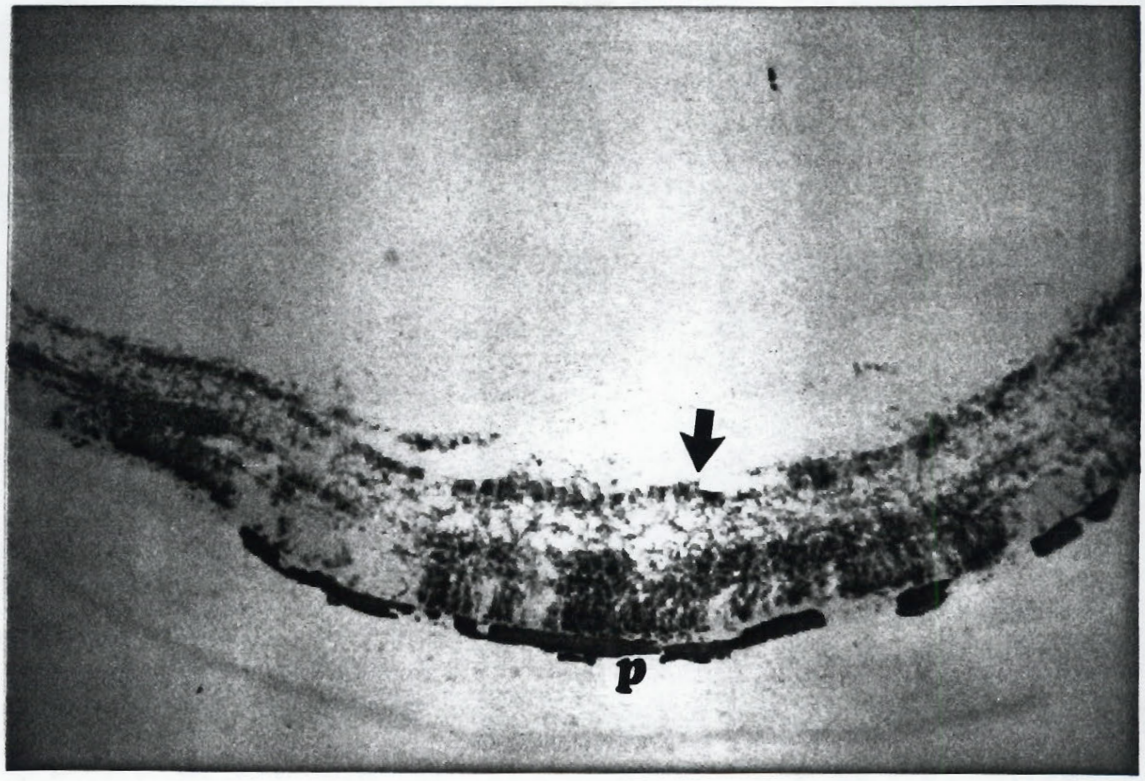

D
V

Figure 12. Photograph illustrating β-galactosidase expressing cells in retina from a GT 9.6 mouse chimera. Upper panel shows whole mount preparation of retina (R) and optic nerves (OP), and lower panel illustrates a 12 μm cryostat retinal section. β-galactosidase expressing cells were observed in all cell layers of retina (arrow). The pigmented cell layer of the retina is indicated by P.


R
OP

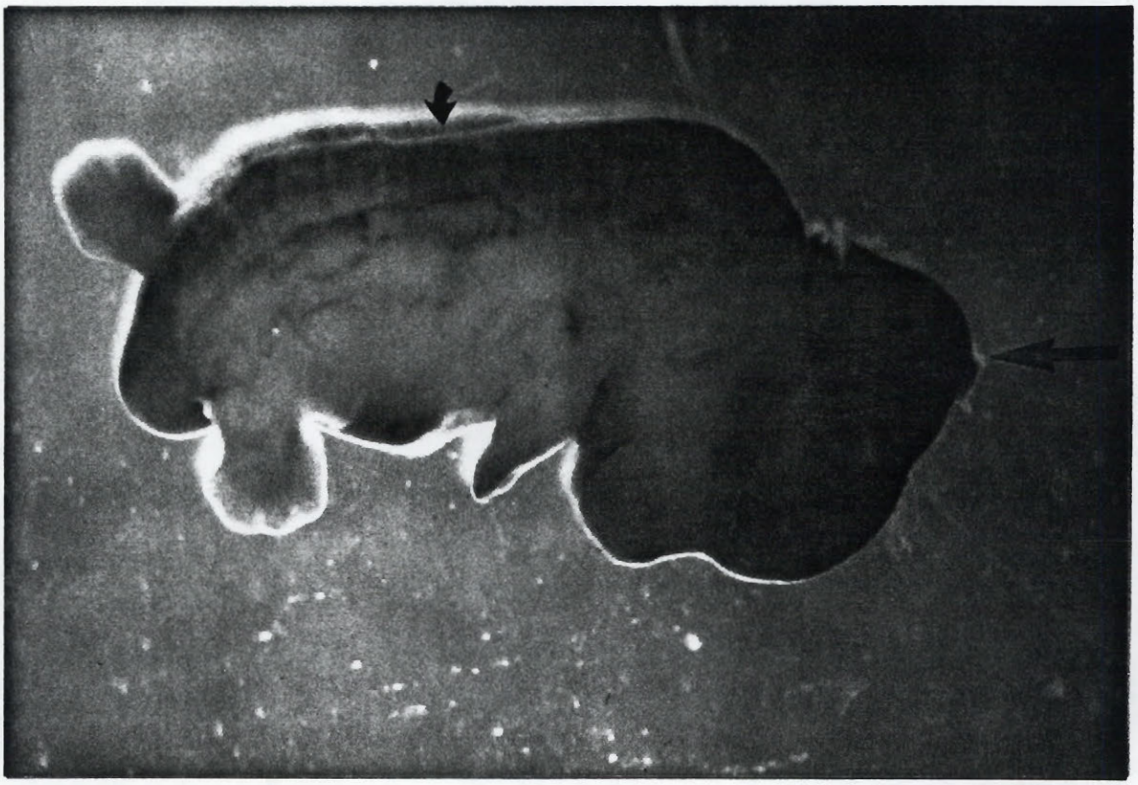

p

pan-neuronal fashion. As a consequence, it may be essential to some aspect of the commitment process to neuronal phenotype, or to the maintenance of neuronal phenotype. β-galactosidase stained cells were also observed in other chimeric tissues, including heart, gut and kidney. These labeled cells may be representative of the innvervating autonomic nervous system, as the parasympathetic nervous systems are known to be closely associated with their target organs of innervation.

3.6 Implication of the trapped gene for neurogenesis

Since β-galactosidase labeled cells were observed in all neuronal populations examined, it was important to establish when this gene was expressed during neurogenesis. One way of addressing this question is to analyze the expression pattern of this gene during embryogenesis. β-galactosidase expressing cells were indeed located throughout the brain and spinal cord of day 12.5 GT9.6 chimeras (Figure 13), indicating that that not only is the trapped gene expressed in all types of neurons (Figure 9-12), but that it may also be expressed in all stages of neurogenesis. An alternative approach of analyzing the expression of the trapped gene during neurogenesis is to study the expression pattern of this gene in the olfactory epithelium of adult chimeras. The olfactory epithelium is composed of multiple layers of sensory neurons and their progenitors. The upper layer of mature neurons is constantly renewed by the underlying neuroprogenitor cells. β-galactosidase labeled cells were detected in all layers of the olfactory epithelium (Figure 14). Thus, it appears that the trapped gene is expressed during the early nervous system development, as well as in the mature nervous system.

Figure 13. Expression of the trapped gene in the mouse developing nervous system. Whole mount day 12.5 embryo from a GT9.6 chimera was histochemically stained for β-galactosidase activity. β-galactosidase expressing cells in the brain, spinal cord and eyes are indicated by long, curved and short arrows, respectively. Note the distribution of coherent patches of the labeled spinal cord cells.



3.7 The trapped gene encodes a nuclear protein

Because the β-galactosidase histochemical reaction product tends to diffuse within cell bodies and adhere to membranous structures, i.e. plasma or nuclear membranes, the simple identification of β-galactosidase activity in different tissues and cells does not reveal accurately the subcellular localization of the gene trap fusion protein (Figure 15). Therefore, immunocytochemistry was also performed on the GT9.6 ES cell clone derived chimeras using an anti β-galactosidase antiserum. In examined sections of cerebellum and spinal cord, the fusion protein was found to be localized predominantly in neuronal nuclei. The entire nucleus of Purkinje cells in the cerebellum and of motor neurons in the spinal cord was labeled with the antibody (Figure 16). Because the β-galactosidase encoded by the gene trap construct does not contain a nuclear localization signal (NLS), the nuclear localization of the fusion protein suggests that the protein product of the trapped gene contains the NLS directing the subcellular localization of the fusion protein.

3.8 Clonal segregation during neurogenesis

Chimeras generated by the ES cell clone GT 9.6 also provide a useful means of analyzing the distribution of cell clones during neurogenesis. This is because in such chimeras, most tissues, including the nervous system, are composed of two cell populations, one derived from the injected ES cells and the other from the inner cell mass of the host blastocyst. Therefore, the pattern of distribution of these two cell populations, as revealed by β-galactosidase stained cells, allows a retrospective analysis of the pattern of generation of tissues during development.

Figure 14. Photograph illustrating β-galactosidase expressing cells in olfactory epithelium from a GT 9.6 mouse chimera. A 12 μm cryostat section of the olfactory epithelium from a GT9.6 mouse chimera. The olfactory epithelium (arrows) and the nasal septum (NS) are shown. The olfactory epithelium is composed of multiple layers of sensory neurons and their progenitors. The outer layer of mature neurons is constantly renewed by the underlying neuroprogenitor cells. β-galactosidase labeled cells were detected in all layers of the olfactory epithelium (arrows).

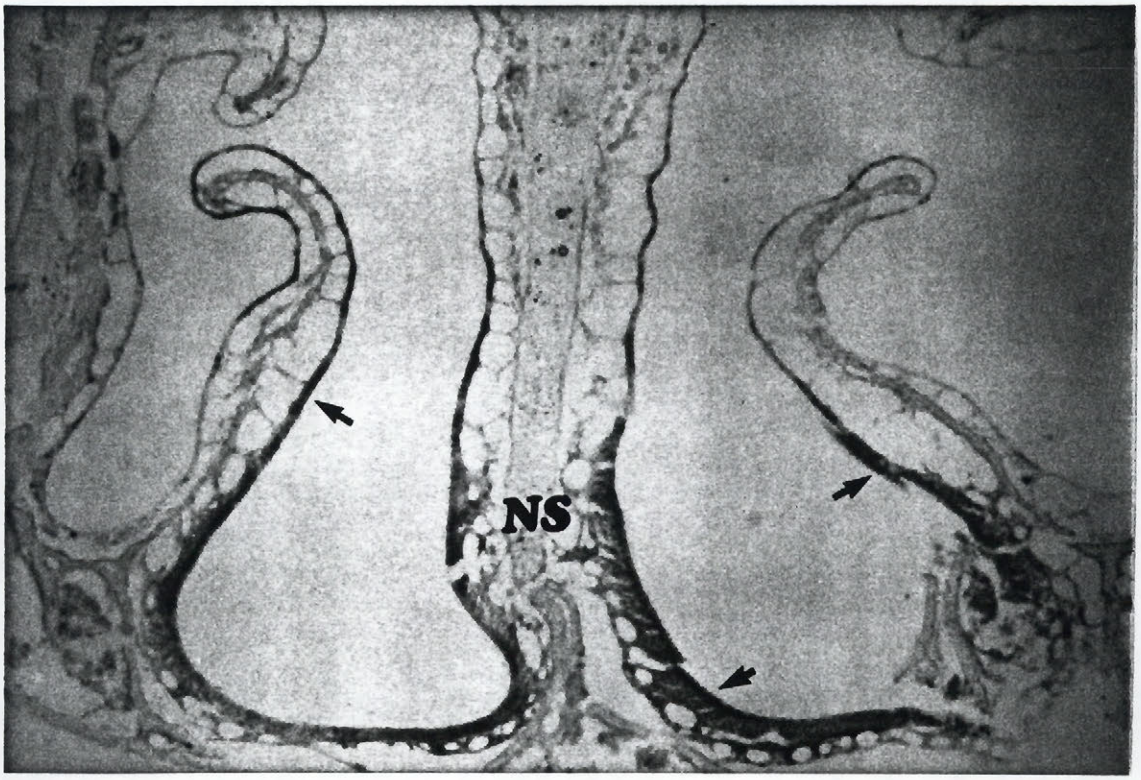
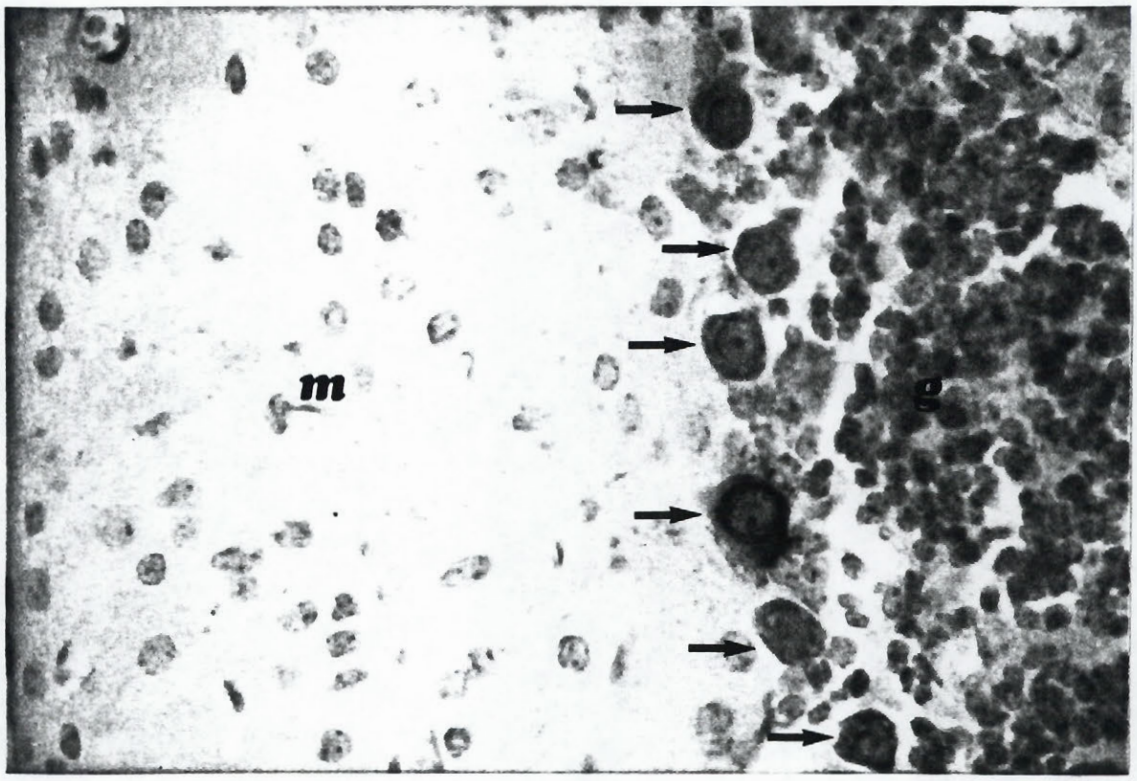

NS

Figure 15. Histochemical determination of β-galactosidase activity in cerebellar Purkinje cells and spinal cord motor neurons of a GT9.6 mouse chimera. Twelve μm thin cryostat section of cerebellum (upper panel) and spinal cord (lower panel) from a GT9.6 mouse chimera is shown. In cerebellar cortex, Purkinje cells (arrows) are located between the molecular cell layer (M) and the granular cell layer (G). In spinal cord, motor neurons are large sized neurons (arrowheads) located in the ventral horn of the gray matter shown in the lower panel. In Purkinje cell and motor neurons that express β-galactosidase activity, the β-galactosidase histochemical reaction product seemed to be localized on the nuclear membranes.


m
g

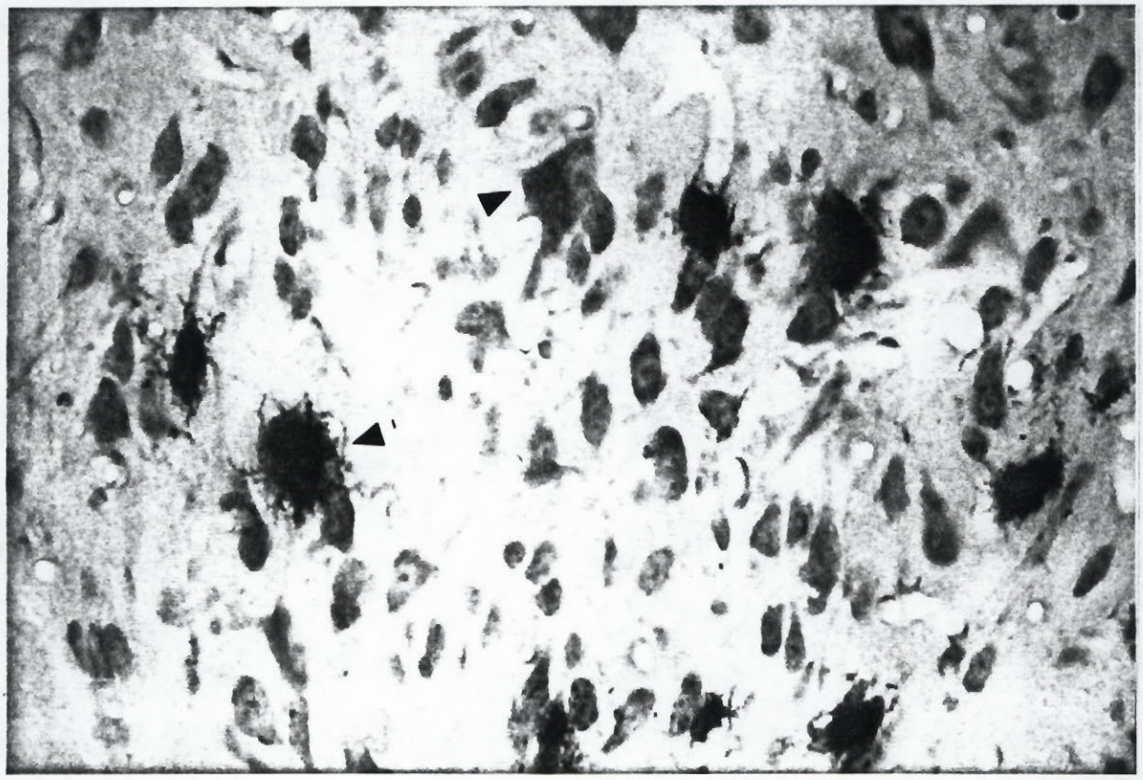
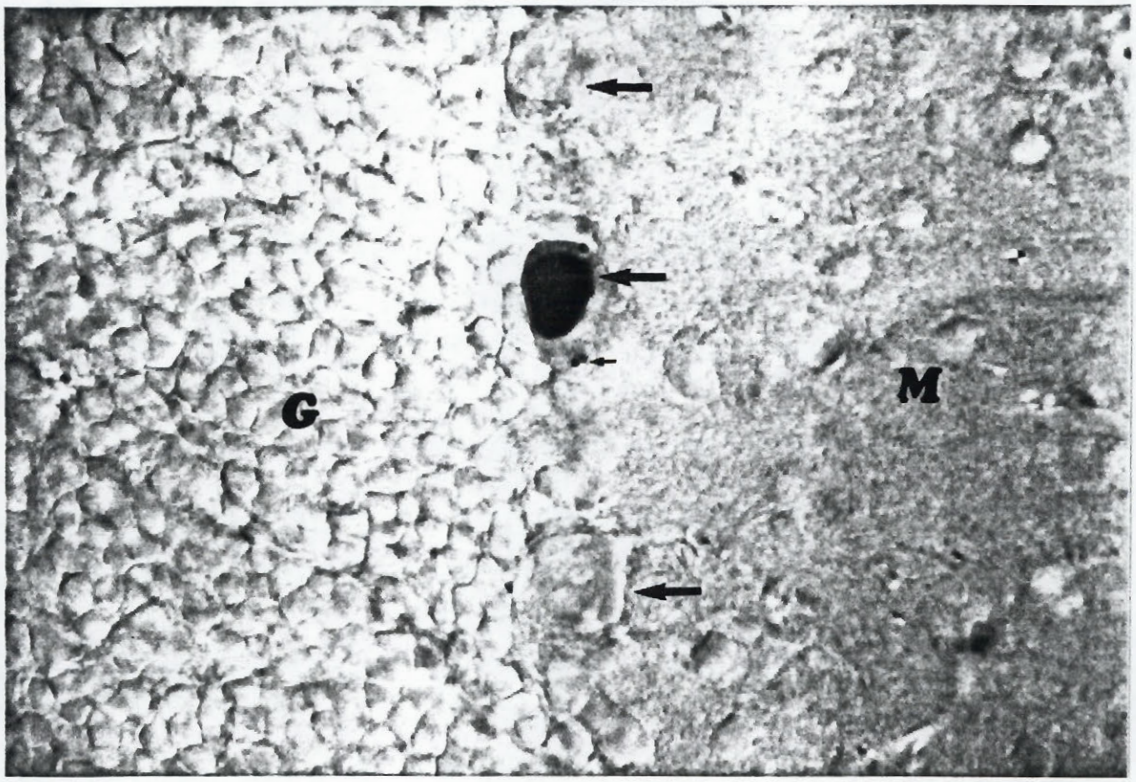
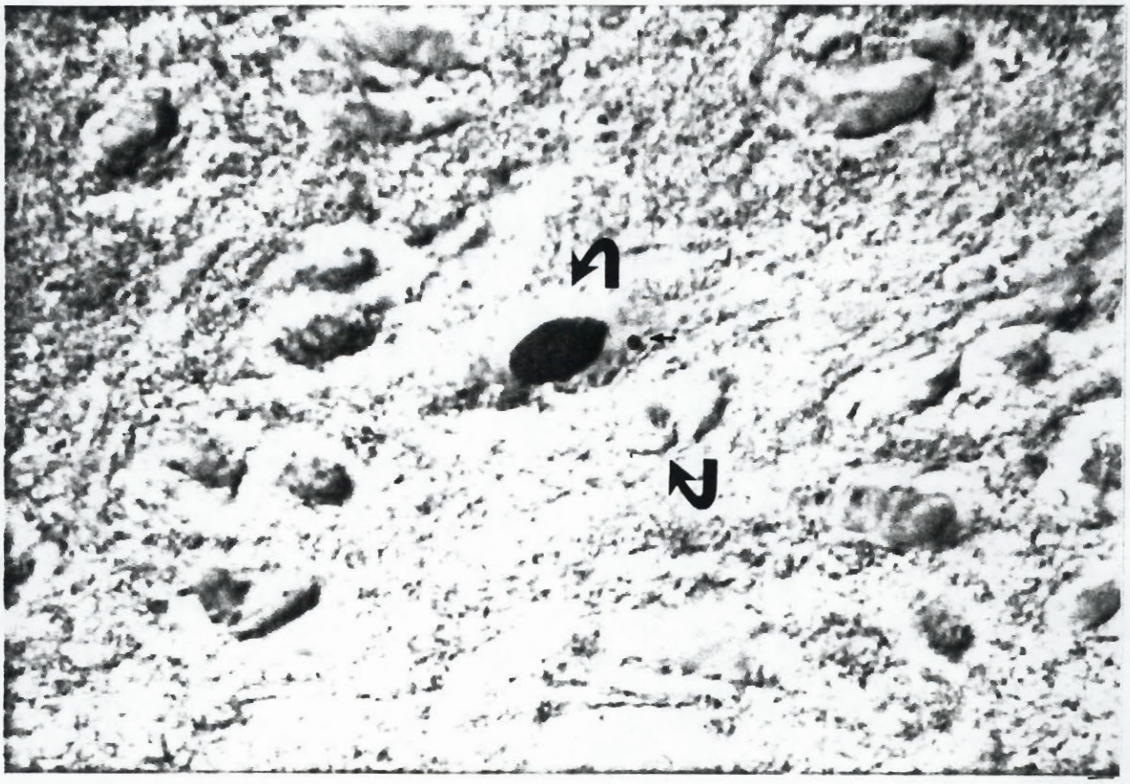


Figure 16. Subcellular localization of β-galactosidase activity in cerebellar Purkinje cells (upper panel) and spinal cord motor neurons (lower panel) of a GT 9.6 mouse chimera determined by immunocytochemistry. Sections of cerebellum and spinal cord were processed as described in materials and methods and were reacted with an anti β-galactosidase antiserum. As in Figure 15, Purkinje cells (arrows) are located between the molecular cell layer (M) and the granular cell layer (G) in the cerebellum (upper panel), and motor neurons are large sized neurons (curved arrows) located in the ventral horn of the gray matter shown in the lower panel. The β-galactosidase fusion protein was localized predominantly in neuronal nuclei. The entire nucleus of Purkinje cells and of motor neurons was labeled by this technique. In addition, small onion-shaped cytoplasmic structures were labeled (arrowheads).


G
M



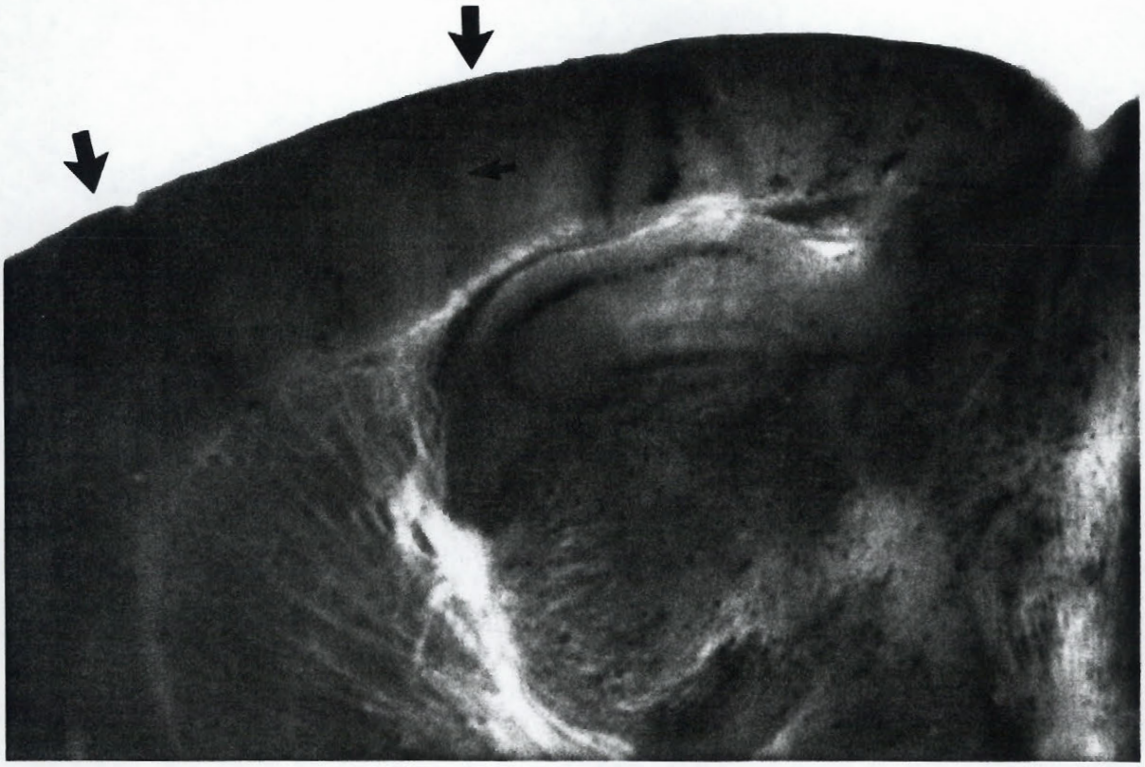
Forty-one adult chimeras produced by ES clone GT 9.6 were analyzed. When the cerebral cortex of each chimera was examined, β-galactosidase expressing cells were found to be organized into radial columns (Figure 17). The width of these columns and the space between them differ from one another in the same chimera, and as well vary among different chimeras. In chimeras with less ES cell contribution, the radial columns are well separated from each other by domains filled entirely with unlabeled cells, while in chimeras that carry an extensive ES cell component, the space between the columns contain both β-galactosidase expressing and unlabeled cells (Figure 17). These results suggest that the development of the cerebral cortex follows a specific pattern. The labeled cells within the radial columns may represent the clonal progeny of cortical neuronal progenitors migrating along their specific pathways during cortical development. The seemingly individual radial stripes, especially those observed in chimeras with low ES cell contribtution in which the radial stripes were separated by domains consisting entirely of unlabeled cells (upper panel, Figure 17), may delineate those cortical neurons that migrate strictly along a radial pathway. On the other hand, the labeled cells scattered within the domains of unlabeled cells, often seen in chimeras with predominant ES cell contribution, suggest that there may be an additional mode of cell dispersion other than radial migration which may play a role during cortical development.

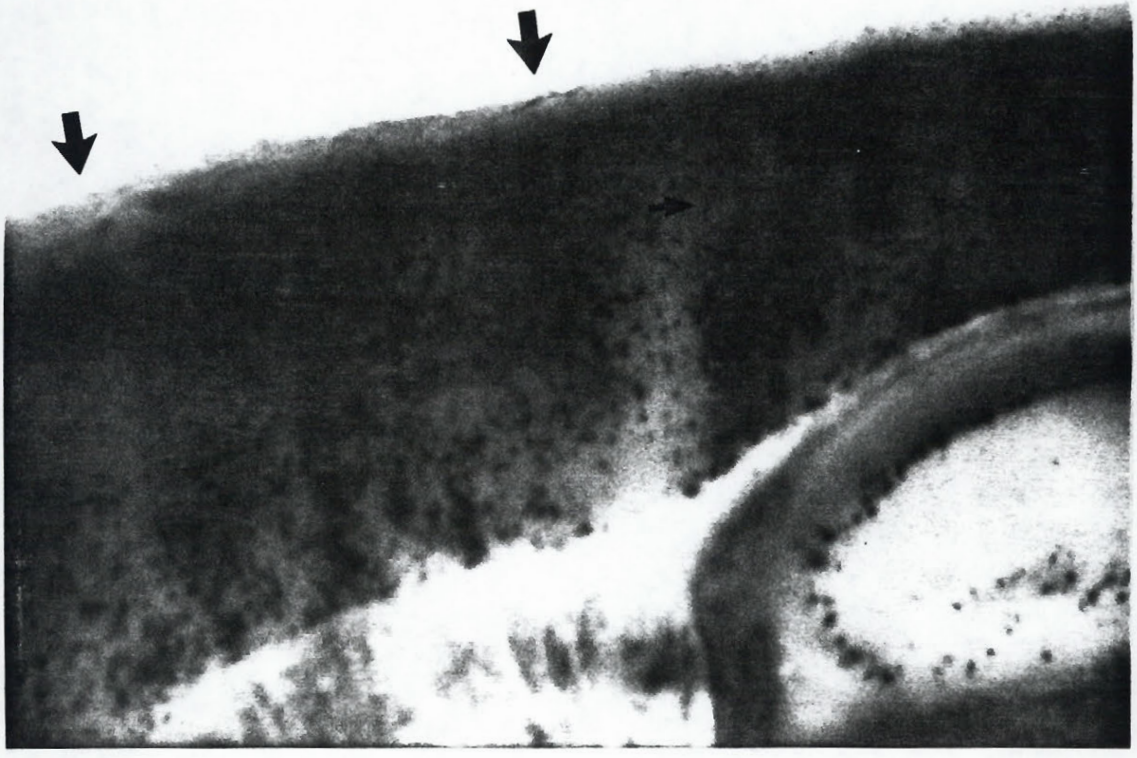
Mouse cortical neurogenesis climaxes around embryonic day 12.5. The pattern of cell migration during cortical genesis was thus further analyzed in day 12.5 GT9.6 chimeric embryos. Prominant radial bands of β-galactosidase labeled cells were found to dominate the parasagittal planes of the day 12.5 embryonic cortex (Figure 20). Strikingly, most radial stripes were either in single cell width, or in a width of no more than a few cells. They were

extending either part or the entire length of the cerebral cortex, perpendicular to the ventricular zone. These results indicate that radial migration may be the predominant pattern of cell migration during early cortical genesis. In addition, there seemed to be a higher density of labeled cells near pia, suggesting that a component of horizontal cell migration may prevail once the radially migrating cells have reached their layer of destination.

In the spinal cord of GT 9.6 ES cell derived adult chimeras, β-galactosidase stained cells were noted along the entire length of the spinal cord (Figure 18). However, in cross sections of the spinal column, labeled cells were organized into homogenous patches on either side of the spinal cord (Figure 11). This striking distribution of seemingly coherent patches is more apparent in chimeras where the ES cell contribution is relatively low. In one chimera with less than 50% of ES cell colonization, the labeled spinal cells were found only on one side of the spinal cord. The contralateral side contained very few, if any, labeled cells (Figure 11). This difference in the distribution of β-galactosidase stained cells on either side of the spinal cord was more apparent on the longitudinal axis, where the labeled cell patches were often absent on the other side of the spinal cord (Figure 18). The rostral-caudal extent of these seemingly homogenous domains appear to be random and does not correspond to any obvious anatomical or functional segments of the spinal column, such as the vertebral level (Figure 18). This phenomenon, however, was less obvious in chimeras with extensive ES cell contribution in which the chimeric spinal cord were predominantly populated with the ES cell derived β-galactosidase expressing cells (Figure 18). Nonetheless, when serial 12 μm spinal cord sections were obtained from one chimera which contained more than 90% of an ES cell component, in one cervical segment of 48 μm long and one thoracic segment of 816 μm long, labeled cells were found

Figure 17. Radial pattern of distribution of β-galactosidase labeled cells in the cerebral cortex of the GT9.6 chimeras. Whole mount sagittal sections of mouse brain from two GT9.6 mouse chimeras histochemically stained with the bluo-gal substrate. The upper and lower panel show brain sections from a GT9.6 chimera with less than 50% ES cell contribution and a GT9.6 chimera with more than 95% ES cell contribution, respectively. In each panel the cerebral cortex is indicated by arrows and in this structure the β-galactosidase expressing cells appeared to be organized into radial columns (arrowheads). The width of the cortical radial stripes and the spacing between them differ in the same chimera, and vary between these two chimeras. In chimeras with low ES cell contribution the radial columns are well separated from each other by unlabeled cells (upper panel), while in chimeras with extensive ES cell component (lower panel) the space between the labeled stripes contain both β-galactosidase expressing and unlabeled cells.





only on one side of the spinal cord, the contralateral side contained very few labeled cells (Figure 19). In the remaining spinal cord, a similar number of β-galactosidase expressing cells were found on both sides of the spinal cord.

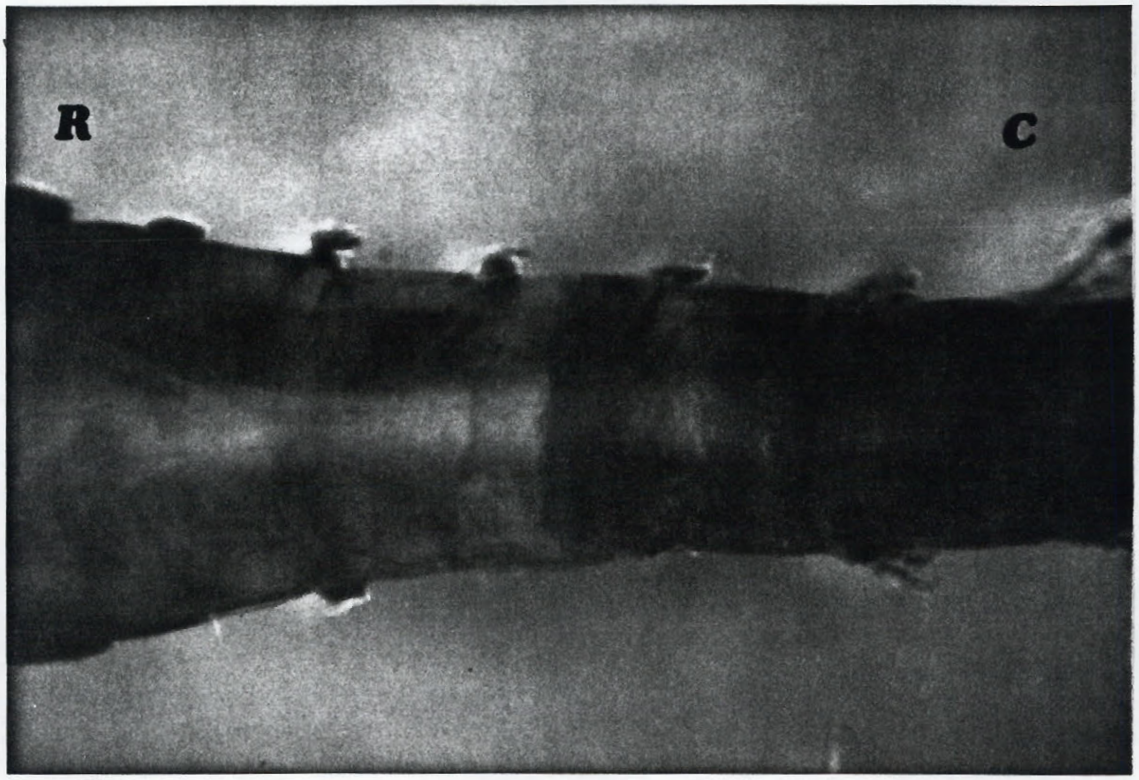
The spatially restricted cell distribution during spinal cord genesis was most evident in the spinal cord of embryonic day 12.5 (E12.5) chimeras. The labeled cells were organized in seemingly coherent patches along the longitudinal axis of the developing spinal cord (Figure 21). Moreover, there were apparent differences in the distribution of the labeled cells on either side of the spinal cord. Segments of labeled cells on one side of the spinal cord were often accompanied by the absence of any labeled cells in the corresponding segments on the opposite side of the spinal cord (Figure 21), although in some other segments of the spinal cord, both sides of spinal cord contained similar numbers of β-galactosidase labeled cells (20).

These results suggest that during spinal cord development, progeny of a single progenitor cell (clone) may not mix extensively with the clonal progeny originating from the other precursors, as suggested by Musci and Mullen (1992). Rather, they may distribute as cohesive patches and respect boundaries of clonal compartments defined by the cohort cell patches.

3.9 Molecular characterization of clone GT 9.6

The gene trap method was used in this investigation to identify new mammalian genes that may possess important developmental functions. Gene trapping is a random approach, and the trapped genes of interest are selected based on their developmental and tissue specific expression pattern. The application of gene trapping thus suggests that the function of the trapped genes may be inferred from their spatial and temporal expression pattern.

Figure 18. Spatially restricted distribution of β-galactosidase expressing spinal cord cells in the adult GT9.6 chimeras. Whole mount spinal cord from two GT9.6 mouse chimeras. In the upper panel, the spinal cord (cervical to thoracic level) is from a chimera with less than 50% ES cell contribution; in the lower panel, the spinal cord (thoracic level) is from a chimera with more than 95% ES cell colonization. Rostral and caudal parts of the spinal cord are indicated by R and C, respectively. β-galactosidase stained cells could found along the longitudinal axis of the spinal cord. In some regions of the spinal cord, especially in chimeras with extensive ES cell contribution (lower panel), β-galactosidase expressing cells were located on both side of the spinal cord. In chimeras with less ES cell component (upper panel) the regions containing the labeled cells were distributed asymmetrically on either side of the spinal cord.


R
C

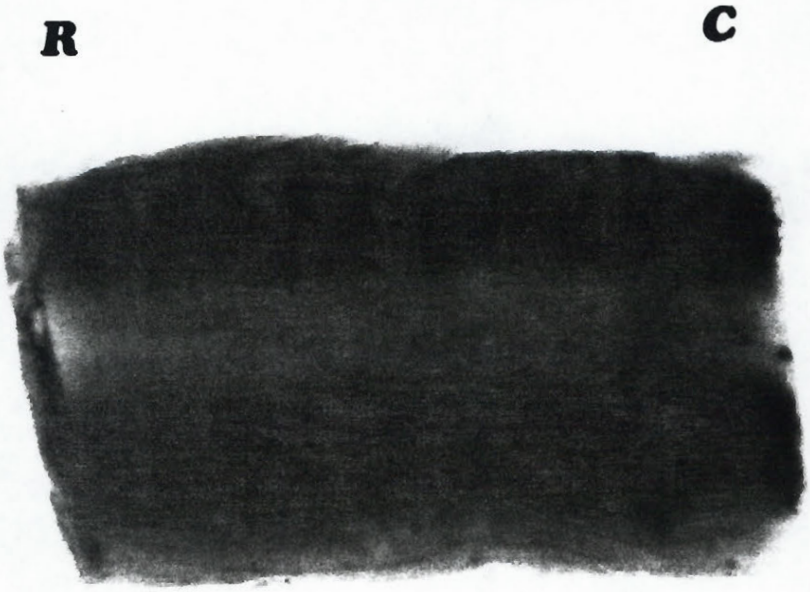
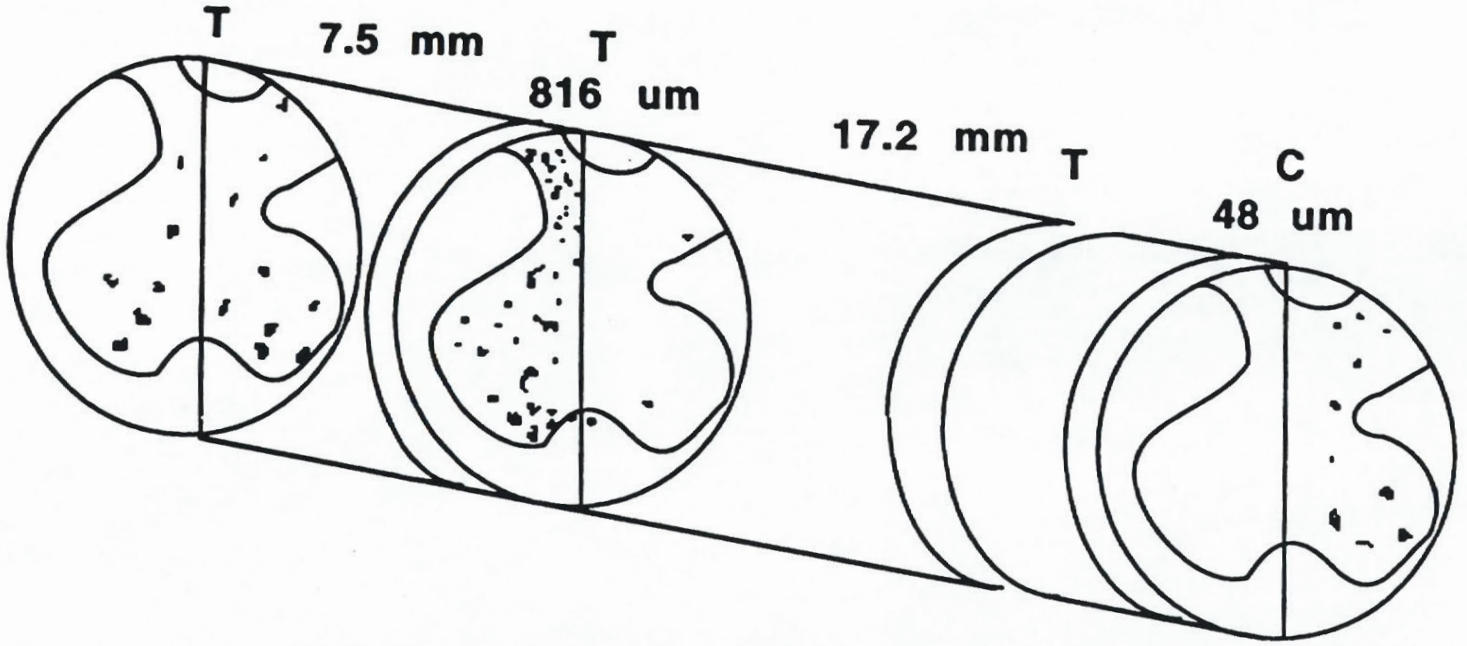

R
C

Figure 19. Schematic illustration of serial spinal cord section and arrangement of coherent spinal cord neurons. Serial 12 μm spinal cord sections were obtained from one chimera in which more than 95% of the spinal cord neurons are of the ES cell genotype, as estimated from the proportion of the β-galactosidase expressing cells in the spinal cord. Despite of the extensive ES cell contributions, in one cervical segment of 48 μm long and one thoracic segment of 816 μm long, labeled cells were found predominantly on one side of the spinal cord. The contralateral side contained much fewer labeled cells. In the remaining spinal cord, the numbers of β-galactosidase expressing cells were evenly distributed on both sides of the spinal cord.

Three dimensional arrangement of coherent spinal cord neurons



While this strategy can be applied to uncover new genes in any system, it was used in this study in an attempt to identify genes that may play important roles during the nervous system development. Although selection of the trapped genes important for neural development is expected to be fortuitous, massive screening of gene trap events in ES cells, and subsequent characterization of the expression pattern of the trapped genes in chimeras have made it possible to choose genes with a spatially and temporally regulated expression pattern in the chimera nervous system. The interesting expression pattern of the trapped endogenous gene in the ES cell clone GT9.6 in the developing and mature nervous system, including the fact that it encodes a nuclear protein that is expressed in a pan-neuronal fashion in the mature mouse nervous system, prompted me to attempt characterization of the trapped gene at the molecular level.

3.9.1 Expression of the gene trap fusion transcript in ES cells

In each of the three β-galactosidase expressing ES cell clones recovered, an actively transcribed endogenous gene must have been trapped and simultaneously mutated in a manner leading to a fusion transcript. To identify the fusion transcripts expressed by the trapped genes in these ES cell clones, Northern blot analysis, using a LacZ gene probe, was performed on RNA isolated from the ES cell clones. Each gene trapped ES cell clone was found to express a distinct fusion message while no transcripts hybridizing to the probe were detected in the control CCE ES cells (Figure 22). The fusion transcript encoded by the ES cell clone GT9.6 is about 4.6 kb. Since 3.2 kb of the sequence in this fusion transcript can be accounted for by LacZ *En*-2 gene sequences, approximately 1.4 kb of 5' sequence must have been encoded by endogenous gene sequences.

3.9.2 RACE-PCR cloning of the trapped endogenous sequence in the fusion message

To obtain the endogenous sequence contained in the fusion transcript of the GT9.6 ES cell clone, rapid amplification of cDNA ends by polymerase chain reaction (RACE-PCR) was used to amplify the 5' cDNA derived from the message encoded by the trapped gene (Figure 22). A LacZ gene specific primer was used to reverse transcribe the fusion transcript present in the GT9.6 ES cell clone, and a polyA homopolymer was then added to the 5' end of the newly synthesized first strand cDNA. PCR amplification was carried out using an outer RACE adapter primer (R_0), which binds to each cDNA at its extreme 5' end, and an *En*-2 gene specific primer (GSP_1). To provide additional specificity, a second round of amplification was performed using an inner RACE adapter primer (R_1) and a second *En*-2 gene specific primer upstream of the first one. Following two rounds of PCR amplification, two abundant fragments of the amplified 5' upstream fusion cDNA sequence were seen. These were 249 bp and 498 bp in length (Figure 24). Only the correct combination of PCR primers (R_o/GSP_1or R_0/GSP_2; R_1/GSP_2) was able to produce the amplified bands. As a control for primer sequences, the combination of primers R_0 and GSP_2 was also used in the first round of amplification and their PCR products were similarly subjected to a second round of amplification. All other combinations of PCR primers were unable to produce any detectable fragments. To verify the authenticity of these two amplified fragments, an agarose gel containing the PCR reaction products were Southern blotted as descirbed in the Materials and Methods. The blot was probed with a radioactively labeled oligo probe (GSP_3) derived from the *En*-2 sequence located further upstream of the GSP_2 primer used for the

Figure 20. Radial migration of cortical neurons in embryonic day 12.5 mouse cerebral cortex. Twelve μm thick cryostat section of cerebral cortex from an embryonic day 12.5 GT9.6 chimera is shown. Histochemical determination of β-galactosidase activity in cerebral cortex revealed a radial pattern of cell migration during cortical genesis (arrowhead). The intensified cell labeling near pia is indicated by an arrow.

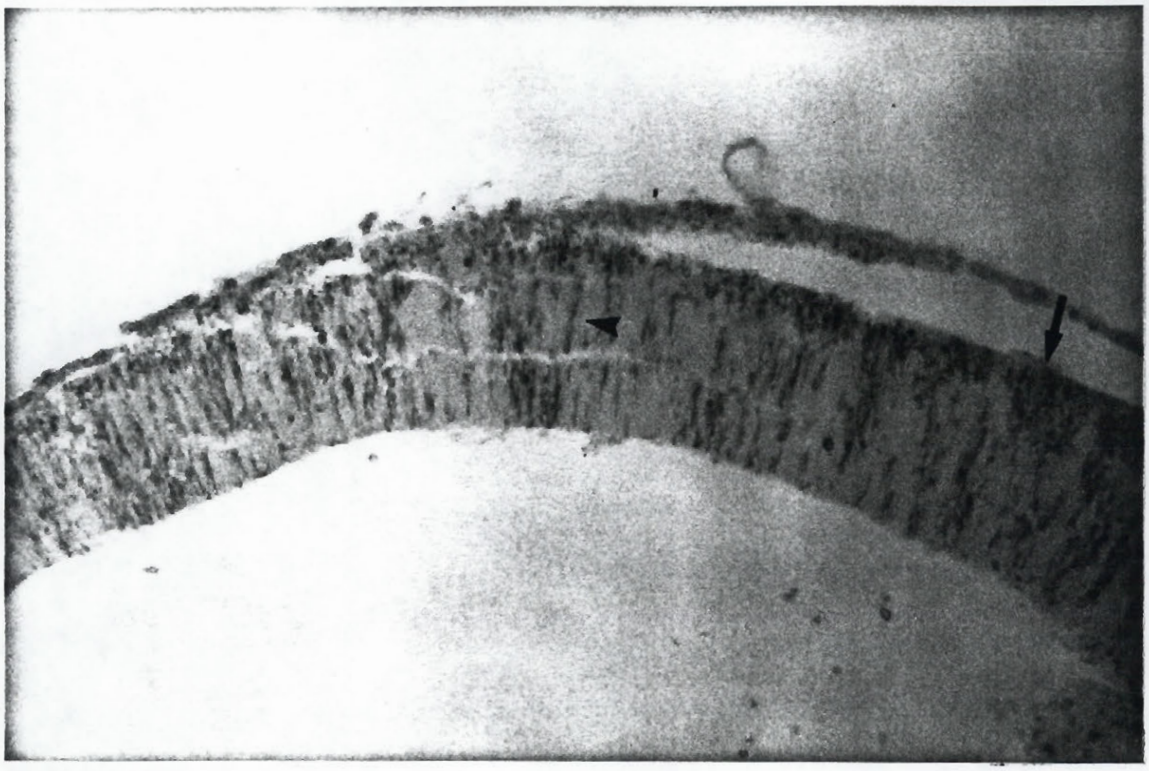
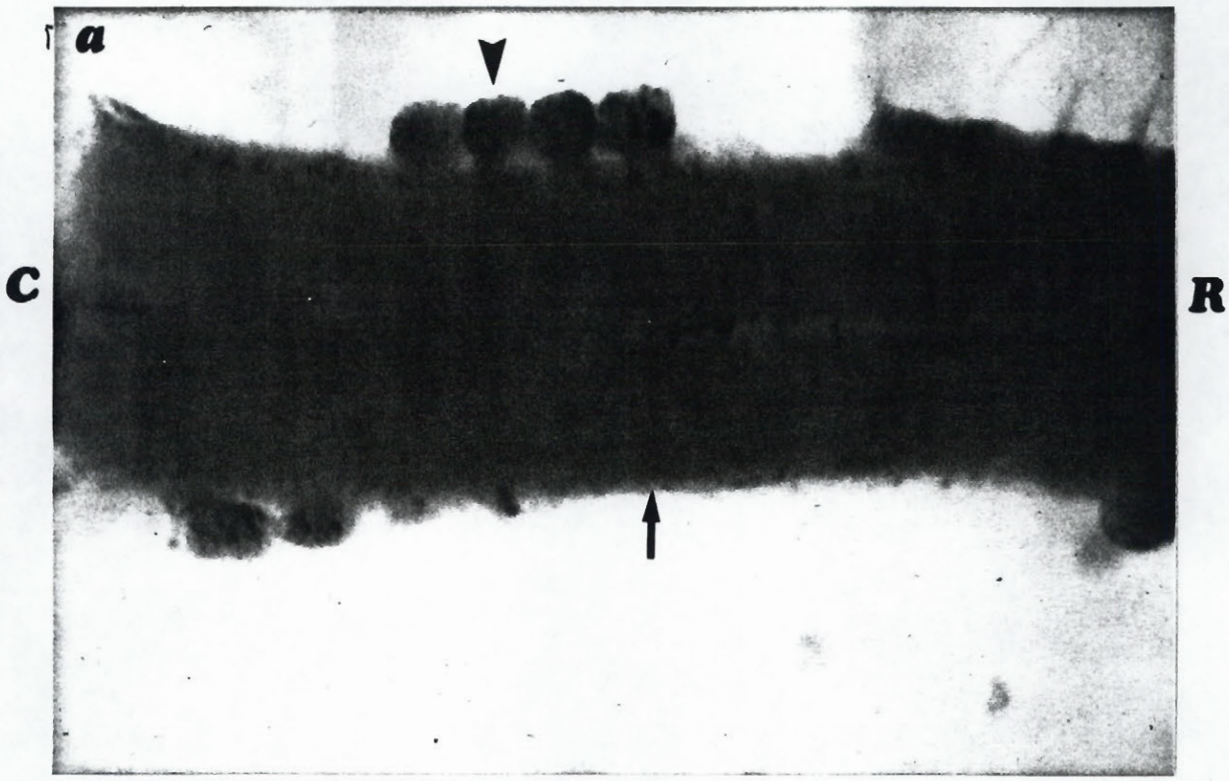


Figure 21. Spatially restricted distribution of β-galactosidase expressing spinal cord cells in the embryonic day 12.5 GT9.6 chimeras. (A) Dorsal view of the spinal cord from a day 12.5 GT9.6 embryonic chimera. (B) Lateral view of the the spinal cord from a day 12.5 GT9.6 embryonic chimera. Coherent patches of labeled spinal cord cells are indicated by arrows, and dorsal root ganglia (DRG) are indicated by arrowheads. Rostral and caudal parts of the spinal cord are indicated by r and c, respectively.


a
C
R

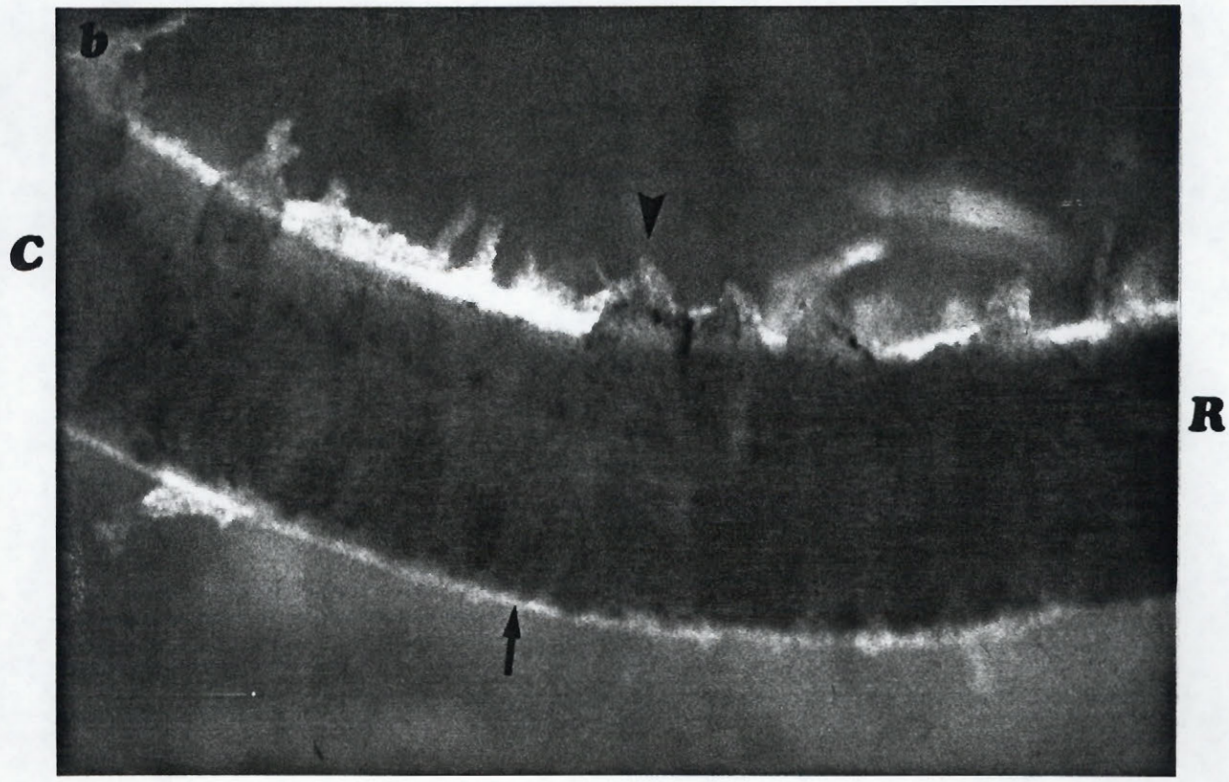

b
C
R

Figure 22. Northern blot analysis of ES cell clones. RNA was isolated from 3 gene trapped ES cell clones (Clones 1, 2, GT9.6) and the northern blot was probed with a LacZ gene probe. Each ES cell clone was found to express a fusion message of a unique size while no hybridizing transcript was detected in the control CCE ES cells. The fusion transcript encoded by the ES cell clone GT9.6 is about 4.6 kb. Since 3.2 kb sequence in this fusion transcript can be accounted for by LacZ gene and *En*-2 exon, sequences encoded by the trapped endogenous gene thus contribute to the remaining approximately 1.4 kb of the sequences.

ES cell
Clone 1
Clone 2
Clone GT9.6
28S
Fusion mRNA
18S

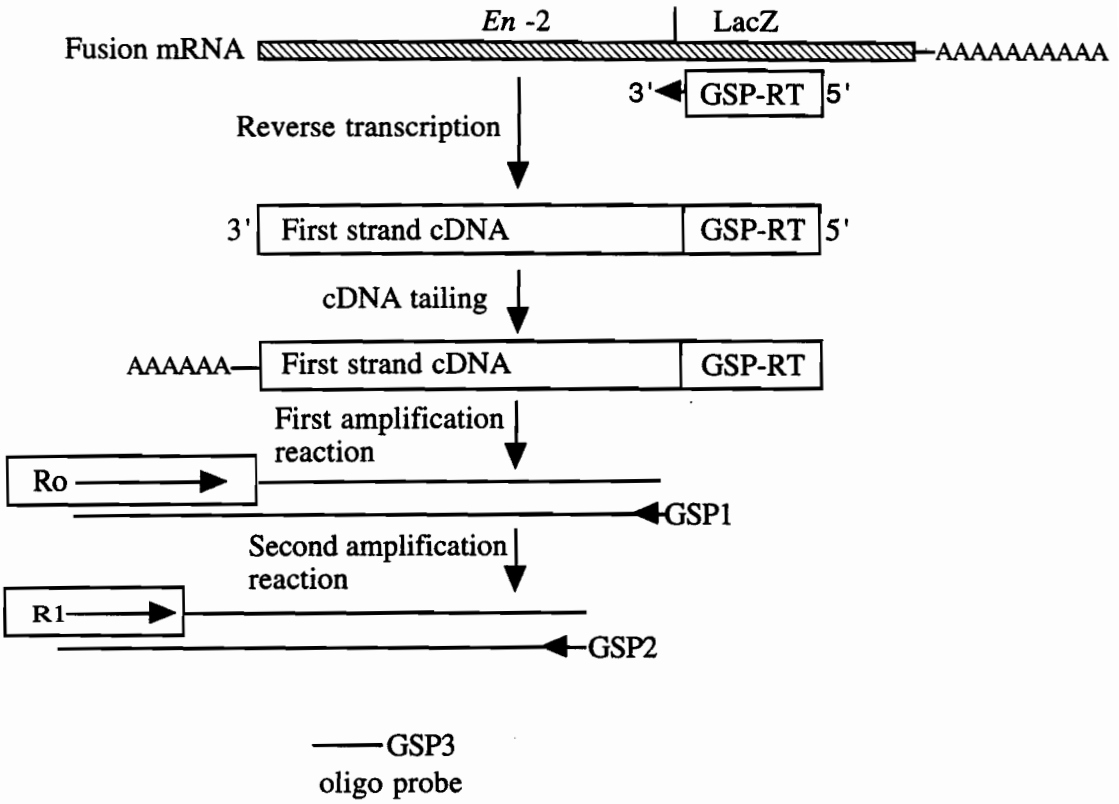
second amplification. Both the 249 bp and 498 bp PCR fragments hybridized to the oligo probe after the first and second round of PCR amplification (Figure 24).

The 249 bp and the 498 bp PCR fragments were subcloned into the TA cloning vector and subsequently sequenced. The 249 bp fragment was found to be composed of a 119 bp novel sequence upstream of the splice junction and a 130 bp *En*-2 exon sequence downstream of the splice junction (Figure 25). The 498 bp fragment appears to be a doublet of the 249 bp fragment in a head-to-head tandem repeat. The *En*-2 intron sequence present in the original gene trap construct was deleted. Therefore, as predicted, the splice junction site in the pGT4.5 gene trap construct was used to join the 119 bp novel sequence, derived from the trapped endogenous gene, to the *En*-2 exon in the gene trap construct. In the processed mRNA, the *En*-2 intron sequences were spliced out and this was reflected in the cDNA, (Figure 25).

There is only one open reading frame (ORF) in this 119 bp DNA sequence, and that ORF is in frame with the 130 bp *En*-2 sequence in the fusion cDNA (Figure 26). Since the LacZ gene did not contain a translation initiation codon and was fused in-frame to the *En*-2 exon sequence at its 5' end, this ORF thus extended beyond the *En*-2 sequence and was in frame with LacZ. Thus, the expression pattern of the trapped endogenous exon sequence is reflected by the expression profile of the β-galactosidase protein. The putative protein sequence of 39 amino acids encoded by the 119 bp RACE-PCR derived cDNA contains consensus sequences for one protein kinase C and two casein kinase II phosphorylation sites. These potential phosphorylation sites are located at amino acid positions 35, 16 and 24, respectively, relative to the first amino acid in the predicted protein sequence. There is also a potential

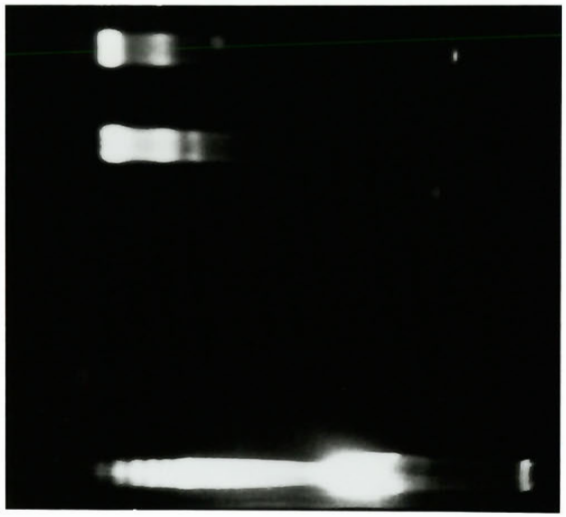
Figure 23. Schematic illustration of rapid amplification of cDNA ends by polymerase chain reaction (RACE-PCR). This procedure was used to amplify the 5' cDNA derived from the message encoded by the trapped gene. The LacZ gene specific primer (GSP-RT) was used to reverse transcribe the fusion transcript present in the GT9.6 ES cell clone, and a polyA homopolymer was then added to the 5' end of the newly synthesized first strand cDNA. PCR amplification was carried out using an outer RACE adapter primer (R_0), which binds to each cDNA at its extreme 5' end, and an *En*-2 gene specific primer (GSP_1 or GSP_2). A second round of amplification was then performed using an inner RACE adapter primer (R_1) and a second *En*-2 gene specific primer (GSP_2) upstream of the first one to provide additional specificity. A radioactively labeled oligo probe (GSP_3) derived from the *En*-2 sequence located further upstream of the GSP_2 primer was used to detect the authenticity of the PCR amplified fragments. The nucleotide sequences of the GSP-RT and various GSP primers and their relative positions within gene trap construct are illustrated in Figure 25.

RACE-PCR



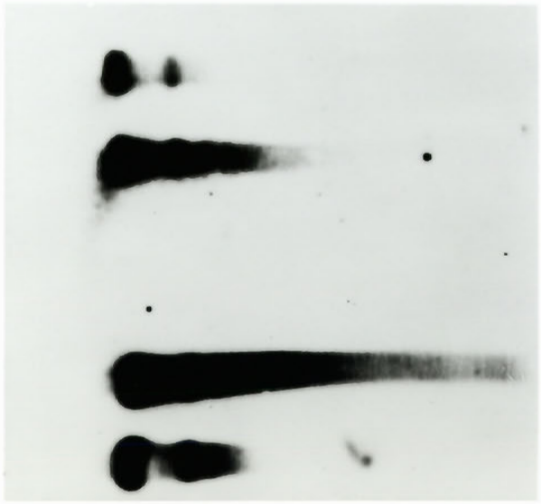
GSP=Gene Specific Primer
GSP-RT=β-galactosidase specific primer used for reverse transcription
GSP1=*En*-2 specific primer
GSP2=*En*-2 specific primer upstream of GSP1
GSP3=*En*-2 specific oligo probe upstream of GSP2
Ro=Xho-BamH1-Cla1-TTTTTTTTTTTT
R1=Xho-BamH1-Cla1

Figure 24. Ethidium bromide stained gel (left) and southern blot (right) of PCR amplification products. Two abundant fragments of the amplified 5' upstream fusion cDNA sequence were seen on a ethidium bromide stained gel after two rounds of PCR amplification. The two amplified PCR fragments were of 249 bp and 498 bp, respectively, as determined by subsequent sequencing experiments (Figure 25). This 1% agarose gel containing the PCR reaction products was blotted. The blot was probed with a radioactively labeled oligo probe (GSP_3) derived from the *En*-2 sequence located further upstream of the GSP_2 primer used for the second amplification. Although the 249 bp and 498 bp PCR fragments were visualized on the ethidium bromide stained gel only after the second round of PCR, both bands were detected by the GSP_3 oligo probe following the first as well as the second round amplification. The photograph of the ethidium bromide stained gel and the blot are presented as mirror images.



Gsp2+R1
No DNA
Gsp2+R1
Gsp2+R1
Gsp1+R1
Gsp2+Ro
No DNA
Gsp1+Ro
DNA 100 bp ladder

Second Amplification
First Amplification



DNA 100 bp ladder
Gsp1+Ro
No DNA
Gsp2+Ro
Gsp1+R1
Gsp2+R1
Gsp2+R1
No DNA
Gsp2+R1

First Amplification
Second Amplification

N-myristoylation site at amino acid position 31 (Figure 27). When this 119 bp endogenous DNA sequence was compared to DNA sequences in the NCBI Database, no significant sequence homology was found with known genes or sequence motifs.

3.9.3 cDNA cloning

To further characterize the trapped endogenous gene, I attempted to obtain cDNA clones containing more of the endogenous sequence. The 119 bp RACE-PCR derived fragment was used as a probe to screen a day 12.5 mouse embryonic cDNA library. Seven independent positive cDNA clones were obtained, and each clone contained an insert with an identical length of 735 bp (Figure 28). This insert was subcloned into the Bluescript KS^+ vector and subsequently sequenced.

The expression of the trapped gene, as reflected by the presence of the fusion protein, in the mature mouse nervous system prompted me to screen a post-natal day 18 mouse brain cDNA library with the purified 735 bp cDNA insert as the probe. Nine independent positive cDNA clones were obtained, all of which contained a 1574 bp cDNA insert. This insert was subcloned into the Bluescript KS^+ vector and subsequently sequenced.

3.9.4 Nucleotide sequence of the cDNA and the inferred amino acid sequence of the putative protein it encodes

Nucleotide sequence analysis of the 735 bp cDNA clone revealed that it contained 67 bp of nucleotide sequence at its 5' end which overlapped with the 3′ end of the 119 bp RACE-PCR derived DNA fragment (Figure 29). Thus, the remaining 668 bp of cDNA sequence are downstream of the splice junction of the *En*-2 exon present in the fusion cDNA. There is only one ORF

Figure 25. Sequence of the *En*-2/LacZ junction in the pGT4.5 gene trap construct and of the 119 bp RACE-PCR derived cDNA. (A) Sequence of the *En*-2/LacZ junction in the pGT4.5 gene trap construct. The *En*-2 splice acceptor region is shown with intron sequences in lowercase letters and exon sequences in bold uppercase letters. The promoterless LacZ coding sequence, also lacking a translation initiation codon, was fused in-frame to *En*-2 with linker sequences. The LacZ coding sequences are shown in italic uppercase letters. The sequences of primers GSP-RT, GSP_1, GSP_2 and GSP_3 used in RACE-PCR procedure are underlined. (B) The amplified 249 bp RACE-PCR fragment was sequenced. The sequence of the 119 bp endogenous cDNA sequences 5' upstream of the gene trap splice acceptor in the LacZ fusion transcript is shown in uppercase letters. The remaining 130 bp of the 249 bp RACE-PCR derived fragment is the *En*-2 exon sequence shown in bold uppercase letters.

A. *En*-2 /LacZ

En-2 intron **SA** *En*-2 exon

tgcacgcgctattctgaccggcggccttcttcag**GTCCCAGGTCCCGAAAACCAAAGAAGAAGAACC**

CTAACAAAGAGGACAAG<u>**CGGCCTCGCACAGCCTTCACTGCTGAGCAGCTCC**</u>

GSP_3

AGAGGCTCAAGGCTGAGTTTCAGA<u>**CCAACAGGTACCTGACAGAGCA**</u>**GCGGC**

GSP_2

GCCAGAGTCTGGCACAGGAGCTCGGTACCC<u>**GGAAGATCTGGACTCTAGAGG**</u>

LacZ GSP_1

ATC*CCGTCGTTT*<u>*TACAACGTCGTGACTGGGAA*</u>*AACCCTGGCGTTACCCAACTTAATCGC*

GSP-RT

CTTGCA

B. Clone GT9.6 (249 bp)

Endogenous cDNA sequences

ATGTTTTATAAGTATTGACATTACACAGTTAACAATGCATCCACAGAGCTTGGCTGAAG

AGGAAATAAAAACTGAGCAGGAGGTGGTGGAGGGAATGGATATCTCTACTCGCTCCAA

En-2 exon

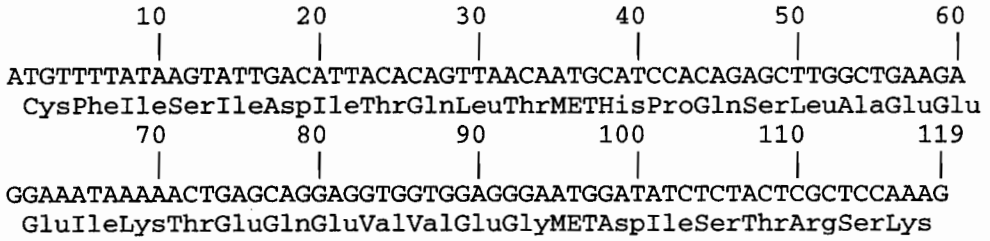
AGGTCCCAGGTCCCGAAAACCAAAGAAGAAGAACCCTAACAAAGAGGACA

AGCGGCCTCGCACAGCCTTCACTGCTGAGCAGCTCCAGAGGCTCAAGGCT

GAGTTTCAGACCAACAGGTACCTGACAGAGCA

Figure 26. The sequence of the 119 bp endogenous cDNA fragment out of the 249 bp RACE-PCR derived fragment, and the putative amino acid sequences of the proteins they respectively encode. (A) RACE-PCR derived 119 bp endogenous cDNA fragment and the predicted amino acid sequence of the protein it encodes. There is only one open reading frame (ORF) in this 119 bp sequence and that ORF is also in frame with the only ORF present in the 130 bp *En*-2-LacZ cDNA sequence derived from the fusion transcript (B). (B) The sequence of the 249 bp RACE-PCR derived fragment and its putative protein sequence. There is an in-frame fusion between the 119 bp endogenous cDNA sequence (uppercase letters) and the *En*-2 exon sequence in the fusion cDNA (italic). The ORF in the 119 bp cDNA continues in the *En*-2 exon.

A. RACE 119bp



B. RACE 119bp/En-2/LacZ (RACE 249bp)

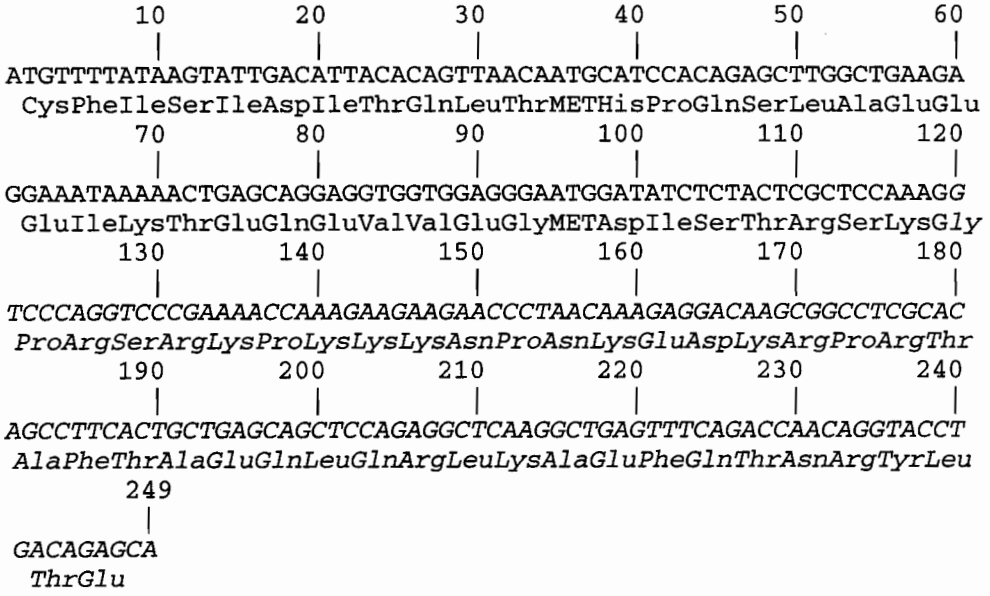


Figure 27. The putative amino acid sequence encoded by the ORF in the 119 bp RACE-PCR derived cDNA is shown. The consensus sequences for one protein kinase C and two casein kinase II phosphorylation sites at amino acid positions 35, 16 and 24, respectively, are indicated. There is also a potential N-myristoylation site at amino acid position 31. The position of the first amino acid (Cys) in the putative protein sequence is designated 1 in this figure.

Protein kinase C phosphorylation site found at 35 : egmdi S trsk

==========

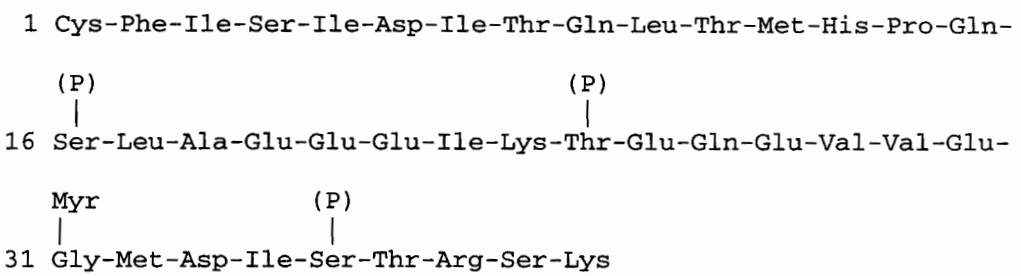
Number of potential casein kinase II phosphorylation sites found: 2

[1] 16 : tmhpq S laeee

[2] 24 : eeeik T eqevv

==========

N-myristoylation site found at 31 : qevve G mdist



Abbreviations:

Myr : stands for a myristyl group.
(P) : stands for a phosphate group.

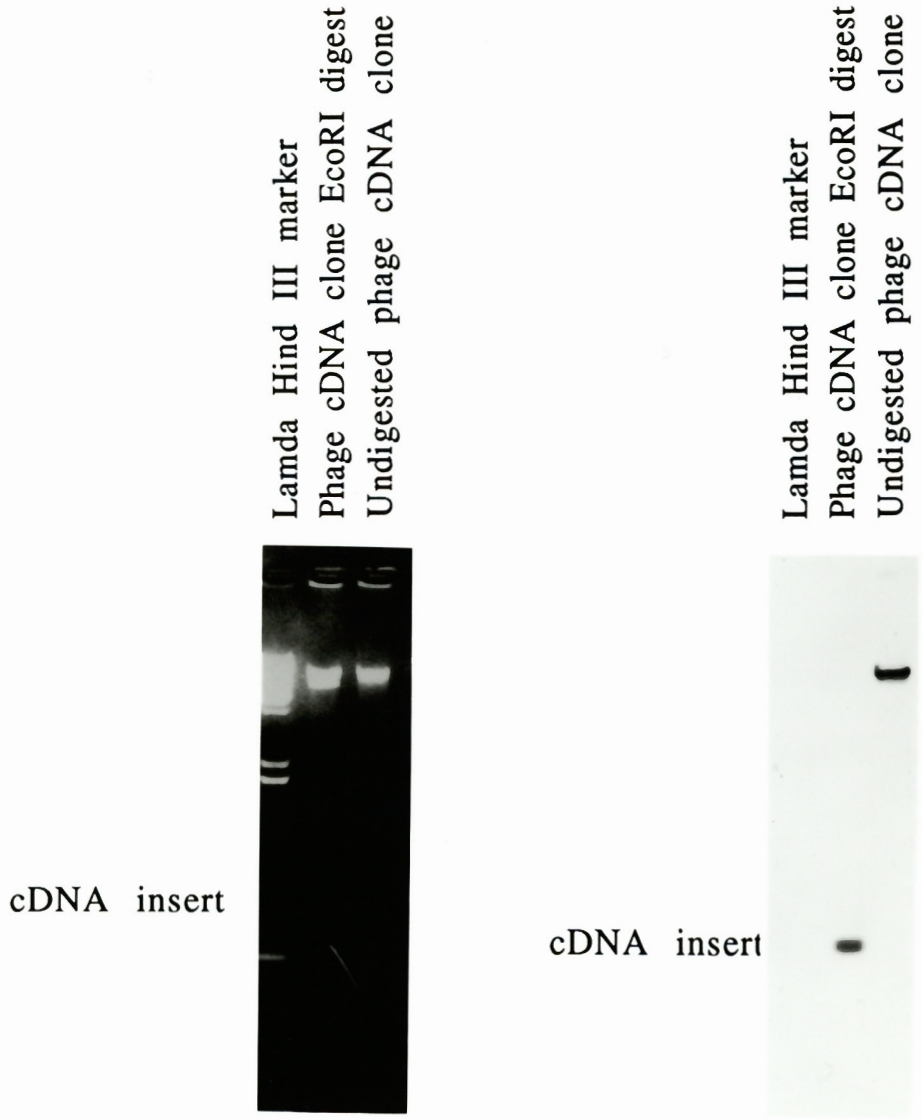
in this cDNA sequence and the first stop codon of that ORF occurred at nucleotide 386 (Figure 30). The other two reading frames have multiple stop codons at the beginning of the cDNA clone. The ORF identified in the 119 bp RACE-PCR derived cDNA sequence continued in this cDNA clone. Thus, it appears that the gene trap insertion disrupted the normal coding sequences of the trapped endogenous gene and created an insertional mutation in the form of a fusion transcript. In addition, it indicated that the splice acceptor site in the pGT4.5 construct was used correctly.

When the protein sequences encoded by the ORF in the 119 bp RACE-PCR derived cDNA or in the 735 bp cDNA clone were examined, no nuclear localization signals were discovered. Given that the fusion protein is localized in the neuronal nuclei, and that only 119 bp out of approximately 1 kb of sequence upstream of the *En*-2 splice junction in the fusion message was identified, it appears likely that the nuclear localization signal in the fusion protein is encoded by the remaining 881 bp message positioned 5' to the 119 bp cDNA sequence.

When the putative protein sequences encoded by the 735 bp cDNA was searched in NCBI Genbank, no protein or functional domains with significant sequence homologies were obtained. However, the putative protein sequence of the 735 bp cDNA clone contains consensus sequences for three protein kinase C and three casein kinase II phosphorylation sites at amino acid positions, 27, 68, 109 and 47, 68, 109, respectively, in addition to the protein kinase C and the casein kinase II potential phosphorylation sites respectively located at amino acid position 18 and 7, which are identical to that found in the overlapping region of the 119 bp RACE-PCR derived DNA (amino acid 24 and 35, respectively) (Figure 27, 31). There is also an additional potential N-

Figure 28. Identification of the phage cDNA clone by TAE agarose gel electrophoresis (left) and Southern blot analysis (right). The 119 bp RACE-PCR amplified fragment was used as a probe to screen a day 12.5 mouse embryonic cDNA library. Seven independent positive cDNA clones were obtained and each clone contained an insert of 735 bp, as determined by subsequent DNA sequencing. Phage DNA from one positive clone was digested with EcoR I and electrophoresed on a 1% TAE agarose gel. A 735 bp cDNA insert was released from the phage DNA following EcoR I digestion. The electrophoresed DNA was blotted, and the filter was probed with the 119 bp RACE-PCR derived DNA probe to confirm the authenticity of the cDNA insert. The uncut phage DNA containing the insert hybridized to the 119 bp DNA probe. Only the 735 bp cDNA insert, but not the cut phage DNA, hybridized to the 119 bp DNA probe following EcoR I digestion.

Phage cDNA clone identification



myristoylation site located at amino acid position 63, besides the N-myristoylation site found at amino acid 14, which is identical to the one identified in the overlapping region of the 119 bp RACE-PCR derived DNA (amino acid 31) (Figure 29). The presence of multiple phosphorylation sites in the putative peptide sequence suggests that the endogenous gene may encode a phosphorylated protein, and that the peptide sequence is close to the C-terminus of the putative protein. No genes with significant sequence homology were found when the sequence of this 735 bp cDNA clone was compared with the NCBI nucelotide Database.

When the sequence of the 735 bp and 1574 bp cDNA clones were examined, a 248 bp nucleotide sequence overlap was found between the 3' end of the 735 bp cDNA clone and the 5' end of the 1574 bp cDNA clone, with two T/A and A/G mismatches at nucleotide position 430 (T) in the 735 bp cDNA and 38 (A) in the 1574 bp cDNA as well as 594 (A) in the 735 bp cDNA and 111 (G) in the 551 bp cDNA. Since the 735 bp cDNA was cloned from a CD1 mouse embryonic cDNA library, whereas the 1574 bp cDNA was cloned from a NIH Swiss mouse brain cDNA library. the two nucleotide mismatches between the two cDNAs could be due to the different mouse strains from which the two cDNA libraries were made. However, the 735 bp cDNA clone contains four unique nucleotides (CCAT) at its most 3' end that diverge from the 1574 bp cDNA sequence (Figure 32). Thus, these two cDNAs may be derived from RNAs produced by alternative splicing. There is only one ORF in the 1574 bp cDNA sequence and the first stop codon of that ORF occurred at nucleotide 1049 (Figure 32). This ORF from the first methionine is predicted to encode a protein of 349 amino acids. In favor of the third in-frame methionine, at amino acid position 46, as the translation start site is the fact that its sequence

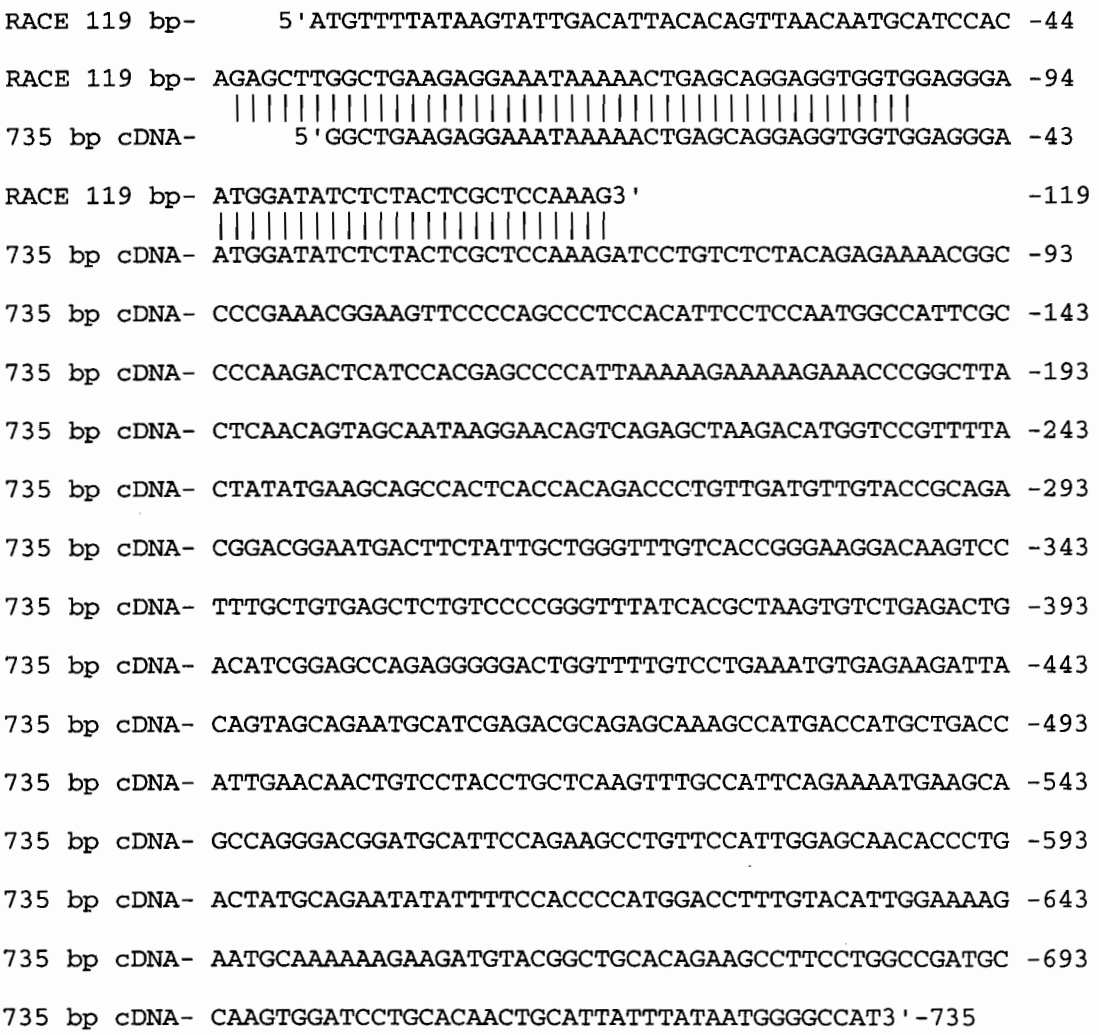

RACE 119 bp- 5'ATGTTTTATAAGTATTGACATTACACAGTTAACAATGCATCCAC -44
RACE 119 bp- AGAGCTTGGCTGAAGAGGAAATAAAAACTGAGCAGGAGGTGGTGGAGGGA -94
735 bp cDNA- 5'GGCTGAAGAGGAAATAAAAACTGAGCAGGAGGTGGTGGAGGGA -43
RACE 119 bp- ATGGATATCTCTACTCGCTCCAAAG3' -119
735 bp cDNA- ATGGATATCTCTACTCGCTCCAAAGATCCTGTCTCTACAGAGAAAACGGC -93
735 bp cDNA- CCCGAAACGGAAGTTCCCCAGCCCTCCACATTCCTCCAATGGCCATTCGC -143
735 bp cDNA- CCCAAGACTCATCCACGAGCCCCATTAAAAAGAAAAAGAAACCCGGCTTA -193
735 bp cDNA- CTCAACAGTAGCAATAAGGAACAGTCAGAGCTAAGACATGGTCCGTTTTA -243
735 bp cDNA- CTATATGAAGCAGCCACTCACCACAGACCCTGTTGATGTTGTACCGCAGA -293
735 bp cDNA- CGGACGGAATGACTTCTATTGCTGGGTTTGTCACCGGGAAGGACAAGTCC -343
735 bp cDNA- TTTGCTGTGAGCTCTGTCCCCGGGTTTATCACGCTAAGTGTCTGAGACTG -393
735 bp cDNA- ACATCGGAGCCAGAGGGGGACTGGTTTTGTCCTGAAATGTGAGAAGATTA -443
735 bp cDNA- CAGTAGCAGAATGCATCGAGACGCAGAGCAAAGCCATGACCATGCTGACC -493
735 bp cDNA- ATTGAACAACTGTCCTACCTGCTCAAGTTTGCCATTCAGAAAATGAAGCA -543
735 bp cDNA- GCCAGGGACGGATGCATTCCAGAAGCCTGTTCCATTGGAGCAACACCCTG -593
735 bp cDNA- ACTATGCAGAATATATTTTCCACCCCATGGACCTTTGTACATTGGAAAAG -643
735 bp cDNA- AATGCAAAAAAGAAGATGTACGGCTGCACAGAAGCCTTCCTGGCCGATGC -693
735 bp cDNA- CAAGTGGATCCTGCACAACTGCATTATTTATAATGGGGCCAT3'-735

Figure 30. The sequence of the 735 bp cDNA clone and its putative protein sequence are illustrated. There is only one open reading frame (ORF) in this cDNA sequence and the first stop codon of that ORF occurred at nucleotide 386. The stop codon (TGA) is shown in bold letters. The other two reading frames have multiple stop codons at the beginning of the cDNA clone. The ORF identified in the 119 bp RACE-PCR derived cDNA sequence continues in this cDNA clone. The shared ORF is underlined.

735 bp cDNA clone

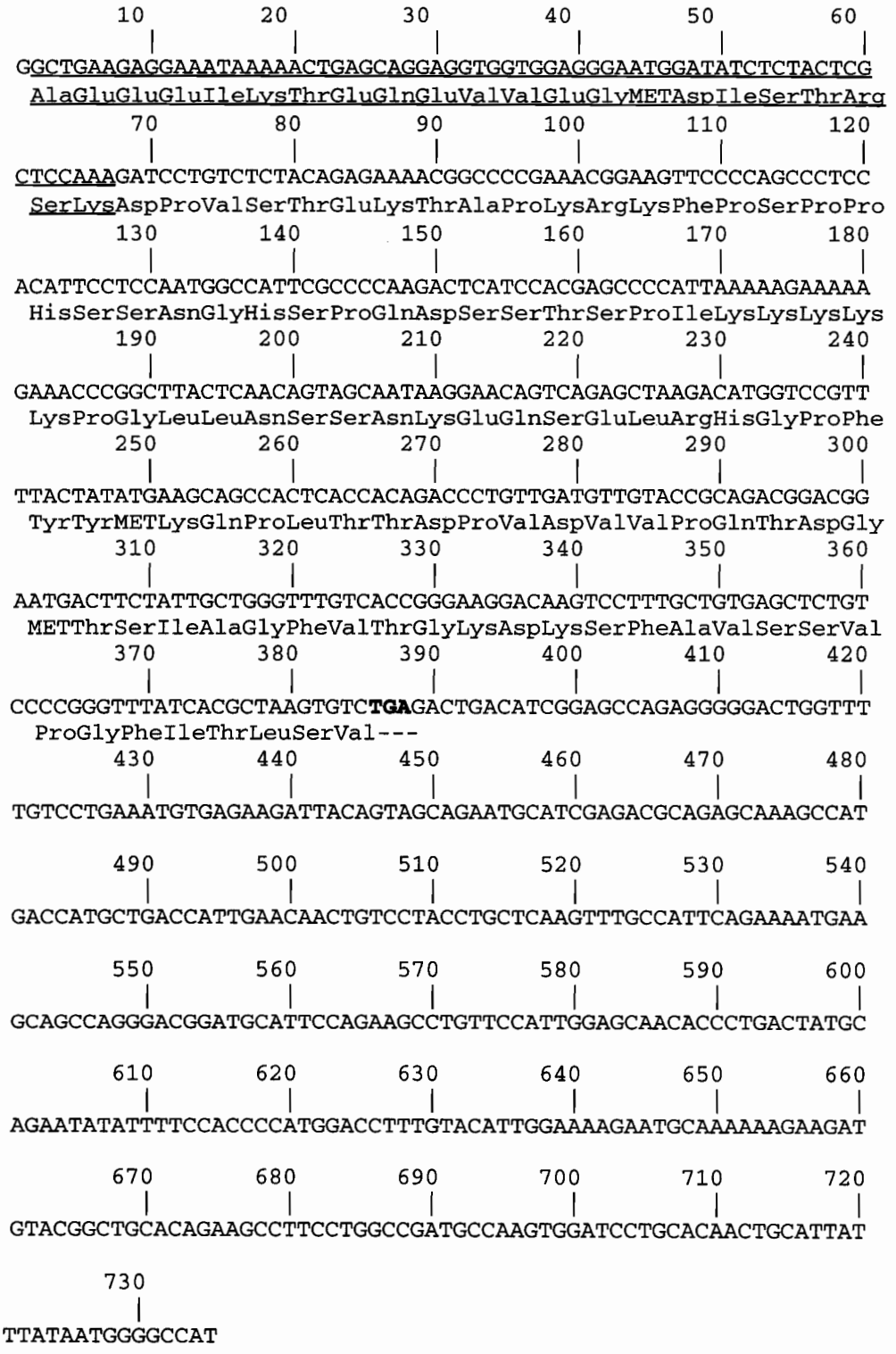


Figure 31. The putative amino acid sequence encoded by the 735 bp cDNA clone is shown. The consensus sequences for four protein kinase C and four casein kinase II phosphorylation sites at amino acid positions 18, 27, 68, 109 and 7, 47, 68, 109, respectively, are indicated. There are also two potential N-myristoylation sites at amino acid positions 14 and 63, respectively, and one N-glycosylation site at amino acid position 66. The position of the first amino acid (Ala) in the putative protein sequence is designated 1 in this figure.

N-glycosylation site found at 66 : kpgll N ssnke

==========

Number of potential protein kinase C phosphorylation sites found: 4

[1] 18 : egmdi S trskd

[2] 27 : kdpvs T ektap

[3] 68 : gllns S nkeqs

[4] 109 : iagfv T gkdks

==========

Number of potential casein kinase II phosphorylation sites found: 4

[1] 7 : eeeik T eqevv

[2] 47 : ssngh S pqdss

[3] 68 : gllns S nkeqs

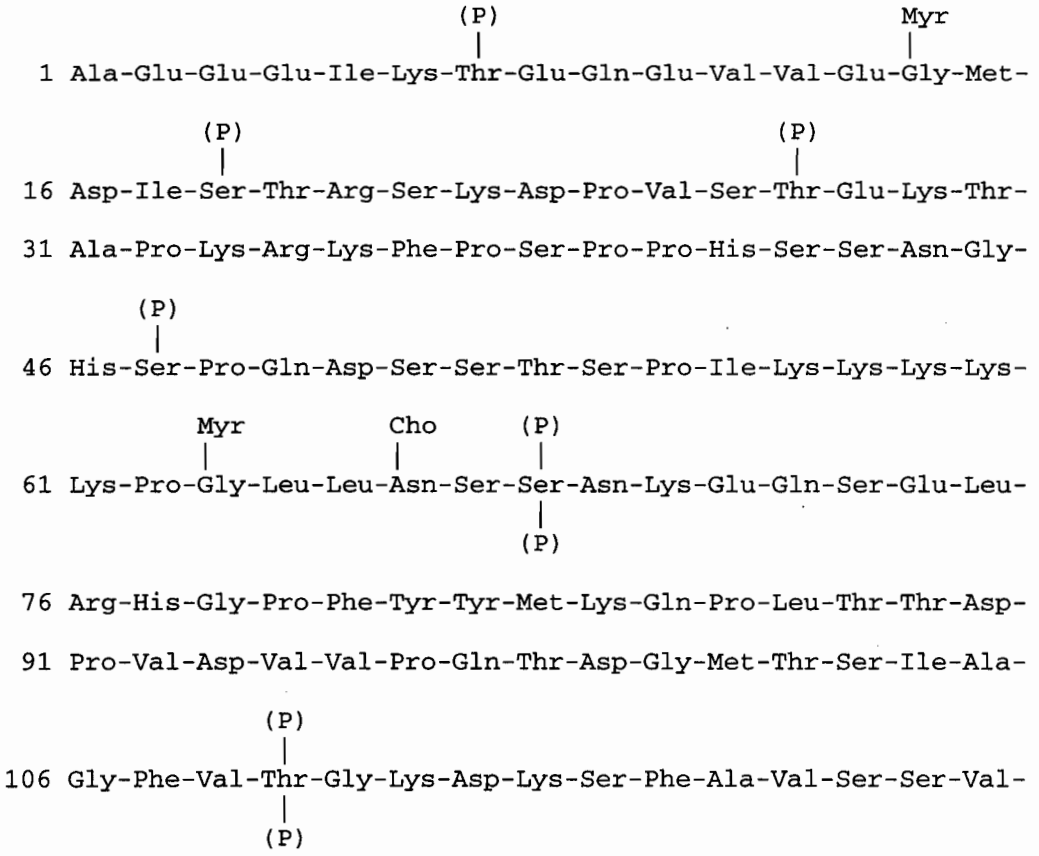
[4] 109 : iagfv T gkdks

==========

Number of potential N-myristoylation sites found: 2

[1] 14 : qevve G mdist

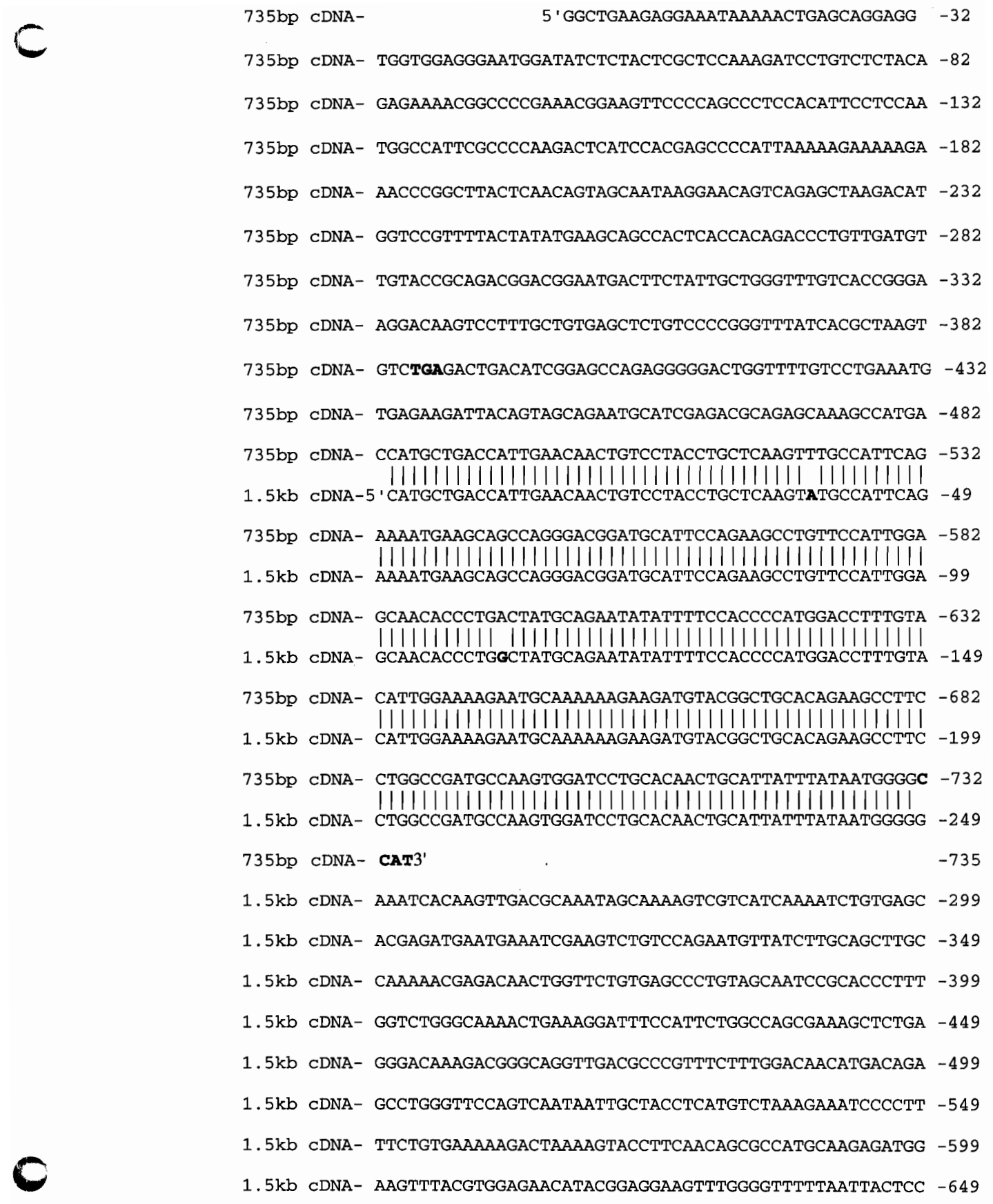
[2] 63 : kkkkp G llnss



Abbreviations:

Cho : stands for a carbohydrate group.
Myr : stands for a myristyl group.
(P) : stands for a phosphate group.

Figure 32. Comparison of the nucleotide sequence of the 735 bp cDNA clone with the 1574 bp cDNA clone. There is a 248 bp nucleotide sequence overlap between the 3' end of the 735 bp cDNA clone and the 5' end of the 1574 bp cDNA clone, with two T/A and A/G mismatches. The two mismatches in the 1574 bp cDNA sequences occurred at nucleotide position 38 (A) and 111 (G) (bold). The four nucleotides CCAT (bold) at the most 3' end of the 735 bp cDNA clone diverge from the 1574 bp cDNA sequence. The stop codon TGA (bold) in the 735 bp cDNA clone appeared before the overlapping region. Thus, the ORF identified in the 1574 bp cDNA clone is independent of that of the 735 bp cDNA clone.


735bp cDNA- 5'GGCTGAAGAGGAAATAAAAACTGAGCAGGAGG -32
735bp cDNA- TGGTGGAGGGAATGGATATCTCTACTCGCTCCAAAGATCCTGTCTCTACA -82
735bp cDNA- GAGAAAACGGCCCCGAAACGGAAGTTCCCCAGCCCTCCACATTCCTCCAA -132
735bp cDNA- TGGCCATTCGCCCCAAGACTCATCCACGAGCCCCATTAAAAAGAAAAAGA -182
735bp cDNA- AACCCGGCTTACTCAACAGTAGCAATAAGGAACAGTCAGAGCTAAGACAT -232
735bp cDNA- GGTCCGTTTTACTATATGAAGCAGCCACTCACCACAGACCCTGTTGATGT -282
735bp cDNA- TGTACCGCAGACGGACGGAATGACTTCTATTGCTGGGTTTGTCACCGGGA -332
735bp cDNA- AGGACAAGTCCTTTGCTGTGAGCTCTGTCCCCGGGTTTATCACGCTAAGT -382
735bp cDNA- GTCTGAGACTGACATCGGAGCCAGAGGGGGACTGGTTTTGTCCTGAAATG -432
735bp cDNA- TGAGAAGATTACAGTAGCAGAATGCATCGAGACGCAGAGCAAAGCCATGA -482
735bp cDNA- CCATGCTGACCATTGAACAACTGTCCTACCTGCTCAAGTTTGCCATTCAG -532
1.5kb cDNA-5'CATGCTGACCATTGAACAACTGTCCTACCTGCTCAAGTATGCCATTCAG -49
735bp cDNA- AAAATGAAGCAGCCAGGGACGGATGCATTCCAGAAGCCTGTTCCATTGGA -582
1.5kb cDNA- AAAATGAAGCAGCCAGGGACGGATGCATTCCAGAAGCCTGTTCCATTGGA -99
735bp cDNA- GCAACACCCTGACTATGCAGAATATATTTTCCACCCCATGGACCTTTGTA -632
1.5kb cDNA- GCAACACCCTGGCTATGCAGAATATATTTTCCACCCCATGGACCTTTGTA -149
735bp cDNA- CATTGGAAAAGAATGCAAAAAAGAAGATGTACGGCTGCACAGAAGCCTTC -682
1.5kb cDNA- CATTGGAAAAGAATGCAAAAAAGAAGATGTACGGCTGCACAGAAGCCTTC -199
735bp cDNA- CTGGCCGATGCCAAGTGGATCCTGCACAACTGCATTATTTATAATGGGGC -732
1.5kb cDNA- CTGGCCGATGCCAAGTGGATCCTGCACAACTGCATTATTTATAATGGGGG -249
735bp cDNA- CAT3' -735
1.5kb cDNA- AAATCACAAGTTGACGCAAATAGCAAAAGTCGTCATCAAAATCTGTGAGC -299
1.5kb cDNA- ACGAGATGAATGAAATCGAAGTCTGTCCAGAATGTTATCTTGCAGCTTGC -349
1.5kb cDNA- CAAAAACGAGACAACTGGTTCTGTGAGCCCTGTAGCAATCCGCACCCTTT -399
1.5kb cDNA- GGTCTGGGCAAAACTGAAAGGATTTCCATTCTGGCCAGCGAAAGCTCTGA -449
1.5kb cDNA- GGGACAAAGACGGGCAGGTTGACGCCCGTTTCTTTGGACAACATGACAGA -499
1.5kb cDNA- GCCTGGGTTCCAGTCAATAATTGCTACCTCATGTCTAAAGAAATCCCCTT -549
1.5kb cDNA- TTCTGTGAAAAAGACTAAAAGTACCTTCAACAGCGCCATGCAAGAGATGG -599
1.5kb cDNA- AAGTTTACGTGGAGAACATACGGAGGAAGTTTGGGGTTTTTAATTACTCC -649

1.5kb cDNA- CCGTTCAGGACGCCCTACACGCCCAACAACCAGTACCAAATGCTGCTGGA -699

1.5kb cDNA- TCCCAGCAACCCCAGCGCGGGCACAGCCAAGACAGACAAACAGGAGAAGG -749

1.5kb cDNA- TGAAGCTTAATTTTGACATGACAGCGTCCCCCAAGATCCTTCTGAGCAAG -799

1.5kb cDNA- CCCCTTCTGAGCGGGGGTGCCGGCCGCAGGATCTCCCTGTCCGACATGCC -849

1.5kb cDNA- TCGCTCCCCCACTACGAACTCTTCCGTGCACACGGGCTCCGATGTGGAGC -899

1.5kb cDNA- AGGACCCCGAGAAGAAGGCCCCGTCCAGCCACTTCAGCGCAAGCGAGGAG -949

1.5kb cDNA- TCCATGGACTTCCTTGATAAGAGCACAGGTCAGCCCCAGTTGGGAGAGTT -999

1.5kb cDNA- GGGTAGGAGGCCACTTGGTGCATTCTGGCGTCCCTGGATGCCTCCCTCAT -1049

1.5kb cDNA- AGGTGCGAGCCGGCAGGCACCACCAGATCCCACAGATGAGGGAGATAGTC -1099

1.5kb cDNA- CCCAAGGAAGGCCAGGTGGGGCCATTCTTAGAAGCGGATGGCAATGGCCG -1149

1.5kb cDNA- GCCACAAACCCTGCAGGGCTCGAGGCCACTGTCAGAGCCTCATGGCTGTT -1199

1.5kb cDNA- CGCAGCAGTGCCCTTGACTCTTTTCTTCTCATCACGGTGTAGTTGAAAGC -1249

1.5kb cDNA- TTAGTTTTTGATACCTTTGTGTTAGGGTTGACTTAGACGTGAGGAGACTG -1299

1.5kb cDNA- AGGTGAAAATTGATCCCCCTTGCCTCCTTCAGGGCTGGCGTTTTCCAGCT -1349

1.5kb cDNA- ATGTGTCGTAGTTTGGAGACAACAACTGGTCTCTAATTGTAGGACTCCCC -1399

1.5kb cDNA- ATCAGTTGATGTGGCTTTGGGGACATATTTAATGTCTTTATGCCTTAGCC -1449

1.5kb cDNA- TTCCCACCTCTACCATGGCGACCGTAACTGGCTACGCCGAGCCTCCCCTA -1499

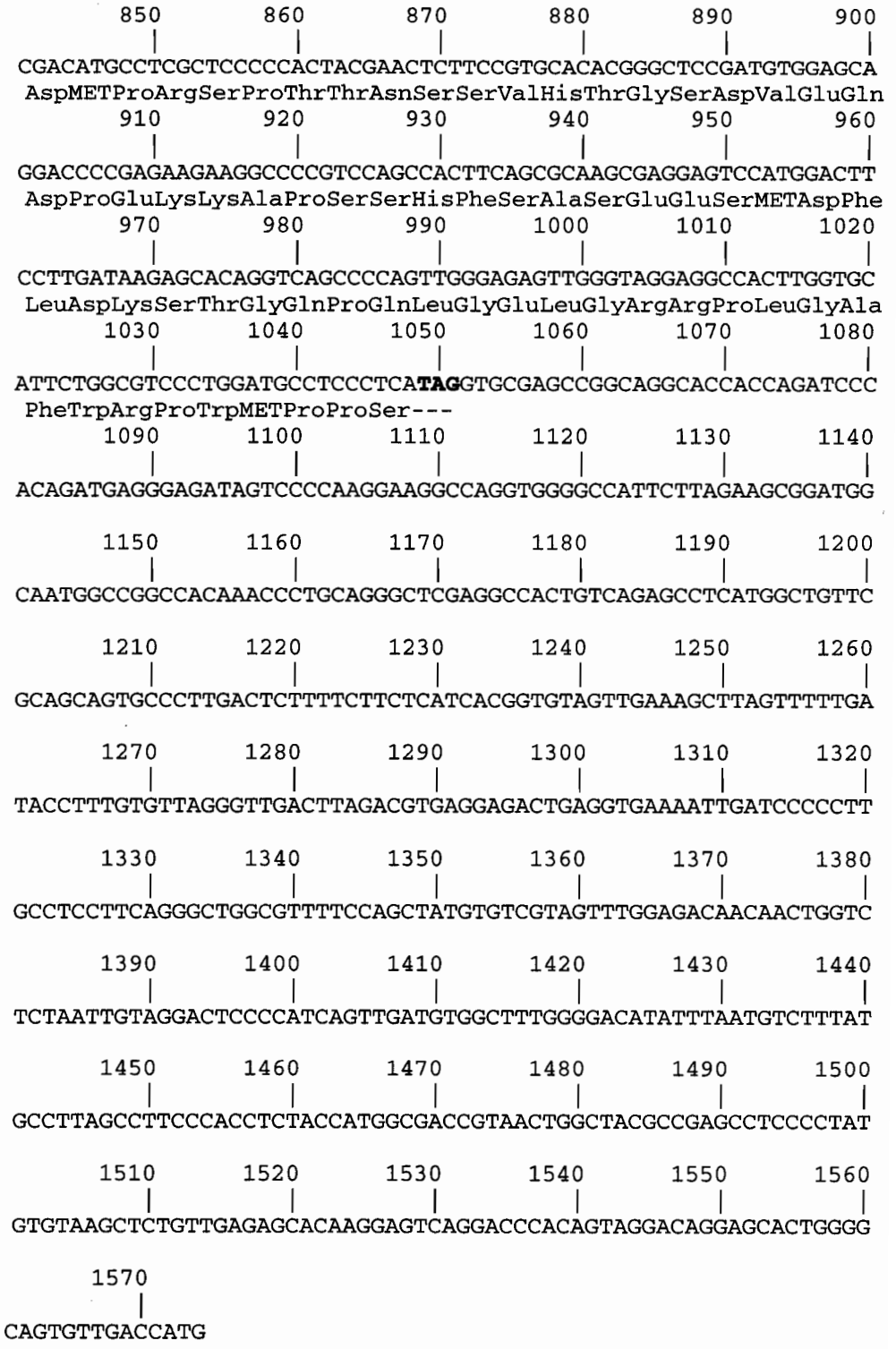
1.5kb cDNA- TGTGTAAGCTCTGTTGAGAGCACAAGGAGTCAGGACCCACAGTAGGACAG -1549

1.5kb cDNA- GAGCACTGGGGCAGTGTTGACCATG3' -1574

Figure 33. The sequence of the 1574 bp cDNA clone and the predicted amino acid sequence it encodes are shown. There is only one ORF in this cDNA sequence and the first stop codon in that ORF occurs at nucleotide 1049. The stop codon is shown in bold letters (TAG). The ORF is shown encoding 349 amino acids from the first methionine. The context of this first AUG start codon does not match the Kozak eukaryotic translation initiation sequence ($CC^{A}/_{G}CCAUGG(G)$), whereas the third AUG is in significantly better context (underlined) and may represent the N terminus of the protein. The other two reading frames have multiple stop codons at the beginning of the cDNA clone.

1574 bp cDNA clone

```
        10         20         30         40         50         60
         |          |          |          |          |          |
CATGCTGACCATTGAACAACTGTCCTACCTGCTCAAGTATGCCATTCAGAAAATGAAGCA
METLeuThrIleGluGlnLeuSerTyrLeuLeuLysTyrAlaIleGlnLysMETLysGln
        70         80         90        100        110        120
         |          |          |          |          |          |
GCCAGGGACGGATGCATTCCAGAAGCCTGTTCCATTGGAGCAACACCCTGGCTATGCAGA
ProGlyThrAspAlaPheGlnLysProValProLeuGluGlnHisProGlyTyrAlaGlu
       130        140        150        160        170        180
         |          |          |          |          |          |
ATATATTTTCCACCCCATGGACCTTTGTACATTGGAAAAGAATGCAAAAAAGAAGATGTA
TyrIlePheHisProMETAspLeuCysThrLeuGluLysAsnAlaLysLysLysMETTyr
       190        200        210        220        230        240
         |          |          |          |          |          |
CGGCTGCACAGAAGCCTTCCTGGCCGATGCCAAGTGGATCCTGCACAACTGCATTATTTA
GlyCysThrGluAlaPheLeuAlaAspAlaLysTrpIleLeuHisAsnCysIleIleTyr
       250        260        270        280        290        300
         |          |          |          |          |          |
TAATGGGGGAAATCACAAGTTGACGCAAATAGCAAAAGTCGTCATCAAAATCTGTGAGCA
AsnGlyGlyAsnHisLysLeuThrGlnIleAlaLysValValIleLysIleCysGluHis
       310        320        330        340        350        360
         |          |          |          |          |          |
CGAGATGAATGAAATCGAAGTCTGTCCAGAATGTTATCTTGCAGCTTGCCAAAAACGAGA
GluMETAsnGluIleGluValCysProGluCysTyrLeuAlaAlaCysGlnLysArgAsp
       370        380        390        400        410        420
         |          |          |          |          |          |
CAACTGGTTCTGTGAGCCCTGTAGCAATCCGCACCCTTTGGTCTGGGCAAAACTGAAAGG
AsnTrpPheCysGluProCysSerAsnProHisProLeuValTrpAlaLysLeuLysGly
       430        440        450        460        470        480
         |          |          |          |          |          |
ATTTCCATTCTGGCCAGCGAAAGCTCTGAGGGACAAAGACGGGCAGGTTGACGCCCGTTT
PheProPheTrpProAlaLysAlaLeuArgAspLysAspGlyGlnValAspAlaArgPhe
       490        500        510        520        530        540
         |          |          |          |          |          |
CTTTGGACAACATGACAGAGCCTGGGTTCCAGTCAATAATTGCTACCTCATGTCTAAAGA
PheGlyGlnHisAspArgAlaTrpValProValAsnAsnCysTyrLeuMETSerLysGlu
       550        560        570        580        590        600
         |          |          |          |          |          |
AATCCCCTTTTCTGTGAAAAAGACTAAAAGTACCTTCAACAGCGCCATGCAAGAGATGGA
IleProPheSerValLysLysThrLysSerThrPheAsnSerAlaMETGlnGluMETGlu
       610        620        630        640        650        660
         |          |          |          |          |          |
AGTTTACGTGGAGAACATACGGAGGAAGTTTGGGGTTTTTAATTACTCCCCGTTCAGGAC
ValTyrValGluAsnIleArgArgLysPheGlyValPheAsnTyrSerProPheArgThr
       670        680        690        700        710        720
         |          |          |          |          |          |
GCCCTACACGCCCAACAACCAGTACCAAATGCTGCTGGATCCCAGCAACCCCAGCGCGGG
ProTyrThrProAsnAsnGlnTyrGlnMETLeuLeuAspProSerAsnProSerAlaGly
       730        740        750        760        770        780
         |          |          |          |          |          |
CACAGCCAAGACAGACAAACAGGAGAAGGTGAAGCTTAATTTTGACATGACAGCGTCCCC
ThrAlaLysThrAspLysGlnGluLysValLysLeuAsnPheAspMETThrAlaSerPro
       790        800        810        820        830        840
         |          |          |          |          |          |
CAAGATCCTTCTGAGCAAGCCCCTTCTGAGCGGGGGTGCCGGCCGCAGGATCTCCCTGTC
LysIleLeuLeuSerLysProLeuLeuSerGlyGlyAlaGlyArgArgIleSerLeuSer
```


850 860 870 880 890 900
CGACATGCCTCGCTCCCCCACTACGAACTCTTCCGTGCACACGGGCTCCGATGTGGAGCA
AspMETProArgSerProThrThrAsnSerSerValHisThrGlySerAspValGluGln
910 920 930 940 950 960
GGACCCCGAGAAGAAGGCCCCGTCCAGCCACTTCAGCGCAAGCGAGGAGTCCATGGACTT
AspProGluLysLysAlaProSerSerHisPheSerAlaSerGluGluSerMETAspPhe
970 980 990 1000 1010 1020
CCTTGATAAGAGCACAGGTCAGCCCCAGTTGGGAGAGTTGGGTAGGAGGCCACTTGGTGC
LeuAspLysSerThrGlyGlnProGlnLeuGlyGluLeuGlyArgArgProLeuGlyAla
1030 1040 1050 1060 1070 1080
ATTCTGGCGTCCCTGGATGCCTCCCTCATAGGTGCGAGCCGGCAGGCACCACCAGATCCC
PheTrpArgProTrpMETProProSer---
1090 1100 1110 1120 1130 1140
ACAGATGAGGGAGATAGTCCCCAAGGAAGGCCAGGTGGGGCCATTCTTAGAAGCGGATGG
1150 1160 1170 1180 1190 1200
CAATGGCCGGCCACAAACCCTGCAGGGCTCGAGGCCACTGTCAGAGCCTCATGGCTGTTC
1210 1220 1230 1240 1250 1260
GCAGCAGTGCCCTTGACTCTTTTCTTCTCATCACGGTGTAGTTGAAAGCTTAGTTTTTGA
1270 1280 1290 1300 1310 1320
TACCTTTGTGTTAGGGTTGACTTAGACGTGAGGAGACTGAGGTGAAAATTGATCCCCCTT
1330 1340 1350 1360 1370 1380
GCCTCCTTCAGGGCTGGCGTTTTCCAGCTATGTGTCGTAGTTTGGAGACAACAACTGGTC
1390 1400 1410 1420 1430 1440
TCTAATTGTAGGACTCCCCATCAGTTGATGTGGCTTTGGGGACATATTTAATGTCTTTAT
1450 1460 1470 1480 1490 1500
GCCTTAGCCTTCCCACCTCTACCATGGCGACCGTAACTGGCTACGCCGAGCCTCCCCTAT
1510 1520 1530 1540 1550 1560
GTGTAAGCTCTGTTGAGAGCACAAGGAGTCAGGACCCACAGTAGGACAGGAGCACTGGGG
1570
CAGTGTTGACCATG

environment matches the Kozak consensus sequence more closely than the upstream methionine. The other two reading frames in the 1574 cDNA clone have multiple stop codons at the beginning of the cDNA clone. The stop codon (TGA) in the 735 bp cDNA clone appeared before the overlapping region (Figure 31). Thus, the ORF identified in the 1574 bp cDNA clone is independent of that of the 735 bp cDNA clone.

The putative protein sequence of the 1574 bp cDNA clone contains consensus sequences for one cAMP and cGMP-dependent protein kinase phosphorylation site located at amino acid position 278 (Met 1). There are also consensus sequences for four protein kinase C and four casein kinase II phosphorylation sites found at amino acid positions 184, 241, 244, 259 and 278, 294, 296 312, respectively. One potential N-glycosylation, three potential N-myristoylation and two potential amidation sites are positioned at amino acid 289, 61, 154, 240 and 273, 333, respectively (Figure 34). When the putative protein sequences encoded by the 1574 bp cDNA was searched in NCBI Genbank, a conserved protein domain of 87 amino acids rich in proline and aromatic residues was discovered (Figure 34). This conserved amino acid sequence motif, known as the bromodomain, has been found in many proteins, most of which are involved in transcription activation (Haynes et al., 1992). The bromodomain containing proteins are present in one, or two copies in a tandem array separated by unrelated sequences, in diverse organisms ranging from yeast to human, suggesting that the functional organization of this domain has been conserved during evolution. The amino acid sequence identity of this domain ranges from 20 to 80% among bromodomain containing proteins of different species. Among the bromodomains presented in Figure 35, there are five invariant amino acid residues, including one phenylalanine (F), two tyrosines (Y) and two

Figure 34. The predicted amino acid sequence of the 1574 bp cDNA clone is shown. Consensus sequences for cAMP amd cGMP-dependent protein kinase phosphorylation sites are located at amino acid position 278. Consensus sequences for four protein kinase C and four casein kinase II phosphorylation sites are found at amino acid positions 184, 241, 244, 259 and 278, 294, 296 312, respectively. One potential N-glycosylation, three potential N-myristoylation and two potential amidation sites are located at positions 289; 61, 154, 240 and 273, 333, respectively. The bromodomain of 85 amino acids from amino acid 22 to amino acid 106 is underlined. The position of the first amino acid (Met) in the putative protein sequence is designated 1 in this figure.

N-glycosylation site found at 289 : rsptt N ssvht

==========

cAMP- and cGMP-dependent protein kinase phosphorylation site found at 278 : agrri S lsdmp

==========

Number of potential protein kinase C phosphorylation sites found: 4

[1] 184 : keipf S vkktk

[2] 241 : npsag T aktdk

[3] 244 : agtak T dkqek

[4] 259 : fdmta S pkill

==========

Number of potential casein kinase II phosphorylation sites found: 4

[1] 278 : agrri S lsdmp

[2] 294 : nssvh T gsdve

[3] 296 : svhtg S dveqd

[4] 312 : psshf S asees

==========

Number of potential N-myristoylation sites found: 3

[1] 61 : kkkmy G cteaf

[2] 154 : lrdkd G qvdar

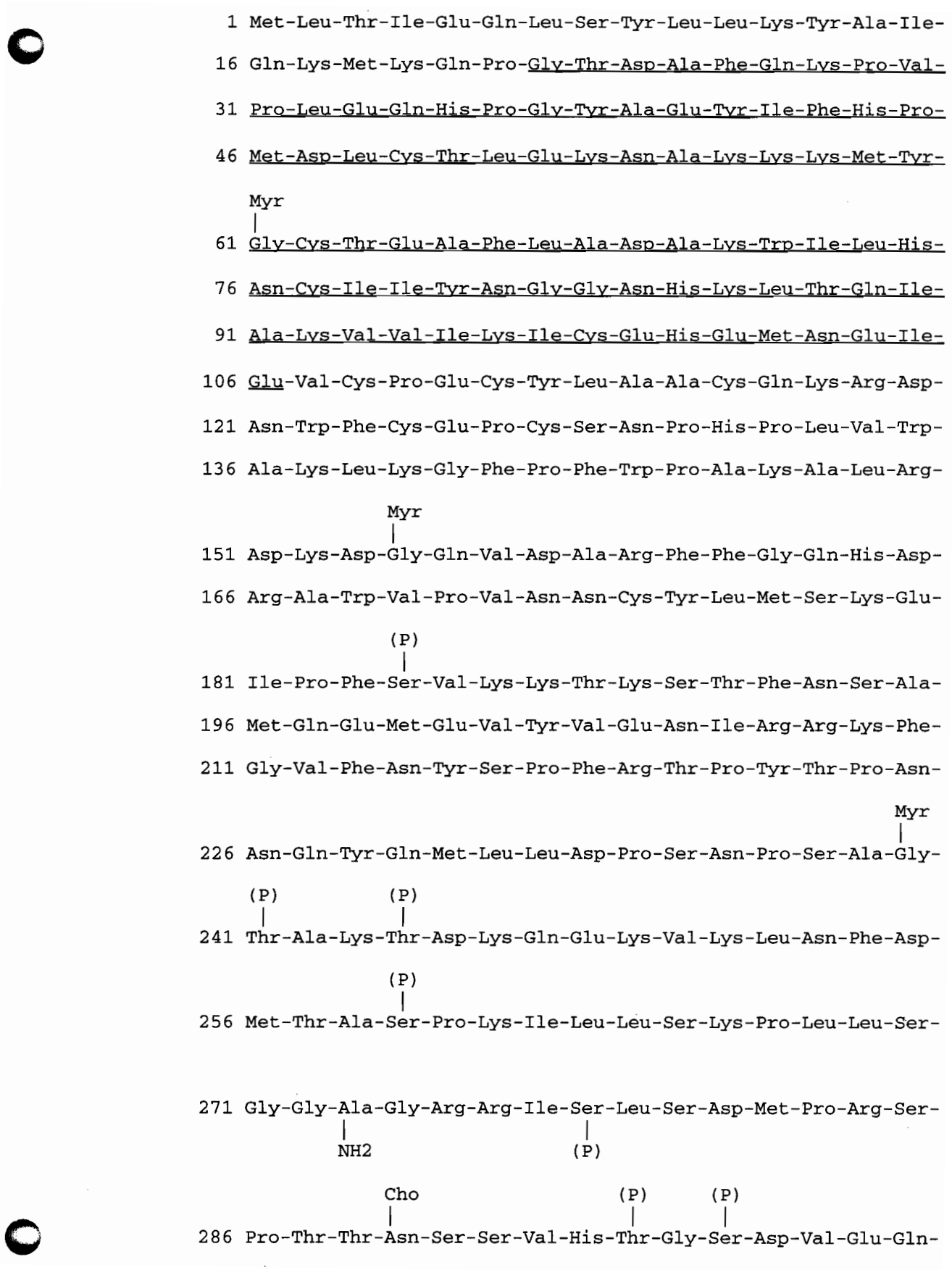
[3] 240 : snpsa G taktd

==========

Number of potential amidation sites found: 2

[1] 273 : llsgg A grris

[2] 333 : pqlge L grrpl


1 Met-Leu-Thr-Ile-Glu-Gln-Leu-Ser-Tyr-Leu-Leu-Lys-Tyr-Ala-Ile-
16 Gln-Lys-Met-Lys-Gln-Pro-Gly-Thr-Asp-Ala-Phe-Gln-Lys-Pro-Val-
31 Pro-Leu-Glu-Gln-His-Pro-Gly-Tyr-Ala-Glu-Tyr-Ile-Phe-His-Pro-
46 Met-Asp-Leu-Cys-Thr-Leu-Glu-Lys-Asn-Ala-Lys-Lys-Lys-Met-Tyr-
Myr
61 Gly-Cys-Thr-Glu-Ala-Phe-Leu-Ala-Asp-Ala-Lys-Trp-Ile-Leu-His-
76 Asn-Cys-Ile-Ile-Tyr-Asn-Gly-Gly-Asn-His-Lys-Leu-Thr-Gln-Ile-
91 Ala-Lys-Val-Val-Ile-Lys-Ile-Cys-Glu-His-Glu-Met-Asn-Glu-Ile-
106 Glu-Val-Cys-Pro-Glu-Cys-Tyr-Leu-Ala-Ala-Cys-Gln-Lys-Arg-Asp-
121 Asn-Trp-Phe-Cys-Glu-Pro-Cys-Ser-Asn-Pro-His-Pro-Leu-Val-Trp-
136 Ala-Lys-Leu-Lys-Gly-Phe-Pro-Phe-Trp-Pro-Ala-Lys-Ala-Leu-Arg-
Myr
151 Asp-Lys-Asp-Gly-Gln-Val-Asp-Ala-Arg-Phe-Phe-Gly-Gln-His-Asp-
166 Arg-Ala-Trp-Val-Pro-Val-Asn-Asn-Cys-Tyr-Leu-Met-Ser-Lys-Glu-
(P)
181 Ile-Pro-Phe-Ser-Val-Lys-Lys-Thr-Lys-Ser-Thr-Phe-Asn-Ser-Ala-
196 Met-Gln-Glu-Met-Glu-Val-Tyr-Val-Glu-Asn-Ile-Arg-Arg-Lys-Phe-
211 Gly-Val-Phe-Asn-Tyr-Ser-Pro-Phe-Arg-Thr-Pro-Tyr-Thr-Pro-Asn-
Myr
226 Asn-Gln-Tyr-Gln-Met-Leu-Leu-Asp-Pro-Ser-Asn-Pro-Ser-Ala-Gly-
(P) (P)
241 Thr-Ala-Lys-Thr-Asp-Lys-Gln-Glu-Lys-Val-Lys-Leu-Asn-Phe-Asp-
(P)
256 Met-Thr-Ala-Ser-Pro-Lys-Ile-Leu-Leu-Ser-Lys-Pro-Leu-Leu-Ser-
271 Gly-Gly-Ala-Gly-Arg-Arg-Ile-Ser-Leu-Ser-Asp-Met-Pro-Arg-Ser-
NH2 (P)
Cho (P) (P)
286 Pro-Thr-Thr-Asn-Ser-Ser-Val-His-Thr-Gly-Ser-Asp-Val-Glu-Gln-

```
                                                          (P)
                                                           |
301 Asp-Pro-Glu-Lys-Lys-Ala-Pro-Ser-Ser-His-Phe-Ser-Ala-Ser-Glu-

316 Glu-Ser-Met-Asp-Phe-Leu-Asp-Lys-Ser-Thr-Gly-Gln-Pro-Gln-Leu-

             NH2
              |
331 Gly-Glu-Leu-Gly-Arg-Arg-Pro-Leu-Gly-Ala-Phe-Trp-Arg-Pro-Trp-

346 Met-Pro-Pro-Ser
```

Abbreviations:

Cho : stands for a carbohydrate group.
Myr : stands for a myristyl group.
NH2 : stands for an amide group.
(P) : stands for a phosphate group.

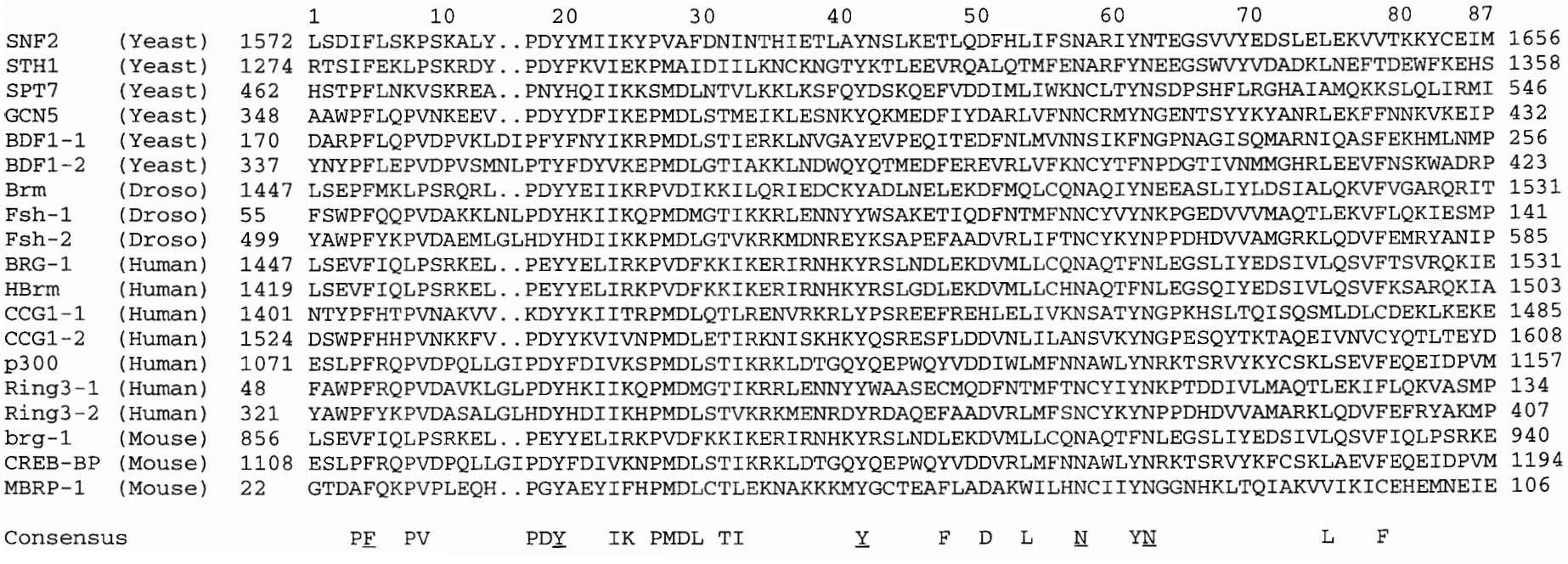
asparagines (N). The proline residue at position 26 is almost invariant except that in SPT7 protein it is substituted by a serine (S) residue. There are also numerous conservative substitutions, as indicated by the consensus sequence in Figure 35. Two other mouse bromodomain containing proteins have been identified. The *brg1* protein contains additional helicase and ATPase domains, and is believed to be a homologue of the *Drosophila brahma* gene (Randazzo et al., 1994). The CBP and the protein encoded by the 1574 bp cDNA appears to belong to another class of bromodomain related proteins, which contain only the bromodomain(s). Many of the proteins in this category, notably the human *CCG1* protein and the mouse CBP protein, have been implicated in the regulation of transcriptional activation (Sekiguchi et al., 1991; Hisatake et al., 1993). Therefore, the protein encoded by the 1574 bp cDNA is a well defined mouse protein in this class, and for the purpose of this thesis, is termed MBRP-1, standing for mouse bromodomain related protein 1. The trapped mouse gene encoding this protein is thus termed *mbrg-1*, standing for mouse bromodomain related gene 1.

3.9.5 The expression pattern of the endogenous gene and the fusion transcript identified by the cDNA clones

To determine the size and the expression pattern of the normal uninterrupted endogenous transcripts, Northern blot analysis of total RNA from CCE ES cells and from the GT9.6 ES cell clone was performed using the cDNA clones as probes. The 735 bp cDNA clone detected two major transcripts of approximately 7 kb and 3 kb in both the CCE ES cells and the GT9.6 ES cell clone (Figure 36). In addtion, both the ES cells and the GT9.6 clone contained a 4.6 kb message located immediately beneath the 28S RNA (4718 bp) and a faint 1.7 kb minor band. Thus, these transcripts are likely to be messages encoded by

Figure 35 Alignment of the bromodomains. The names of the bromodomain containing proteins are given, and the names of the species from which these proteins are derived are included in brackets. The positions of the first and the last amino acid sequence of the bromodomain in a given protein are indicated by the numbers preceding and following the bromodomain, respectively. For proteins with two bromodomains, the numerical order of the bromodomain is illustrated by the numbers (1 or 2) following the name of the protein. The consensus sequence suggests that the amino acids in those positions are more than 60% conserved. The invariant amino acids are underlined.

Bromodomain Alignment

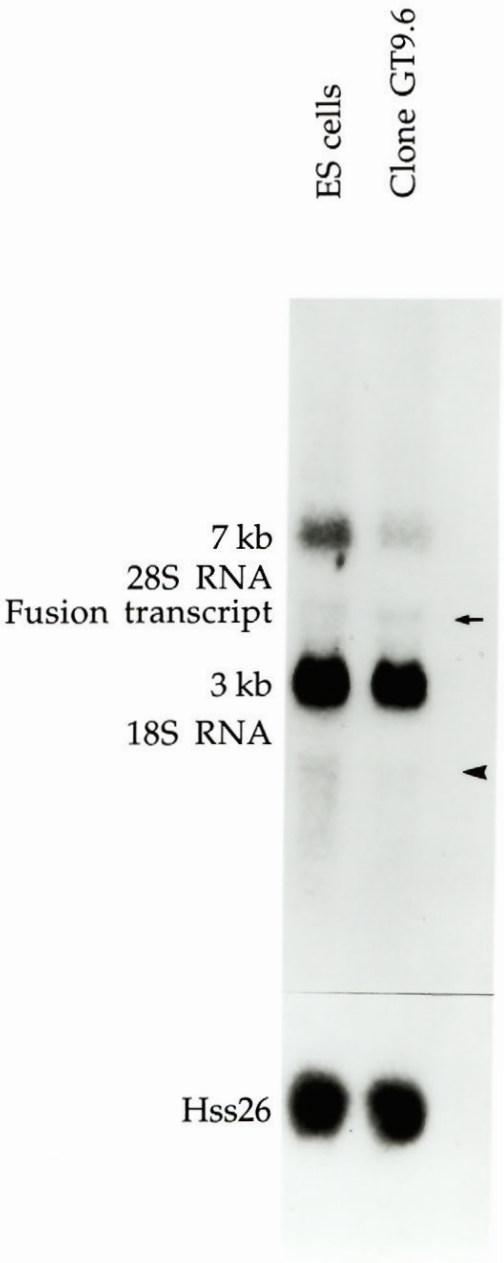


the endogenous gene. Interestingly, the 3 kb transcript was expressed at a higher level than the 7 kb transcript in ES cells. In the GT9.6 ES cell clone, however, the intensity of the 7 kb band was reduced to approximately half compared to that detected in control ES cells. Furthermore, a unique 4.6 kb band was identified in the GT9.6 clone when the blot was rehybridized with the LacZ gene probe (Figure 37). This suggests that the 4.6 kb message represents the fusion transcript in the GT9.6 ES cell clone and that it co-migrates with an authentic 4.6 kb message encoded by the endogenous gene. These results suggest that one of the alleles encoding the 7 kb message was disrupted by the gene trap insertion and a 4.6 kb fusion transcript was produced as a result of the splice acceptor site in the gene trap construct.

The 1574 bp cDNA clone only contains cDNA sequences downstream of the *En*-2 splice junction at the site of the gene trap insertion, these sequences were not present in the LacZ fusion transcript. Therefore, when the same Northern blot is probed with the 1574 bp cDNA, the 4.6 kb LacZ fusion transcript should not be detected. However, the 4.6 kb endogenous message co-migrating with the fusion transcript, along with the endogenous 7 kb, 3 kb, and 1.7 kb transcripts, were clearly visible (Figure 38). Moreover, as observed for the blot hybridized with the 735 bp cDNA probe, the intensity of the 7 kb message was reduced to half of that detected in the control ES cell samples. Thus, I infer from this that common sequences are shared by the 7 kb transcript and the 1574 bp cDNA clone and that disruption of one copy of the 7 kb message results in this quantitative difference.

In addition to the four endogenous messages, the 1574 bp cDNA probe also hybridized to a novel transcript of approximately 800 bp (Figure 38). The level of expression of the 1.7 kb and the 800 bp transcripts appeared to be constant in normal ES cells versus that in the GT9.6 ES cell clones. Both the

Figure 36. A northern blot analysis of total RNA from CCE ES cells and from the GT9.6 ES cell clone probed with the 735 bp cDNA clone is shown. The 735 bp cDNA clone detected two major transcripts of approximately 7 kb and 3 kb in both the CCE ES cells and the GT9.6 ES cell clone. The 3 kb transcript was expressed at a higher level than the 7 kb transcript in ES cells. In the GT9.6 ES cell clone, however, the intensity of the 7 kb message was reduced to half compared to that in control ES cells. In addtion, both the ES cells and the GT9.6 clone expressed a 4.6 kb message (arrow) located immediately beneath the 28S RNA (4718 bp) and a 1.7 kb minor band (arrowhead). The HSS26 ribosomal protein RNA hybridizing to a 513 bp Hind III - EcoR I fragment derived from the human HSS26 cDNA provided an internal control for the amount of RNA loaded (bottom strip).


ES cells
Clone GT9.6
7 kb
28S RNA
Fusion transcript
3 kb
18S RNA
Hss26

1.7 kb and the 800 bp transcripts hybridized with the 1574 bp cDNA probe, although the 1.7 kb transcript hybridized to this probe at a much higher intensity (Figure 38). In contrast, the 1.7 kb transcript was barely and the 800 bp message was not at all detected by the 735 bp probe (Figure 36). Given also the fact that the sequence of the 735 bp cDNA begins to differ from that of the 1574 bp cDNA clone at nucleotide position 730 (Figure 32), it appears plausible that the two unique transcripts hybridized by the 1574 bp probe were generated by alternative splicing.

3.9.6 The expression pattern of the endogenous gene in different tissues and during P19 cell differentiation

The expression profile of the endogenous gene in different tissues was determined by Northern blot analysis of RNAs isolated from various adult mouse tissues. In brain, heart, kidney and gut, the 735 bp cDNA probe identified the 7 kb and 3 kb species as major transcripts. As seen in the ES cells and the GT9.6 ES cell clone, the 4.6 kb transcript was significantly less abundant. However, the 7 kb message is expressed at a significantly higher level than the 3 kb message in the mouse brain, while the 3 kb species is more abundant than the 7 kb transcript in ES cells (Figure 39). Since only the 7 kb transcript was interrupted by the splicing event of the gene trap construct, to generate a 4.6 kb fusion transcript in the GT 9.6 ES cell clone (Figure 36, 37, 38), the β-galactosidase expression seen in the nervous system of the GT9.6 ES cell derived chimeras suggests that the 7 kb transcript is expressed at higher levels in neurons. The nuclear localization of the fusion protein identified in the mouse nervous system by the immunocytochemical analysis indicates that the 7 kb message encodes a protein containing a nuclear localization signal. In non-neuronal tissues, the expression level of the 7 kb, as well as the 3 kb

Figure 37. Northern blot of RNA from unmanipulated ES cells and the GT9.6 clone probed with the LacZ gene. The ES cell RNA blot probed with the 735 bp cDNA probe was stripped and rehybridized with the LacZ gene probe. The GT9.6 ES cell clone, but not the control ES cells, expressed the 4.6 kb fusion transcript, indicating that in the GT9.6 ES cell clone, one of the allele encoding the 7 kb message was disrupted by the insertion of the gene trap construct, and a unique 4.6 kb fusion transcript was expressed.

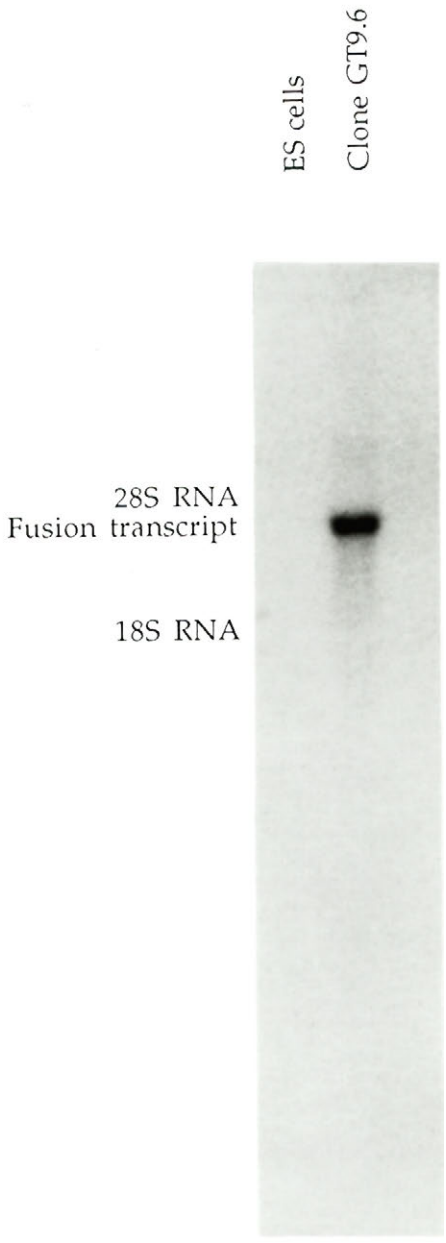
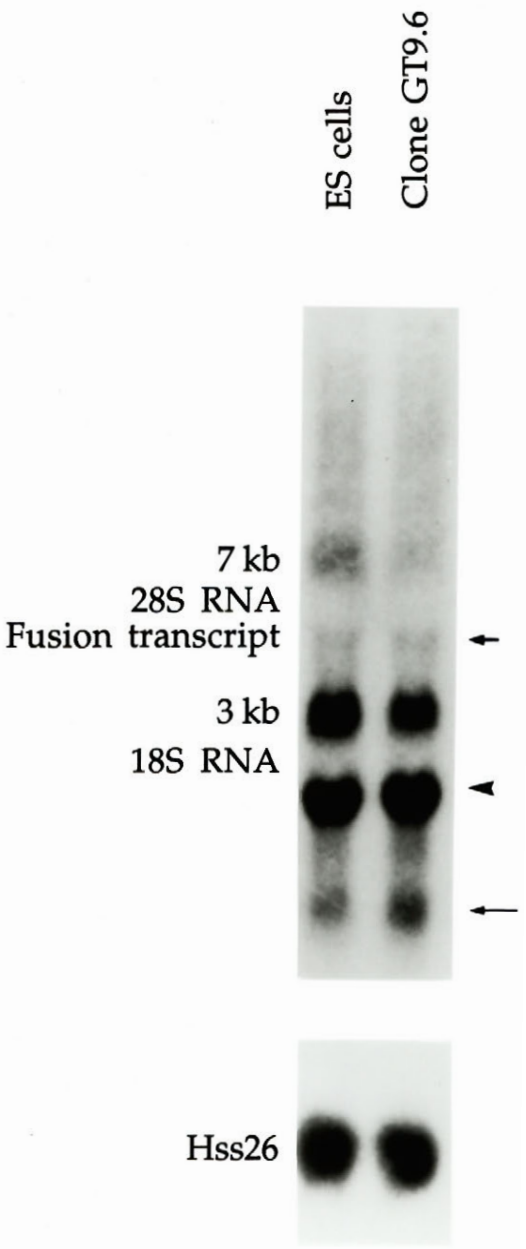

ES cells
Clone GT9.6
28S RNA
Fusion transcript
18S RNA

Figure 38. Northern blot of RNA from unmanipulated ES cells and clone GT9.6 probed with the 1574 bp cDNA. The ES cell Northern blot shown in Fig. 33 was re-stripped and probed with the 1574 cDNA probe. The 1574 bp cDNA clone only contains cDNA sequences downstream of the *En*-2 splice junction at the site of gene trap insertion. These sequences were not present in the LacZ fusion transcript. Nonetheless, as with the blot probed with the 735 bp cDNA, the endogenous 7 kb, 4.6 kb (arrow), 3kb and 1.7 kb (arrowhead) transcripts, were detected in the control ES cells and the GT9.6 ES cell clone. Similarily, the intensity of the 7 kb message was reduced to half of that observed in the control ES cell sample. Furthermore, the 1.7 kb transcript hybridized intensely with the 1574 bp cDNA probe. In addition to the four endogenous messages, the 1574 bp cDNA probe also hybridized to a transcript of approximately 800 bp (long arrow). The 1.7 kb and the 800 bp transcripts appeared to be expressed at equal levels in normal ES cells and in the GT9.6 ES cell clones. The HSS26 ribosomal protein RNA hybridizing to a 513 bp Hind III - EcoR I fragment derived from the human HSS26 cDNA provided an internal control for the amount of RNA loaded (bottom strip).


ES cells
Clone GT9.6
7 kb
28S RNA
Fusion transcript
3 kb
18S RNA
Hss26

messages appear to be constant (Figure 39). The extra bands seen higher than the 7 kb message may be due to the low stringency conditions as evidenced by the higher background hybridization on the blot.

Given the high abundance of the 7 kb transcript in the nervous system compared to the other messages, I attempted to further investigate the differential expression pattern of the endogenous transcripts during neuronal differentiation, Northern blot analysis was performed on total RNA isolated from P19 cells and from cells at different stages of P19 cell differentiation induced by retinoic acid treatment with the 1574 bp cDNA probe. As for the ES cell samples, the 3 kb transcript was more abundant than the 7 kb message in undifferentiated P19 cells. However, upon retinoic acid treatment when P19 cells are induced to differentiate into neuronal phenotype, the relative level of the 7 kb transcript in comparison to the 3 kb transcript was increased. This unique pattern of differential expression of the two transcripts during neuronal differentiation was first observed at day 1, and then day 2, day 5 and up to ten days following retinoic acid treatment. The RAC65 cell line, a mutant embryonic carcinoma cell line that does not undergo neuronal differentiation upon retinoic acid treatment (Campione-Piccardo et al., 1985; Pratt et al., 1990), did not show any significant changes in the level of expression of the 7 kb transcript relative to that of the 3 kb transcript following retinoic acid treatment. The absolute and relative ratios of RNA levels of P19 and RAC65 cell lines were quantitated by densitometry (data not shown). Interestingly, a unique 1 kb message present was detected in all the samples obtained from the P19 cells. This transcript may have been obscured by the intense hybridization signal of the 1.7 kb message seen in the ES cell northern blot (Figure 38). Thus, the abundance of the 7 kb and 3 kb transcripts as well

their relative levels appear to change during neuronal differentiation of P19 cells and vary in different tissues of the mature mouse.

3.9.7 Single site insertion of the gene trap construct in the GT9.6 ES cell genome

To exclude the possibility that the β-galactosidase expression pattern seen in the GT9.6 ES cell derived chimeras could be due to the expression of multiple insertions of the gene trap construct in the genome of the GT9.6 ES cell clone, Southern blot analysis was performed on DNA isolated from ES cells and the GT9.6 clone. High molecular weight DNA was digested with Kpn I and the DNA blot was hybridized with the LacZ gene probe. Only one fragment of 6 kb was identified in the GT9.6 clone whereas no signal was detected in the control CCE ES cells (Figure 41). Since the pGT4.5 gene trap construct does not contain an internal Kpn I site, the integration of the gene trap construct in the GT9.6 clone must therefore have been a single insertion event and the expression of the β-galactosidase protein observed in the GT 9.6 chimeras reflects the expression pattern of one trapped gene.

3.9.8 Conservation of trapped gene sequences in different species

A search for conservation of the trapped endogenous gene sequences among different species in the animal kingdom was carried out by Southern blot analysis. Genomic DNA were isolated from human, mouse, rat, zebrafish, snake, quail, *Drosophila*, *C. elegans* and yeast. Genomic DNA of human, mouse and *Drosophila* were restriction digested with both Kpn I and EcoR I, whereas genomic DNA of rat, zebrafish, snake and quail were restriction digested with EcoR I, and that of *C. elegans* and yeast with Kpn I. The enzyme digested genomic DNA from different species were blotted and

appear to change during neuronal differentiation of P19 cells and vary in different tissues of the mature mouse.

3.9.7 Single site insertion of the gene trap construct in the GT9.6 ES cell genome

To exclude the possibility that the β-galactosidase expression pattern seen in the GT9.6 ES cell derived chimeras could be due to the expression of multiple insertions of the gene trap construct in the genome of the GT9.6 ES cell clone, Southern blot analysis was performed on DNA isolated from ES cells and the GT9.6 clone. High molecular weight DNA was digested with Kpn I and the DNA blot was hybridized with the LacZ gene probe. Only one fragment of 6 kb was identified in the GT9.6 clone whereas no signal was detected in the control CCE ES cells (Figure 41). Since the pGT4.5 gene trap construct does not contain an internal Kpn I site, the integration of the gene trap construct in the GT9.6 clone must therefore have been a single insertion event and the expression of the β-galactosidase protein observed in the GT 9.6 chimeras reflects the expression pattern of one trapped gene.

3.9.8 Conservation of trapped gene sequences in different species

A search for conservation of the trapped endogenous gene sequences among different species in the animal kingdom was carried out using a zooblot. Genomic DNA were isolated from human, mouse, rat, zebrafish, snake, quail, *Drosophila, C. elegans* and yeast. Genomic DNA of human, mouse and *Drosophila* were restriction digested with both Kpn I and EcoR I, whereas genomic DNA of rat, zebrafish, snake and quial were restriction digested with EcoR I, and that of *C. elegans* and yeast with Kpn I. The enzyme digested genomic DNA from different species were Southern blotted and

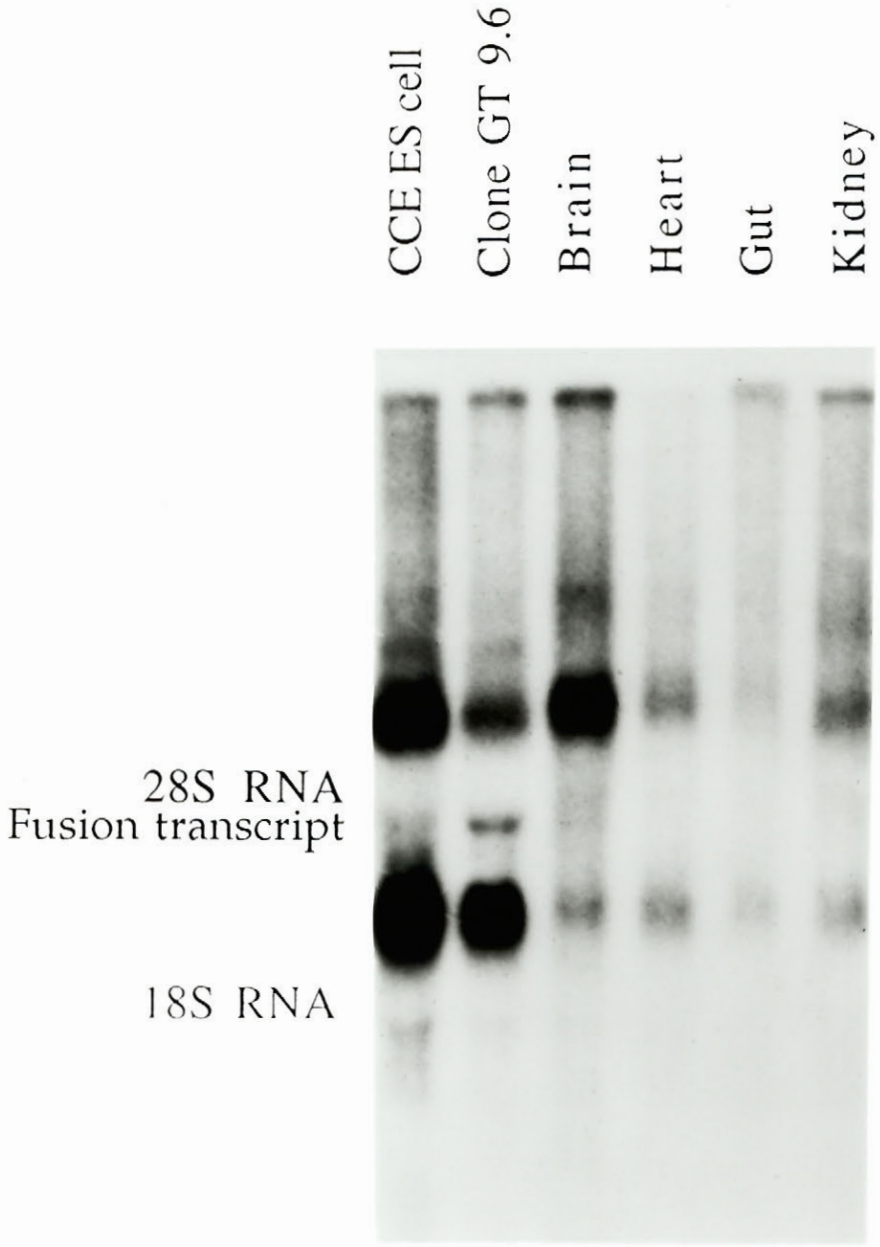

CCE ES cell
Clone GT 9.6
Brain
Heart
Gut
Kidney
28S RNA
Fusion transcript
18S RNA

Figure 40. Northern blot analysis of total RNA isolated from undifferentiated P19 cells and from P19 cells at different stages of differentiation induced by retinoic acid treatment. The 1574 bp cDNA was used as a probe. As observed for the ES cell samples, the 3 kb transcript was more abundant than the 7 kb message in undifferentiated P19 cells (Day 0). One, two, five and ten days following retinoic acid treatment when P19 cells are induced to differentiate into neurons, the relative level of the 7 kb transcript was increased compared with the level of the 3 kb transcript. The RAC65 cell line, a mutant embryonic carcinoma cell line that does not undergo neuronal differentiation upon retinoic acid treatment, did not show any significant changes in the relative level of the 7 kb transcript versus that of the 3 kb transcript following retinoic acid treatment. The absolute and relative ratios of RNA levels of P19 and RAC65 cell lines were quantitated by densitometry (data not shown). In addition, a new 1 kb message (arrow) was detected in all the P19 cell samples. This transcript may have been obscured by the intense hybridization signal of the 1.7 kb message seen in the ES cell northern blot probed with the 1574 bp cDNA probe (Figure 38).

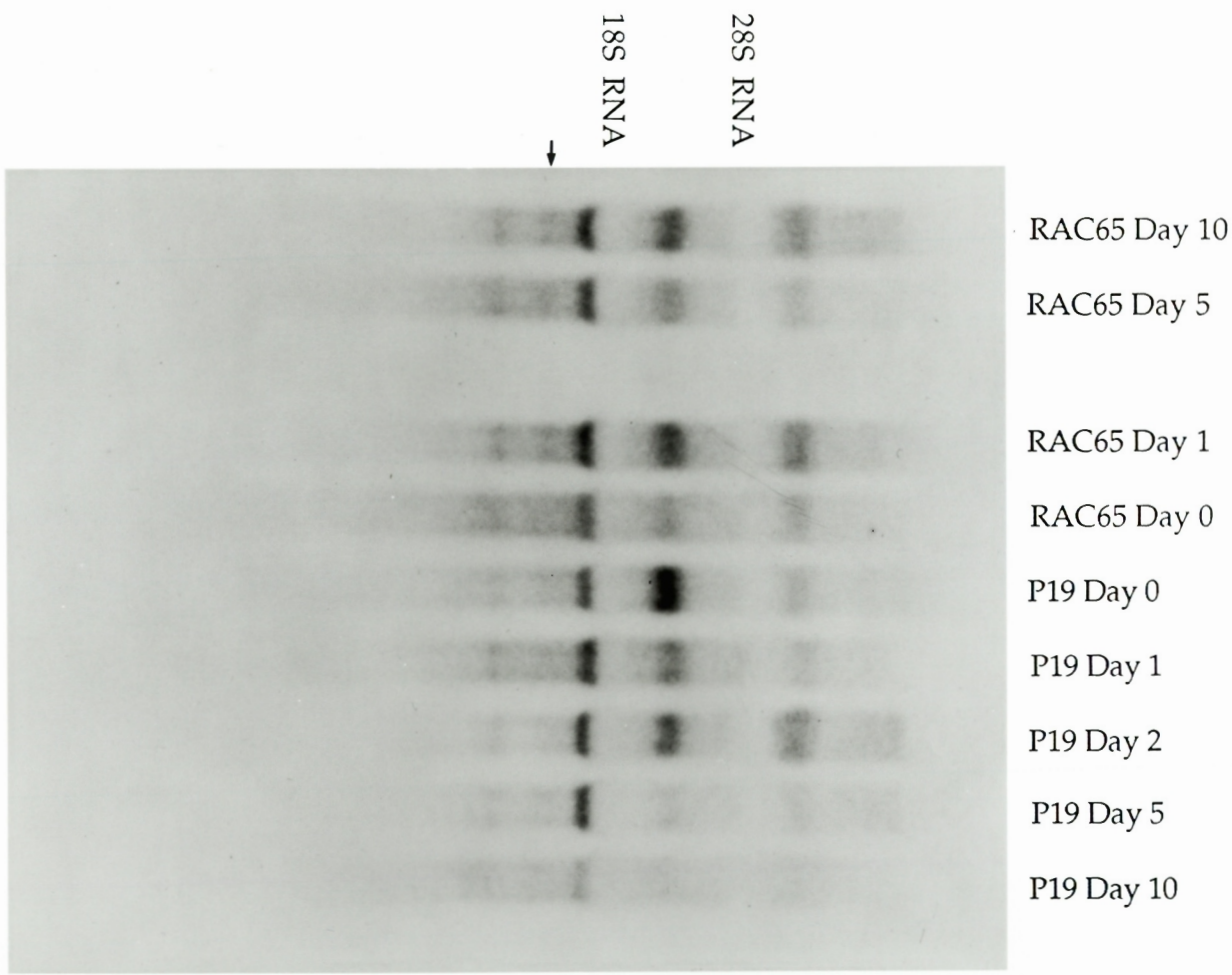
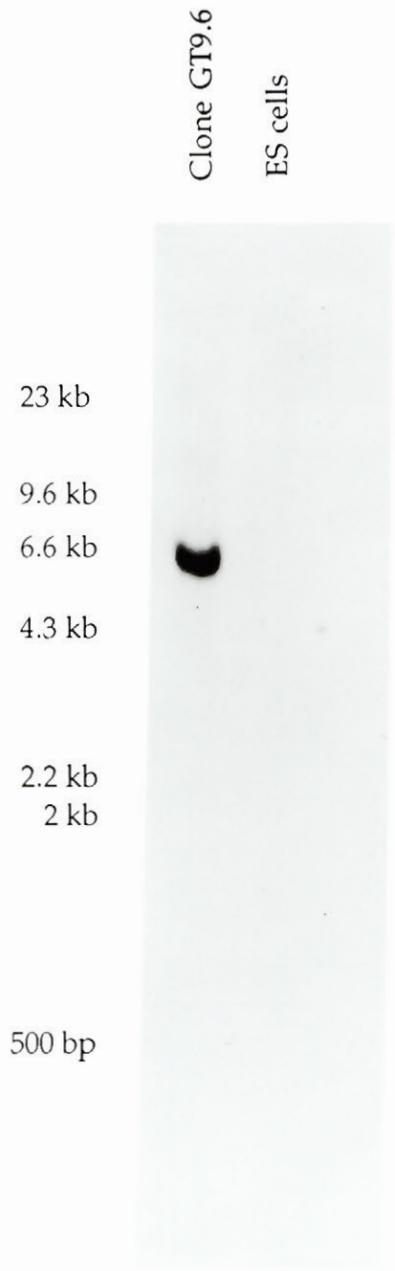

18S RNA
28S RNA
RAC65 Day 10
RAC65 Day 5
RAC65 Day 1
RAC65 Day 0
P19 Day 0
P19 Day 1
P19 Day 2
P19 Day 5
P19 Day 10

Figure 41. Southern blot of DNA isolated from ES cells and the GT9.6 clone probed with the LacZ gene. High molecular weight DNA was digested with Kpn I. A single band of approximately 6 kb was identified in the GT9.6 clone whereas no signal was detected in the control CCE ES cells.


Clone GT9.6
ES cells
23 kb
9.6 kb
6.6 kb
4.3 kb
2.2 kb
2 kb
500 bp

probed with the 1574 bp cDNA probe. Two or more hybridized bands were detected from human, rat and mouse DNA, whereas no detectable signal was observed when DNA from zebrafish, snake, fruitfly, bird, nematode and yeast was hybridized under the stringent condition (0.5XSSC) provided in this experiment (Figure 42). Thus, the endogenous gene identified by gene trap in mouse ES cells appears to contain homologous sequence motifs that exist in other mammalian species, including human and rat. Similar sequence homologies do not seem to exist in organisms such as zebrafish, snake, fruitfly, bird, nematode and yeast.

3.9.9 Chromosomal localization of *mbrg-1*

To determine the chromosomal location of the *mbrg-1* gene, flourescent in situ hybridization (FISH) was performed in collaboration with Drs. X. Zhang and Dr. P. Eydoux at Montreal Children's Hospital using the 1574 bp cDNA clone as a probe. The mouse *mbrg-1* gene was found to be localized on mouse chromosome 2q and its human counterpart on human chromosome 20q.

Figure 42. Southern blot of genomic DNA isolated from human, mouse, rat, zebrafish, python, quail, *Drosophila, C.elegants* and yeast probed with the 1574 bp cDNA. The human, mouse and *Drosophila* genomic DNA were digested with either Kpn 1 or EcoR 1. The zebrafish, python, quail genomic DNA were digested with EcoR I only, and the *C.elegants* and yeast genomic DNA were digested with Kpn I only. While one or more bands were detected from the human (arrow in Kpn I digested DNA blot), mouse and rat genome, no detectable signal was observed in the genomic DNA from the rest of the non-mammalian species. The blot was washed in 0.5 X SSC, 0.1% SDS at 65°C for 15 minutes.

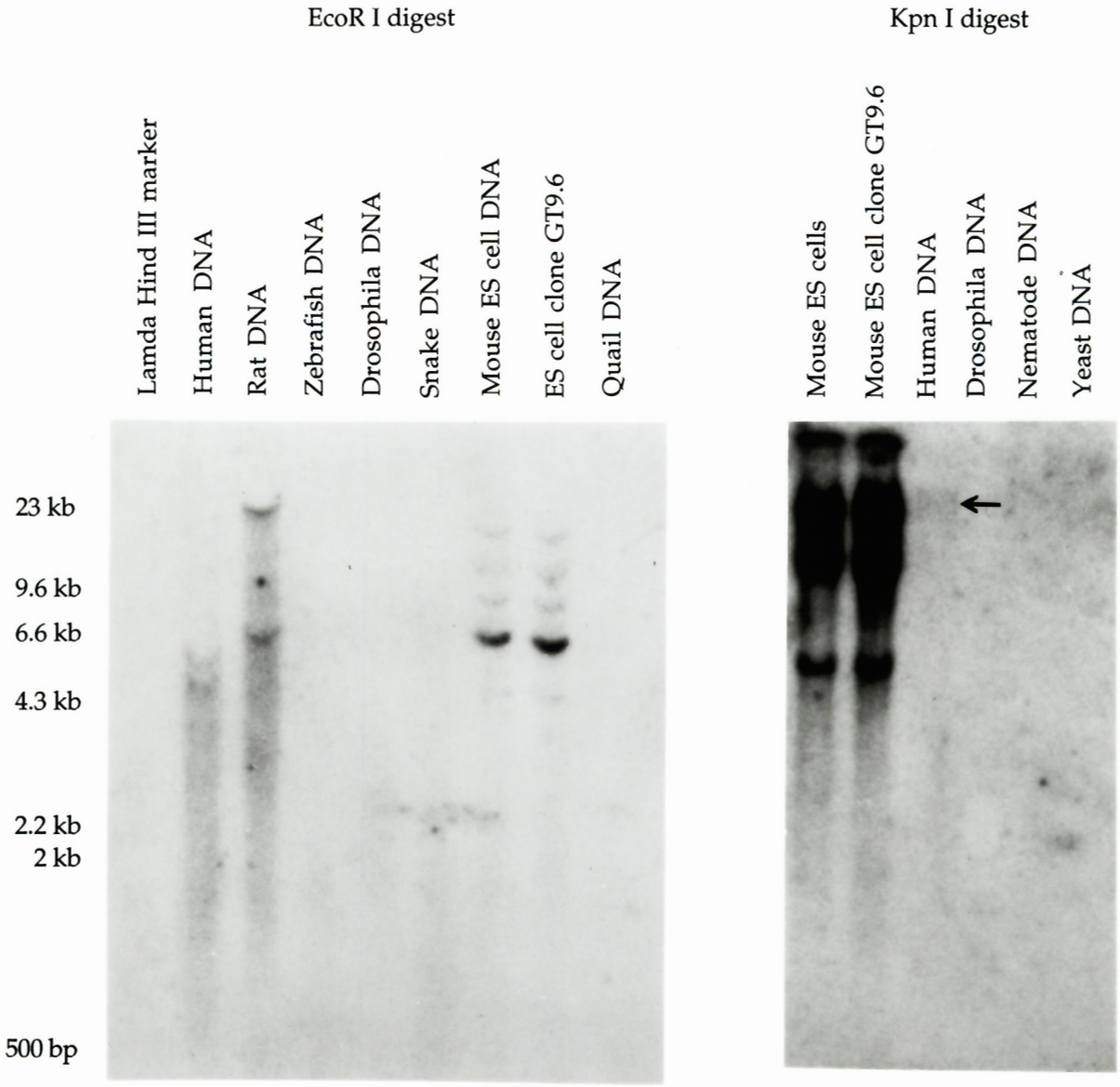
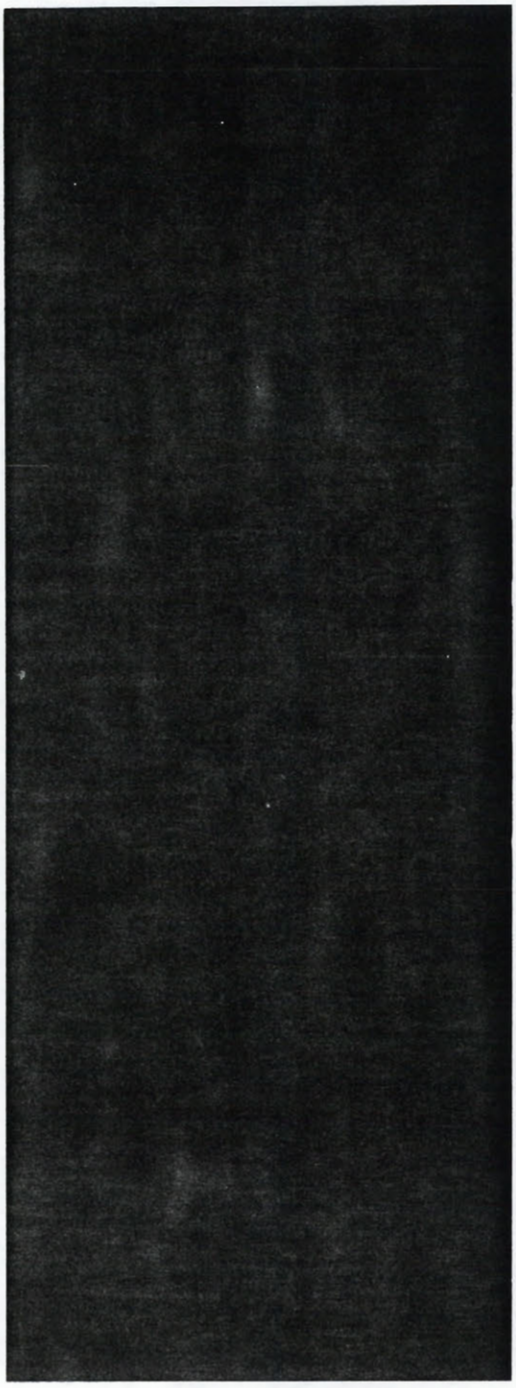
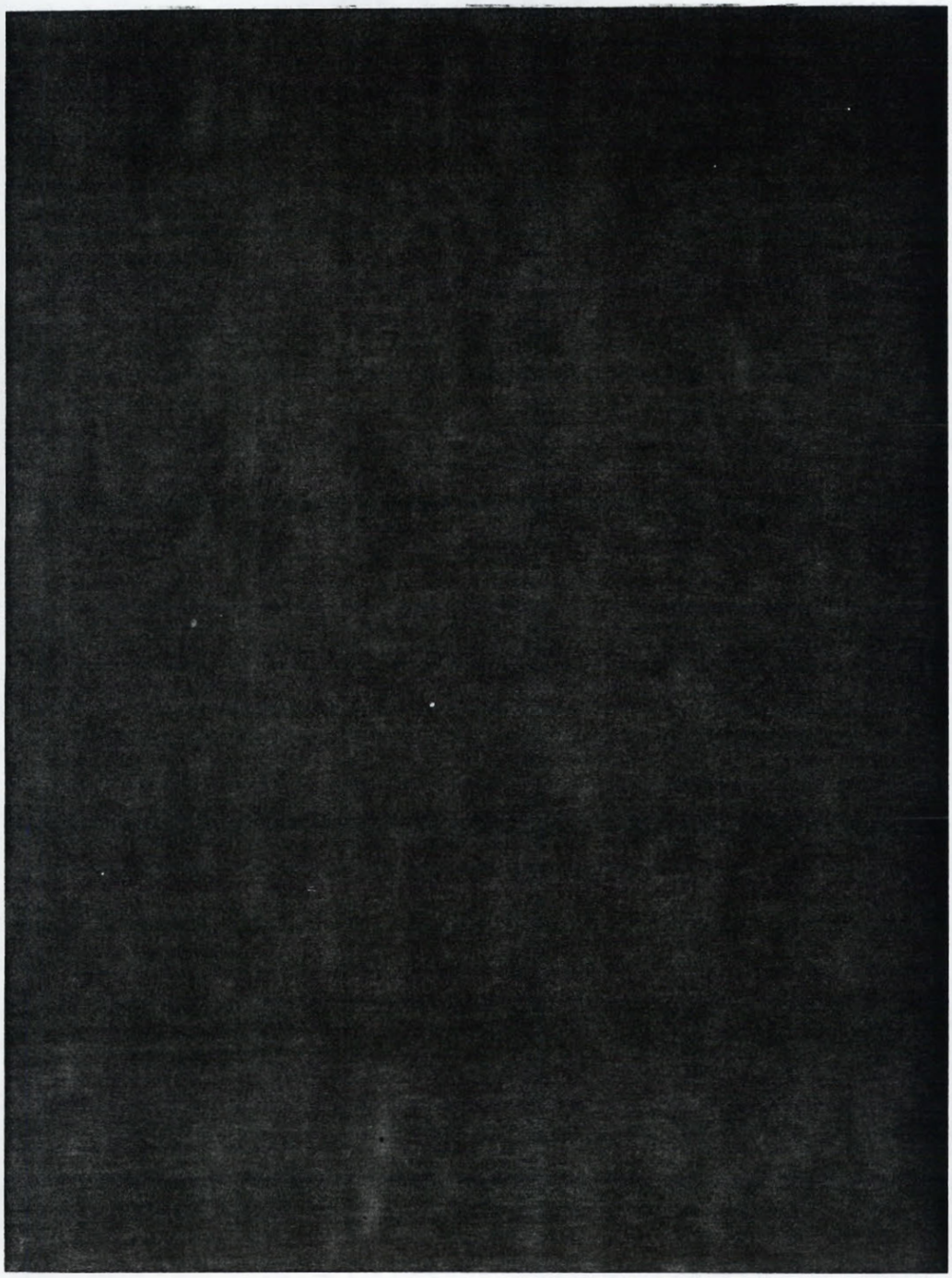

EcoR I digest
Kpn I digest
Lamda Hind III marker
Human DNA
Rat DNA
Zebrafish DNA
Drosophila DNA
Snake DNA
Mouse ES cell DNA
ES cell clone GT9.6
Quail DNA
Mouse ES cells
Mouse ES cell clone GT9.6
Human DNA
Drosophila DNA
Nematode DNA
Yeast DNA
23 kb
9.6 kb
6.6 kb
4.3 kb
2.2 kb
2 kb
500 bp

Figure 43 Chromosomal localization of *mborg-1*. (A) Mouse metaphase chromosome spread hybridized with the 1574 bp cDNA probe using FISH. (B) Human metaphase chromosome spread hybridized with the 1574 bp cDNA probe using FISH. Strong signals on mouse chromosome 2q (A) and human chromosome 20q (B) were detected.

A



B



Discussion

4.1 Gene trapping as a means of mammalian gene identification

Gene trapping strategy was used in this investigation to screen for genes that, on the basis of their expression pattern, play important roles during mammalian nervous system development. The availability of the ES cell-chimera-transgenic system provides a unique advantage that allows large scale screening for mammalian gene trap events *in vitro* and subsequent determination of the temporal and spatial pattern of expression of the trapped genes *in vivo*. Thus, in contrast to the genetic screening performed in *Drosophila*, where selections are based on mutational phenotypes, gene trapping offers a screening procedure that identifies genes according to their expression pattern. This method allows genes with interesting tissue specific developmental expression patterns to be selected at will for further molecular characterization. In this investigation, this gene trapping strategy was applied with the purpose of identifying new genes that may play significant roles in the development of the nervous system.

Several limitations exist in using this approach for mammalian gene identification. Gene trapping is a random approach, and the selection of trapped ES cell clones requires chimera analysis. Massive screening for interesting expression patterns in chimeras can be tedious and labor intensive. Selection of gene trapped ES cell clones in this study is limited to those that express the gene trap construct (blue clones). As a consequence, only genes that are expressed during early embryogenesis at the blastocyst stage were investigated. Genes expressed at later embryonic stages would not be included in this study. Screening for these LacZ negative ES cell clones (white clones) for their activated expression pattern in the developing

transcript to form the fusion transcript, whereas other alternatively spliced RNA isoforms remain intact (Figure 36, 38). Thus, there are no apparent changes in the overall pattern of the alternative splicing due to the gene trap event. The 7 kb transcript is expressed in a neuron specific fashion (Figure 39), and its level is significantly upregulated following neuronal differentiation (Figure 40) whereas other RNA isoforms are expressed ubiquitously. Thus, there appears to be a unique splicing mechanism shared by all neurons in producing the 7 kb transcript, suggesting an important role played by the 7 kb transcript in neuronal function.

Regional or cell type specific variation (Laurie and Seeburg, 1994; Smith and O'Dowd, 1994; O'Connor et al., 1994) and developmental heterogeneity (Laurie and Seeburg, 1994) of alternative spliced RNA isoforms have been observed in the nervous system. In this investigation, the 7 kb transcript (Figures 37, 38), and the protein it encodes (Figure 9-15), are highly abundant in neurons. This regulated pattern of neuron-specific expression suggests that it may play an important role in formation, maintenance or gene regulation in neuronal nuclei.

4.5 Bromodomain in transcriptional activation

The 1574 bp cDNA clone corresponding to the 1.7 kb endogenous RNA has been found to encode a putative protein containing a bromodomain. This is a partial cDNA clone since it does not contain a polyadenylation signal and a polyA tail, nor does it have an unambiguous 5' end of the coding sequence. The cDNA clone (1574 bp) is about 130 bp short of the entire sequence of the endogenous transcript (1.7 kb). The close resemblance of the sequence surrounding the third methionine residue to the Kozak consensus indicates that this methionine may be used as the translation start codon of the

putative protein. The putative protein should then be composed of 304 amino acids, instead of 349 amino acids were the first methionine used as the translation start site. Evidently, the definitive proof for the translation initiation site requires the cloning of the 5' sequence upstream of the first methionine, with the identification of an in-frame upstream stop codon preceding the starting methionine.

The bromodomain has been implicated in transcriptional activation. Proteins containing the bromodomain seem to fall into two categories. The first class of these proteins contains ATPase and helicase motifs, in addition to the downstream bromodomain(s). The second class of the bromodomain containing genes consists of only the bromodomain. The prototype of this class of proteins is the human *CCG1* protein, which was found to be the 250k subunit of the transcription factor IID (TFIID) complex (Rupport et al., 1993). This subunit is a potential binding site for activators, and is essential in inducing activator-dependent transcription (Hisatake et al., 1993). In addition to the bromodomain, the *CCG1* protein also contains four potential phosphorylation sites for casein kinase II, two for cAMP-dependent protein kinase and four for Ca^{++}-calmodulin-dependent protein kinase. The putative protein encoded by the 1574 bp cDNA appears to bear some resemblance to the human *CCG1* protein. It also contains only a bromodomain without the ATPase and helicase motifs. In addition, the putative protein sequence contains consensus sequences for one cAMP and cGMP-dependent protein kinase, four protein kinase C and four casein kinase II phosphorylation sites. Despite some structural similarity, however, the *CCG1* protein (1872 aa) (Sekiguchi et al., 1991) is about six times the size of the protein encoded by the 1574 cDNA (best estimate is 349 aa). Moreover, *CCG1* is a nuclear protein which contains a sequence similar to the putative DNA-binding domain of

sequences in mammalian species such as rat and human (Figure 42). Thus, the sequence contained in the 1.7 kb transcript is only conserved among mammals. One simple explanation is that the bromodomain conservation occurs at the protein level with amino acid sequence homologies ranging from 20% to 80%. The level of difference in nucleotide sequences encoding these divergent amino acids is so large that there are not enough homologous DNA sequences to be detected by nucleic acid hybridization. It is thus possible that the degenerate DNA sequences encoding the bromodomains are more closely related among mammals than other species. One interesting question needed to be addressed is whether cDNA sequences derived from the other RNA isofoms have homologous sequences present in species other than mammals.

Transcription factors encoded by the same locus but with different pattern of evolutionary conservation have been reported (Albagli et al., 1994). The chicken c-ets-1 locus encodes two transcription factors that differ in their N-termini. One of the transcription factors, p54c-ets-1, is widely conserved in vertebrates ranging from amphibians to mammals, whereas the other, p68c-ets-1 is only found in birds and reptiles. Thus, the I54 and αβ exons that respectively encoded the two transcription factors apparently were subjected to different degrees of evolutionary conservation. Similarly, it is possible that different exons encoding the various alternatively spliced transcripts of the trapped gene locus adopt different patterns of cross species conservation. Sequences in other RNA isoforms may be conserved in species that are distinct from the 1.7 kb RNA. The differential use, either gain or loss, of exons encoding different protein isoforms by the same gene locus during evolution may greatly facilitate our understanding of the evolutionary importance of various transcriptionally functional domains.

progeny of a cortical progenitor different from the precursors of the labeled cells within the same radial column (Figure 17). Tangential migration could account for the opposite colored cells seen in the radial columns in chimeras with predominant ES cell contribution. Alternatively, a single progenitor cell in the VZ may divide asymmetrically, with one daughter cell migrating radially, and the other dispersing horizontally in the VZ for a limited distance before moving out of the VZ. (Fishell et al., 1993). These daughter cells may still follow a radial migrating path, but they differ from their siblings in the onset of migration because of asymmetrical division, and in the initial location of migration because of horizontal movements within the VZ. These cells could presumably account for the opposite colored cells seen within the radial columns, and even those retroviral labeled cells thought to be dispersing tangentially (Walsh and Cepko, 1993). The individually separated radial columns seen in the GT9.6 chimeras with low ES cell contributions evidently contradict this theory. Due to asymmetrical division and horizontal cell migration within the VZ, radial columns would have always been accompanied by labeled cells present in the neighbouring regions of unlabeled cells, as were observed in the GT9.6 chimeras with dominant ES cell contributions. Thus, these results support the notion that there are two patterns of cell migration during cortical genesis. While tangential dispersion has been elegantly identified by the retroviral labeling experiments (Walsh and Cepko, 1992, 1993), implicated in the studies of X-inactivation of transgenes (Tan and Breen, 1993; Tan et al., 1995), and further supported by the chimera analysis conducted in this investigation, there is also a indisputable pattern of radial migration during cortical genesis, which was clearly shown by the individually separated radial columns in the developing cortex of the GT9.6 chimeras. Since the width of the radial stripes vary

randomly within a chimera, it is unlikely that these blue stripes in the cortex represent an endogenous pattern of expression of the trapped gene in the GT9.6 ES cell clone.

Cell distribution during the development of the spinal cord, like that of the cerebral cortex, is an issue of controversy. The GT 9.6 chimeras in this investigation have demonstrated that the cell mixing during early spinal cord development is not at all random. Clonally related cells are distributed in a spatially restricted fashion. There is a great deal of variation in cell distribution in left versus right side of the spinal cord. This pattern of asymmetrical clonal cell distribution could be clearly seen on the cross section of the spinal cord (Figure 11), but more evidently, from the longitudinal view of the whole mount chimeric spinal cord (Figure 18 and 20). Clear clonal compartments and boundaries are indicated by labeled cell patches. The coherent patches of clonally related cells along the longitudinal axis are most striking in E12.5 chimeric spinal cord (Figure 20). Each fine segment may represent clonal progeny of a single progenitor with a restricted rostral-caudal distribution. Limited longitudinal expansion of the clonally related spinal neurons has been shown in the developing chick spinal cord by retroviral labeling of spinal neurons precursors (Leber et al., 1990; Leber and Sanes, 1995). Thus, in contrast to the aggregation chimera experiments in which the two spinal cord cell populations were thought to mix extensively, which suggest a lack of clonal barriers during the development of the spinal cord (Musci and Mullen, 1992), this investigation has revealed a spatially restricted pattern of organization during the development of the spinal cord. In early neurogenesis, cell mixing and distribution are not random, but rather follow a pattern of clonal segregation within coherent clonal patches in all three dimensions. Cells that contribute to the left half of the spinal cord do not mix

with the cells that give rise to the right half of the spinal cord and vice versa. There is also apparent dorsal-ventral patches that suggest a clonal segregation in that direction. Most strikingly, there is a segmented pattern of generation of the spinal cord along the longitudinal axis. The dorsal root ganglia (DRG) corresponding to each labeled segment was labeled approximately to the same extent in the E12.5 chimeras, suggesting that the neural crest cells that give rise to the DRG neurons are derived from the neural tube in the corresponding segment in the spinal cord, and are representative of the genotype proportion in the remainder of the neural tube. The longitudinal cohort cell patches in the spinal cord were found to be bigger in the mature chimera than in E12.5 chimeras . This could be due to the expansion of labeled segments during the rostral-caudal extension of the spinal cord.

The difference in early spinal cord cell mixing observed between the chimera studies of Musci and Mullen (1992) and that conducted in this investigation may be due to different mouse strains used. One could also argue that the GT9.6 ES cell clone used to construct chimeras in this study may not be entirely normal. Nonetheless, ES cells have been shown to mix extensively with the inner cell mass (ICM) of the injected blastocysts (Robertson, 1986). The developmental potential of the GT9.6 clone could be complemented by the normal ICM of the recipient blastocysts. Normal early cell mixing has also been reported for various mouse mutant aggregation chimeras, despite the mutant abnormality of one chimeric component (Musci and Mullen, 1990; Goldowitz, 1989). Short of any identifiable mutant phenotype of the GT9.6 ES cell clone, the pattern of cell distribution in the GT9.6 chimeras may not be grossly altered.

The neuron-specific expression of the fusion transcript was based on the morphological identification in histochemical and immunocytochemical

analysis. Therefore, it can not be ruled out definitively that some other cell types may express the fusion protein. Formal proof that the labeled cell types are all neurons would require a double labeling with the β-galactosidase antibody and cell-type-specific markers.

4.8 Summary and future directions

In summary, using the gene trap strategy in ES cells and chimera transgenesis, I have identified a novel mouse gene that encodes multiple alternatively spliced transcripts. One of the RNA isoform that corresponds to the 1574 kb cDNA, possibly the 1.7 kb message, was characterized and found to encode a putative protein with a bromodomain, a feature of proteins implicated in transcriptional regulation. While the bromodomain is conserved in proteins of many species, the novel gene identified in this study appears to be restricted to mammals. Thus, this gene may play a role in the mechanism of transcriptional regulation shared only by mammals. The major transcript (7 kb) encoded by the trapped gene locus was disrupted and the interrupted message created a fusion transcript with the inserted gene trap sequence. The formation of other RNA isoforms, however, was not apparently altered by the gene trap event. The fusion transcript encodes a fusion protein that is localized in neuronal nuclei. The endogenous 7 kb transcript that contributes to the fusion transcript is highly abundant in neural tissues, and is developmentally upregulated following embryonic cell differentiation to the neuronal lineage. The nuclear localization of the protein encoded by the 7 kb transcript, along with the fact that one of the RNA isoforms, possibly the 1.7kb transcript, encodes a bromodomain containing protein, is consistent with the notion that the novel gene identified in this study may encode a family of functionally related

components of mammalian transcriptional complexes. While proteins encoded by the minor transcripts may play important roles in ubiquitous mammalian transcriptional regulation, the major transcript (7 kb) encodes a protein that may be involved in transcriptional activation shared only by neurons. It is also conceivable that while the 1.7 kb RNA isoform encodes a bromodomain containing protein, other transcripts, including the neuron-specific 7 kb major transcript, may encode domains such as the ATPase and helicase motifs that are involved in transcriptional activation.

Complete understanding of the structure and the function of the endogenous gene requires that cDNAs corresponding to the different alternatively spliced isoforms be characterized. To this end, the 119 bp RACE derived DNA, the 735 bp and 1574 bp cDNA clones can be used as probes to screen mouse cDNA libraries. Since the 7 kb message is highly abundant in neural tissues, whereas the minor transcripts are more prominent in ES cell and at embryonic stage, mouse brain cDNA libraries and ES cell or embryonic cDNA libraries should be used in the cDNA cloning. This would increase the opportunity to obtain full cDNA clones that correspond to the 7 kb major transcript and the other RNA isoforms, respectively. Moreover, the endogenous pattern of tissue-specific and developmentally regulated expression of different transcripts could be determined by *in situ* hybridization using cDNA probes characteristic for each transcript. Identification of the endogenous expression pattern of different transcripts would not only provide clues to their developmental function, but also offers ultimate proof of the pan-neuronal expression of the 7 kb RNA.

The failure of germline transmission of the GT9.6 chimeras has precluded the analysis of insertional mutation in transgenic mice. However, functional investigation of the identified gene can be achieved by a gene

targeting approach. Different domains of interest, such as the bromodomain, can be selectively knocked out in an attempt to determine their respective functions. A prerequisite for the knock out experiments is the genomic cloning and characterization of the gene locus. Given the availability of the cDNA probes and various mouse genomic libraries, especially the ES cell genomic library derived from the 129 inbred strain that is genotypically identical to the genetic background of the CCE ES cell used in this study, cloning the entire endogenous locus identified in this investigation does not seem to be a formidable task.

In collaboration with Drs. X. Zhang, and P. Eydoux, the novel mouse gene has been localized on mouse chromosome 2q, and its human counterpart on human chromosome 20q. The availability of the chromosome specific mouse and human genomic DNA libraries will greatly facilitate the genomic cloning. The potential transcriptional activation function exerted by proteins encoded by the trapped gene could be tested *in vitro* once the cDNA corresponding to each endogenous message has been fully characterized. Since the bromodomain has been shown to potentiate transcriptional activation of a number of nuclear receptors (Muchardt and Yaniv, 1993; Singh et al., 1995), similar test could be performed on the *in vitro* translated proteins of the endogenous gene. In addition, footprinting and gel retardation experiments could be carried out to identify the potential downstream targets of the proteins. A number of potential kinase phosphorylation sites have been identified along with the bromodomain in the protein encoded by the 1.7 kb RNA. It would be interesting to see whether the phosphorylation status of the *in vitro* translated protein regulates its transcriptional activation function. It could be achieved by modifying the phosphorylation status of the *in vitro* translated protein and analyzing the effects of altered

phosphorylation of the proteins on their capacity to interact with other proteins or DNA.

Chimeras contain the GT9.6 ES cell clone, and therefore a convenient pan-neuronal marker. They were utilized in this investigation to analyze the pattern of neurogenesis in the mouse cerebral cortex and the spinal cord. This study further demonstrated the spatially restricted pattern of cell mixing and migration may be a common mechanism of cortical and spinal cord genesis. A component of radial migration pattern during corticalgenesis was demonstrated without obvious "contamination" by the mode of tangential dispersion. These radial stripes may represent pathways that the clonally related neurons within a compartment follow during their migration. Moreover, a highly spatially regulated pattern of generation was demonstrated during spinal cord genesis. It was demonstrated that during early spinal cord development, cells do not mix randomly and extensively with each other. Clonal progeny of a common progenitor tend to stay together as cohort patches on all three dimensions. The neural crest cells that give rise to the DRG seem to be most representative of their clonal origin in the neural tube. The DRGs and their closely associated spinal cord segments were labeled approximately to the same degree. Evidently, the GT9.6 chimeras could be used to analyze the developmental pattern of other parts of the nervous system, notably the cerebellum and the retina.

In conclusion, the gene trap strategy has provided a powerful screening method to identify genes that may play important roles during mouse development. Combined with the ES-chimera-transgenic technology, it offers an alternative but effective approach to uncover mammalian developmentally important genes. Chimeras produced during the application

of this method may offer an extremely useful means to study developmental pattern formation during organogenesis.

Contribution to original knowledge

In the study conducted in this thesis, I have identified a novel mammalian gene by gene trapping in mouse embryonic stem cells and chimera transgenesis. The newly discovered gene shared the following features in its molecular structure and expression pattern, and is believed to play a role in transcriptional regulation of mammalian nervous system:

1) It encodes multiple transcripts generated by alternative pre-mRNA splicing.

2) The different RNA isoforms are differentially expressed during embryogenesis and neurogenesis, with one major transcript exhibiting neuron-specific expression in the developing and mature nervous system.

3) The neuron-specific expression of the major transcript suggests the existence of pan-neuronal splicing machinery with unique properties, and the presence of neural specific components involved in transcriptional regulation.

4) The protein encoded by the major transcript is localized in neuronal nuclei, providing a useful pan-neuronal nuclear marker, and is consistent with a role it may play in transcriptional regulation.

5) One of the minor transcripts characterized in this investigation encodes a bromodomain containing protein. This putative protein thus shared the characteristics of many transcription regulators involved in transcriptional activation.

6) The novel mouse gene identified in this study is conserved among mammals, but has not been revealed in other organisms. The gene locus has been localized on mouse chromosome 2q and human syntenic chromosome 20q. No previously identified genes on the syntenic chromosome locations

share any known structural features nor expression patterns consistent with the newly identified mouse gene.

7) Analyses of chimeras constructed by the gene trapped ES cell clone revealed a spatially restricted pattern of neurogenesis during the development of the cerebral cortex and the spinal cord. Some of these patterns were predicted by previous investigations, but others suggest novel patterns of cell mixing and cell migration during the development of the central nervous system.

References

Aebi M., Weissmann C. Precision and orderliness in splicing. Trends in Genetics (3)4:102-107, 1987.

Albagli O. Soudant N. Ferreira E. Dhordain P. Dewitte F. Begue A. Flourens A. Stehelin D. Leprince D. A model for gene evolution of the ets-1/ets-2 transcription factors based on structural and functional homologies. Oncogene. 9(11):3259-71, 1994.

Allen ND., Cran DG., Barton SC., Hettle S., Reik W., Surani AM. Transgene as probes for active chromosomal domains in mouse development. Nature 333:852-855, 1988.

Austin CP. Cepko CL. Cellular migration patterns in the developing mouse cerebral cortex. Development. 110(3):713-32, 1990.

Aviv, H., Leder, P. Purification of biologically active globin messenger RNA by chromatography on oligothymidylic acid-cellulose. Proceedings of the National Academy of Sciences of the United States of America 69:1408-1412, 1972.

Axelrod HR. Embryonic stem cell lines derived from blastocyst by a simple technique. Developmental Biology 101:225-228, 1984.

Balling R., Deutsch U., Gruss P. *undulated*, a mutation affecting the development of the mouse skeleton, has a point mutation in the paired box of *Pax 1*. Cell 55:531-535, 1988

Beck S. Hanson I. Kelly A. Pappin DJ. Trowsdale J. A homologue of the Drosophila female sterile homeotic (fsh) gene in the class II region of the human MHC. DNA Sequence. 2(4):203-10, 1992.

Beddington RSP., Morgernstern J., Land H., Hogan A. An *in situ* transgenic enzyme marker for the midgestation mouse embryo and the visualization of inner cell mass clones during early embryogenesis. Development 106:37-46, 1989

Bellen HJ. O'Kane CJ. Wilson C. Grossniklaus U. Pearson RK. Gehring WJ. P-element-mediated enhancer detection: a versatile method to study development in Drosophila. Genes & Development. 3(9):1288-300, 1989.

Bellomy G. R., Record M. T. Jr. A convenient moderate-scale procedure for obtaining DNA from bacteriophage lamda. Biotechniques 7 1: 21-23, 1989.

Bier E. Vaessin H. Shepherd S. Lee K. McCall K. Barbel S. Ackerman L. Carretto R. Uemura T. Grell E. et al. Searching for pattern and mutation in the Drosophila genome with a P-lacZ vector. Genes & Development. 3(9):1273-87, 1989.

Boncinelli E. Simeone A. Acampora D. Mavilio F. HOX gene activation by retinoic acid. Trends in Genetics. 7(10):329-34, 1991.

Bonnerot C., Grimber G., Briand P., Nicolas JF. Patterns of expression of position-dependent integrated transgenes in mouse embryo. Proceedings of the National Academy of Sciences of the United States of America 87:6331-6335, 1990.

Bopp D. Burri M. Baumgartner S. Frigerio G. Noll M. Conservation of a large protein domain in the segmentation gene paired and in functionally related genes of Drosophila. Cell.47(6):1033-40, 1986

Bowes C. Li T. Danciger M. Baxter LC. Applebury ML. Farber DB. Retinal degeneration in the rd mouse is caused by a defect in the beta subunit of rod cGMP-phosphodiesterase. Nature. 347(6294):677-80, 1990.

Bradley A., Evans M., Kaufman MH., Robertson R. Formation of germ-line chimeras from embryo-derived teratocarcinoma cell lines. Nature 309:255-256, 1984

Brenner DG. Lin-Chao S. Cohen SN. Analysis of mammalian cell genetic regulation in situ by using retrovirus-derived "portable exons" carrying the Escherichia coli lacZ gene. Proceedings of the National Academy of Sciences of the United States of America. 86(14):5517-21, 1989.

Brown A. Copeland NG. Gilbert DJ. Jenkins NA. Rossant J. Kothary R. The genomic structure of an insertional mutation in the Dystonia musculorum locus. Genomics. 20(3):371-6, 1994.

Bultman SJ. Michaud EJ. Woychik RP. Molecular characterization of the mouse agouti locus. Cell. 71(7):1195-204, 1992.

Butner K., Lo CW. Modulation of *tk* expression in mouse pericentromeric heterochromatin. Molecular and Cell Biology 6:12:4440-4449, 1986.

Camper SA. Saunders TL. Katz RW. Reeves RH. The Pit-1 transcription factor gene is a candidate for the murine Snell dwarf mutation. Genomics. 8(3):586-90, 1990.

Campione-Piccardo J. Sun JJ. Craig J. McBurney MW. Cell-cell interaction can influence drug-induced differentiation of murine embryonal carcinoma cells. Developmental Biology. 109(1):25-31, 1985.

Capecchi MR. The new mouse genetics: altering the genome by gene targeting. Trends in Genetics 5:70-76, 1989.

Capobianco AJ. Simmons DL. Gilmore TD. Cloning and expression of a chicken c-rel cDNA: unlike p59v-rel, p68c-rel is a cytoplasmic protein in chicken embryo fibroblasts. Oncogene. 5(3):257-65, 1990.

Casadaban MJ., Chou J., Cohen SN. In vitro gene fusions that join an enzymatically active β-galactosidase segment to amino-terminal fragments of exogenous proteins: *Escherichia Coli* plasmid vectors for the detection and cloning of translational initiation signals. Journal of Bacteriology 143:971-980, 1980.

Chalepakis G. Fritsch R. Fickenscher H. Deutsch U. Goulding M. Gruss P. The molecular basis of the undulated/Pax-1 mutation. Cell. 66(5):873-84, 1991.

Chambers CA. Kang J. Pawling J. Huber B. Hozumi N. Nagy A. Exogenous Mtv-7 superantigen transgene expression in major histocompatibility complex class II I-E- mice reconstituted with embryonic stem cell-derived hematopoietic stem cells. Proceedings of the National Academy of Sciences of the United States of America. 91(3):1138-42, 1994.

Chrivia JC. Kwok RP. Lamb N. Hagiwara M. Montminy MR. Goodman RH. Phosphorylated CREB binds specifically to the nuclear protein CBP. Nature. 365(6449):855-9, 1993 Oct 28.

Chomczynski, P., Sacchi, N. Single-step method of RNA isolation by guanidium thiocyanate-phenol-chloroform extraction. Analytical Bochemistry 162:156-159, 1987.

Clerc RG. Corcoran LM. LeBowitz JH. Baltimore D. Sharp PA. The B-cell-specific Oct-2 protein contains POU box- and homeo box-type domains. Genes & Development. 2(12A):1570-81, 1988.

Conlon FL. Lyons KM. Takaesu N. Barth KS. Kispert A. Herrmann B. Robertson EJ. A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. Development. 120(7):1919-28, 1994.

Copeland NG. Gilbert DJ. Cho BC. Donovan PJ. Jenkins NA. Cosman D. Anderson D. Lyman SD. Williams DE. Mast cell growth factor maps near the steel locus on mouse chromosome 10 and is deleted in a number of steel alleles. Cell. 63(1):175-83, 1990.

Cote J. Quinn J. Workman JL. Peterson CL. Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. Science. 265(5168):53-60, 1994.

Crandall JE. Herrup K. Patterns of cell lineage in the cerebral cortex reveal evidence for developmental boundaries. Experimental Neurology. 109(1):131-9, 1990.

Crick F. Split genes and RNA splicing. Science. 204(4390):264-71, 1979.

Darcangelo G. Miao GG. Chen SC. Soares HD. Morgan JI. Curran T. A protein related to extracellular matrix proteins deleted in the mouse mutant reeler Nature. 374(6524):719-723, 1995.

Davis AC., Noble-Topham SE., Rossant J., Joyner AL. Expression of the homeo box-containing gene En-2 delineates a specific region of the developing mouse brain. Genes & Development 2:361-371, 1988.

Davis CA., Joyner A. Expression patterns of the homeo box-containing genes *En*-1 and *En*-2 and the proto-oncogene *int*-1 diverge during mouse development. Genes & Development 2:1736-1744, 1988.

Davis RL., Weintraub H., Lassar AB. Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell 51:987-1000, 1987

De Robertis EM. Longthorne RF. Gurdon JB. Intracellular migration of nuclear proteins in Xenopus oocytes. Nature. 272(5650):254-6, 1978.

DeChiara TM., Efstratiadis A., Robertson EJ. A growth-deficiency phenotype in heterozygous mice carrying an insulin-like growth factor II gene disrupted by targeting. Nature 345:7880, 1990.

Delmas V. Stokes DG. Perry RP. A mammalian DNA-binding protein that contains a chromodomain and an SNF2/SWI2-like helicase domain. Proceedings of the National Academy of Sciences of the United States of America. 90(6):2414-8, 1993.

Deutsch U. Dressler GR. Gruss P. Pax 1, a member of a paired box homologous murine gene family, is expressed in segmented structures during development. Cell. 53(4):617-25, 1988.

Dodd J., Jessell TM. Axon guidance and the patterning of neuronal projections in vertebrates.

Science 242:692-699, 1988.

Doe CQ., Hiromi Y., Gehring WJ., Goodman CS. Expression and function of the segmentation gene *fushi tarazu* during drosophila neurogenesis. Science 239:170-175, 1988.

Doetschman T., Eistetter H., Katz M., Schmidt W., Kemler R. The *in vitro* development of blastocyst-derived embryonic stem cell lines: formation of visceral yolk sac, blood island and myocardium. Journal of Embryology and Experimental Morphology 87:27-45, 1985.

Doetschman T., Gossler A., Serfling E., Schaffner W., Marcu K., df Stanton L., and Kemler R. Introduction of genes into mouse embryonic stem cells. Progress in Developmental Biology Part A:47-50, 1986.

Dolle P. Castrillo JL. Theill LE. Deerinck T. Ellisman M. Karin M. Expression of GHF-1 protein in mouse pituitaries correlates both temporally and spatially with the onset of growth hormone gene activity. Cell. 60(5):809-20, 1990.

Dolle P. Izpisua-Belmonte JC. Falkenstein H. Renucci A. Duboule D. Coordinate expression of the murine Hox-5 complex homoeobox-containing genes during limb pattern formation. Nature. 342(6251):767-72, 1989.

Dunaief JL. Strober BE. Guha S. Khavari PA. Alin K. Luban J. Begemann M. Crabtree GR. Goff SP. The retinoblastoma protein and BRG1 form a complex and cooperate to induce cell cycle arrest. Cell. 79(1):119-30, 1994.

Dynan WS. Modularity in promoters and enhancers. Cell 58:1-4, 1989.

Dynan WS., Tjian R. Control of eukaryotic messenger RNA synthesis by sequence-specific DNA binding proteins. Nature 316:774-778, 1985.

Eckner R. Ewen ME. Newsome D. Gerdes M. DeCaprio JA. Lawrence JB. Livingston DM. Molecular cloning and functional analysis of the adenovirus E1A-associated 300-kD protein (p300) reveals a protein with properties of a transcriptional adaptor. Genes & Development. 8(8):869-84, 1994.

Eisenmann DM. Arndt KM. Ricupero SL. Rooney JW. Winston F. SPT3 interacts with TFIID to allow normal transcription in Saccharomyces cerevisiae. Genes & Development. 6(7):1319-31, 1992.

Elfring LK. Deuring R. McCallum CM. Peterson CL. Tamkun JW. Identification and characterization of Drosophila relatives of the yeast transcriptional activator SNF2/SWI2. Molecular & Cellular Biology. 14(4):2225-34, 1994.

Epstein DJ. Vekemans M. Gros P. Splotch (Sp2H), a mutation affecting development of the mouse neural tube, shows a deletion within the paired homeodomain of Pax-3.Cell. 67(4):767-74, 1991.

Evans MJ., Kaufman MH. Establishment in culture of pluripotential cells from mouse embryos. Nature 292:154-156, 1981.

Finney M. Ruvkun G. Horvitz HR. The C. elegans cell lineage and differentiation gene unc-86 encodes a protein with a homeodomain and extended similarity to transcription factors. Cell. 55(5):757-69, 1988.

Fishell G. Mason CA. Hatten ME. Dispersion of neural progenitors within the germinal zones of the forebrain [published erratum appears in Nature 1993 May 20;363(6426):286. Nature. 362(6421):636-8, 1993.

Fishell G. Rossant J. van der Kooy D. Neuronal lineages in chimeric mouse forebrain are segregated between compartments and in the rostrocaudal and radial planes. Developmental Biology. 141(1):70-83, 1990.

Flanagan JG. Leder P. The kit ligand: a cell surface molecule altered in steel mutant fibroblasts. Cell. 63(1):185-94, 1990.

Forrester LM. Bernstein A. Rossant J. Nagy A. Long-term reconstitution of the mouse hematopoietic system by embryonic stem cell-derived fetal liver. Proceedings of the National Academy of Sciences of the United States of America. 88(17):7514-7, 1991.

Friedrich G. Soriano P. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. Genes & Development. 5(9):1513-23, 1991.

Frohman MA. Rapid amplification of complementary DNA ends for generation of full-length complementary DNAs: thermal RACE. Methods in Enzymology. 218:340-56, 1993.

Frohman MA., Dush MK., Martin G. Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. Proceedings of the National Academy of Sciences of the United States of America 85:8998-9002, 1988.

Gavin BJ., McMahon JA., McMahon AP. 1990 Expression of multiple novel Wnt-1/int-1-related genes during fetal and adult mouse development. Genes & Development 4:2319-2332

Gearing DP., Nicholas MG., King JA., Hilton DJ., Nicola NA., Simpson RJ., Nice EC., Kelso A., Metcalf D. Molecular cloning and expression of cDNA encoding a murine myeloid leudaemia inhibitory factor LIF. EMBO 6:13:3993-4002, 1987.

Gehring WJ. Qian YQ. Billeter M. Furukubo-Tokunaga K. Schier AF. Resendez-Perez D. Affolter M. Otting G. Wuthrich K. Homeodomain-DNA recognition. Cell. 78(2):211-23, 1994.

Geissler EN. Cheng SV. Gusella JF. Housman DE. Genetic analysis of the dominant white-spotting (W) region on mouse chromosome 5: identification of cloned DNA markers near W. Proceedings of the National Academy of Sciences of the United States of America. 85(24):9635-9, 1988a.

Geissler EN. McFarland EC. Russell ES. Analysis of pleiotropism at the dominant white-spotting (W) locus of the house mouse: a description of ten new W alleles. Genetics. 97(2):337-61, 1981.

Geissler EN. Ryan MA. Housman DE. The dominant-white spotting (W) locus of the mouse encodes the c-kit proto-oncogene. Cell. 55(1):185-92, 1988b.

Georgakopoulos T. Thireos G. Two distinct yeast transcriptional activators require the function of the GCN5 protein to promote normal levels of transcription. EMBO Journal. 11(11):4145-52, 1992.

Ghysen A., Dambly-Chaudiere C. Genesis of the *Drosophila* peripheral nervous system. Trends in Genetics 5:251-253, 1989.

Ghysen A., O'Kane C. Neural enhancer-like elements as specific cell markers in Drosophila. Development 105, 35-52, 1989.

Gibson F. Walsh J. Mburu P. Varela A. Brown KA. Antonio M. Beisel KW. Steel KP. Brown SD. A type VII myosin encoded by the mouse deafness gene shaker-1. Nature. 374(6517):62-4, 1995.

Goldowitz D. Cell partitioning and mixing in the formation of the CNS: analysis of the cortical somatosensory barrels in chimeric mice. Brain Research. 432(1):1-9, 1987.

Goldowitz D. The weaver granuloprival phenotype is due to intrinsic action of the mutant locus in granule cells: evidence from homozygous weaver chimeras. Neuron. 2(6):1565-75, 1989.

Goodfellow PJ. Darling SM. Thomas NS. Goodfellow PN. A pseudoautosomal gene in man. Science. 234(4777):740-3, 1986.

Gorbalenya AE. Koonin EV. Donchenko AP. Blinov VM. A novel superfamily of nucleoside triphosphate-binding motif containing proteins which are probably involved in duplex unwinding in DNA and RNA replication and recombination. FEBS Letters. 235(1-2):16-24, 1988.

Gorbalenya AE. Koonin EV. Donchenko AP. Blinov VM. Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. Nucleic Acids Research. 17(12):4713-30, 1989.

Gossler A. Joyner AL. Rossant J. Skarnes WC. Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. Science. 244(4903):463-5, 1989.

Gossler A., Doetschman T., Korn R., Serfling E., Kemler R. Transgenesis by means of blastocyst-derived embryonic stem cell lines. Proceedings of the National Academy of Sciences of the United States of America 83:9065-9069, 1986.

Gough NM., Gearing DP., KIng JA., Willson TA., Hilton DJ., Nicola NA., Metcalf D. Molecular cloning and expression of the human homologue of the murine gene encoding myeloid leukemia inhibitory factor. Proceedings of the National Academy of Sciences of the United States of America 85:2623-2627, 1988

Graziadei PP. Monti Graziadei AG. Regeneration in the olfactory system of vertebrates. American Journal of Otolaryngology. 4(4):228-33, 1983.

Green EL. Biology of the laboratory mouse. Jackson laboratory, Green EL editor, 2nd ed. New York, Blakiston Division, McGraw-Hill, 1966.

Green MR. Pre-mRNA splicing. Annual Review of Genetics 20:671-708, 1986.

Gridley T. Gray DA. Orr-Weaver T. Soriano P. Barton DE. Francke U. Jaenisch R. Molecular analysis of the Mov 34 mutation: transcript disrupted by proviral integration in mice is conserved in Drosophila. Development. 109(1):235-42, 1990.

Gridley T. Jaenisch R. Gendron-Maguire M. The murine Mov-34 gene: full-length cDNA and genomic organization. Genomics. 11(3):501-7, 1991.

Grosveld F., van Assendelft GB., Greaves DR., Kollias G. Position-independent, high-level

expression of the human β-globin gene in transgenic mice. Cell 51:975-985, 1987.

Grove EA. Williams BP. Li DQ. Hajihosseini M. Friedrich A. Price J. Multiple restricted lineages in the embryonic rat cerebral cortex. Development. 117(2):553-61, 1993.

Gruss C. Sogo JM. Chromatin replication. Bioessays. 14(1):1-8, 1992.

Gubbay J. Collignon J. Koopman P. Capel B. Economou A. Munsterberg A. Vivian N. Goodfellow P. Lovell-Badge R. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. Nature. 346(6281):245-50, 1990.

Hanley M. Proto-oncogenes in the nervous system Neuron 1:175-182, 1988.

Hannink M. Temin HM. Transactivation of gene expression by nuclear and cytoplasmic rel proteins. Molecular & Cellular Biology. 9(10):4323-36, 1989.

Hartman SC. Mulligan RC. Two dominant-acting selectable markers for gene transfer studies in mammalian cells. Proceedings of the National Academy of Sciences of the United States of America. 85(21):8047-51, 1988.

Hartung S. Jaenisch R. Breindl M. Retrovirus insertion inactivates mouse alpha 1(I) collagen gene by blocking initiation of transcription. Nature. 320(6060):365-7, 1986.

Hartung S., Jaenisch R., Breindl M. Retrovirus insertion inactivates mouse a1I collagen gene by blocking initiation of transcription. Nature 320:365-367, 1986.

Haynes SR. Dollard C. Winston F. Beck S. Trowsdale J. Dawid IB. The bromodomain: a conserved sequence found in human, Drosophila and yeast proteins. Nucleic Acids Research. 20(10):2603, 1992.

Haynes SR. Mozer BA. Bhatia-Dey N. Dawid IB. The Drosophila fsh locus, a maternal effect homeotic gene, encodes apparent membrane proteins. Developmental Biology. 134(1):246-57, 1989.

Henikoff S. Transcriptional activator components and poxvirus DNA-dependent ATPases comprise a single family. Trends in Biochemical Sciences. 18(8):291-2, 1993.

Herr W. Sturm RA. Clerc RG. Corcoran LM. Baltimore D. Sharp PA. Ingraham HA. Rosenfeld MG. Finney M. Ruvkun G. et al. The POU domain: a large conserved region in the mammalian pit-1, oct-1, oct-2, and Caenorhabditis elegans unc-86 gene products. Genes & Development. 2(12A):1513-6, 1988.

Herrmann BG., Labeit S., Poustka A., King TR., Lehrach H. Cloning of the T gene required in mesoderm formation in the mouse. Nature 343:617-622, 1990.

Hill RE. Favor J. Hogan BL. Ton CC. Saunders GF. Hanson IM. Prosser J. Jordan T. Hastie ND. van Heyningen V. Mouse small eye results from mutations in a paired-like homeobox-containing gene [published erratum appears in Nature 1992 Feb 20;355(6362):750]. Nature. 354(6354):522-5, 1991.

Hisatake K. Hasegawa S. Takada R. Nakatani Y. Horikoshi M. Roeder RG. The p250 subunit of native TATA box-binding factor TFIID is the cell-cycle regulatory protein CCG1. Nature.

362(6416):179-81, 1993.

Hodgkinson CA. Moore KJ. Nakayama A. Steingrimsson E. Copeland NG. Jenkins NA. Arnheiter H. Mutations at the mouse microphthalmia locus are associated with defects in a gene encoding a novel basic-helix-loop-helix-zipper protein. Cell. 74(2):395-404, 1993.

Hodgman TC. A new superfamily of replicative proteins [published erratum appears in Nature 1988 Jun 9;333(6173):578]. Nature. 333(6168):22-3, 1988.

Hogan B., Costantini F., Lacy E. Manipulating the mouse embryo - A laboratory manual. Cold Spring Harbor Laboratory, 1986

Hosler BA. LaRosa GJ. Grippo JF. Gudas LJ. Expression of REX-1, a gene containing zinc finger motifs, is rapidly reduced by retinoic acid in F9 teratocarcinoma cells. Molecular & Cellular Biology. 9(12):5623-9, 1989.

Hui CC. Joyner AL. A mouse model of greig cephalopolysyndactyly syndrome: the extra-toesJ mutation contains an intragenic deletion of the Gli3 gene. Nature Genetics. 3(3):241-6, 1993.

Ingraham HA. Chen RP. Mangalam HJ. Elsholtz HP. Flynn SE. Lin CR. Simmons DM. Swanson L. Rosenfeld MG. A tissue-specific transcription factor containing a homeodomain specifies a pituitary phenotype. Cell. 55(3):519-29, 1988.

Ingraham HA. Flynn SE. Voss JW. Albert VR. Kapiloff MS. Wilson L. Rosenfeld MG. The POU-specific domain of Pit-1 is essential for sequence-specific, high affinity DNA binding and DNA-dependent Pit-1-Pit-1 interactions. Cell. 61(6):1021-33, 1990.

Jackson IJ. Chambers D. Rinchik EM. Bennett DC. Characterization of TRP-1 mRNA levels in dominant and recessive mutations at the mouse brown (b) locus. Genetics. 126(2):451-9, 1990.

Jackson-Grusby L. Kuo A. Leder P. A variant limb deformity transcript expressed in the embryonic mouse limb defines a novel formin. Genes & Development. 6(1):29-37, 1992.

Jaenisch R. Germ line integration and Mendelian transmission of the exogenous Moloney leukemia virus. Proceedings of the National Academy of Sciences of the United States of America. 73(4):1260-4, 1976.

Jaenisch R. Germ line integration of moloney leukemia virus: effect of homozygosity at the m-mulV locus. Cell. 12(3):691-6, 1977.

Jaenisch R. Harbers K. Schnieke A. Lohler J. Chumakov I. Jahner D. Grotkopp D. Hoffmann E. Germline integration of moloney murine leukemia virus at the Mov13 locus leads to recessive lethal mutation and early embryonic death. Cell. 32(1):209-16, 1983.

Jaenisch R. Jahner D. Nobis P. Simon I. Lohler J. Harbers K. Grotkopp D. Chromosomal position and activation of retroviral genomes inserted into the germ line of mice. Cell. 24(2):519-29, 1981.

Jaenisch R. Retroviruses and embryogenesis: microinjection of Moloney leukemia virus into midgestation mouse embryos. Cell. 19(1):181-8, 1980.

Jaenisch R. Transgenic animals. Science 240:1468-1474, 1988.

Jenkins NA. Copeland NG. Taylor BA. Lee BK. Dilute (d) coat colour mutation of DBA/2J mice is associated with the site of integration of an ecotropic MuLV genome. Nature. 293(5831):370-4, 1981.

Joyner AL., Martin GR. En-1 and En-2, two mouse genes with sequence homology to the Drosophila engrailed gene: expression during embryogenesis. Genes & Development 1:29-38, 1987.

Kalderon D. Roberts BL. Richardson WD. Smith AE. A short amino acid sequence able to specify nuclear location. Cell. 39(3 Pt 2):499-509, 1984.

Kessel M. Gruss P. Homeotic transformations of murine vertebrae and concomitant alteration of Hox codes induced by retinoic acid. Cell. 67(1):89-104, 1991.

Kessel M., Gruss P. Murine developmental control genes. Science 249:374-379, 1990.

Keynes R., Lumsden A. Segmentation and the origin of regional diversity in the vertebrate central nervous system. Neuron 2:1-9, 1990.

Khavari PA. Peterson CL. Tamkun JW. Mendel DB. Crabtree GR. BRG1 contains a conserved domain of the SWI2/SNF2 family necessary for normal mitotic growth and transcription. Nature. 366(6451):170-4, 1993.

Kirkwood TB. Price J. Grove EA. The dispersion of neuronal clones across the cerebral cortex [letter; comment]. Science. 258(5080):317-20, 1992.

Klar AJ. A model for specification of the left-right axis in vertebrates. Trends in Genetics. 10(11):392-6, 1994.

Korn R. Schoor M. Neuhaus H. Henseling U. Soininen R. Zachgo J. Gossler A. Enhancer trap integrations in mouse embryonic stem cells give rise to staining patterns in chimaeric embryos with a high frequency and detect endogenous genes. Mechanisms of Development. 39(1-2):95-109, 1992.

Kothary R., Clapoff S., Brown A., Campbell R., Peterson A., Rossant J. A transgene containing lacZ inserted into the dystonia locus is expressed in neural tube. Nature 335:435-437, 1988.

Kothary R., Clapoff S., Darling S., Perry MD., Moran L., Rossant J. Inducible expression of an hsp68-lacZ gene in transgenic mice. Development 105:707-714, 1989.

Kratochwil K. von der Mark K. Kollar EJ. Jaenisch R. Mooslehner K. Schwarz M. Haase K. Gmachl I. Harbers K. Retrovirus-induced insertional mutation in Mov13 mice affects collagen I expression in a tissue-specific manner. Cell. 57(5):807-16, 1989.

Krumlauf R. Hox genes in vertebrate development. Cell. 78(2):191-201, 1994.

Krushel LA. Johnston JG. Fishell G. Tibshirani R. van der Kooy D. Spatially localized neuronal cell lineages in the developing mammalian forebrain. Neuroscience. 53(4):1035-47, 1993.

Kuehn MR., Bradley A., Robertson EJ., Evans MJ. A potential animal model for Lesch-Nyhan syndrome through introduction of HPRT mutations into mice. Nature 326:295-298, 1987.

Kwok RP. Lundblad JR. Chrivia JC. Richards JP. Bachinger HP. BrennanRG. Roberts SG.

Green MR. Goodman RH. Nuclear protein CBP is a coactivator for the transcription factor CREB [see comments]. Nature. 370(6486):223-6, 1994

Kwon BS. Haq AK. Pomerantz SH. Halaban R. Isolation and sequence of a cDNA clone for human tyrosinase that maps at the mouse c-albino locus [published erratum appears in Proc Natl Acad Sci U S A 1988 Sep;85(17):6352]. Proceedings of the National Academy of Sciences of the United States of America. 84(21):7473-7, 1987.

Lanford RE. Butel JS. Construction and characterization of an SV40 mutant defective in nuclear transport of T antigen. Cell. 37(3):801-13, 1984.

LaRosa GJ., Gudas LJ. Early retinoic acid-induced F9 teratocardinoma stem cell gene ERA-1: alternate splicing creats transcripts for a homeobox-containing protein and one lacking the homeobox. Molecular and Cell Biology 89:3906-3917, 1988.

Latchman DS. Dent CL. Lillycrop KA. Wood JN. POU family transcription factors in sensory neurons. Biochemical Society Transactions. 20(3):627-31, 1992.

Laurent BC. Carlson M. Yeast SNF2/SWI2, SNF5, and SNF6 proteins function coordinately with the gene-specific transcriptional activators GAL4 and Bicoid [published erratum appears in Genes Dev 1992 Nov;6(11):2233]. Genes & Development. 6(9):1707-15, 1992.

Laurent BC. Treich I. Carlson M. The yeast SNF2/SWI2 protein has DNA-stimulated ATPase activity required for transcriptional activation. Genes & Development. 7(4):583-91, 1993.

Laurent BC. Treitel MA. Carlson M. Functional interdependence of the yeast SNF2, SNF5, and SNF6 proteins in transcriptional activation. Proceedings of the National Academy of Sciences of the United States of America. 88(7):2687-91, 1991.

Laurent BC. Yang X. Carlson M. An essential Saccharomyces cerevisiae gene homologous to SNF2 encodes a helicase-related protein in a new family. Molecular & Cellular Biology. 12(4):1893-902, 1992.

Laurie DJ. Seeburg PH. Regional and developmental heterogeneity in splicing of the rat brain NMDAR1 mRNA. Journal of Neuroscience. 14(5 Pt 2):3180-94, 1994.

Leber SM. Breedlove SM. Sanes JR. Lineage, arrangement, and death of clonally related motoneurons in chick spinal cord. Journal of Neuroscience. 10(7):2451-62, 1990.

Leber SM. Sanes JR. Migratory paths of neurons and glia in the embryonic chick spinal cord Journal of Neuroscience. 15(2):1236-1248, 1995.

Leber SM., Breedlove SM., Sanes JR. Lineage, arrangement, and death of clonally related motoneruons in chick spinal cord. The Journal of Neuronscience 107:2451-2462, 1990.

Lemieux N., Dutrillaux B., Viegas-Pequignot E. A simple method for simultaneous R- or G-banding and flourescence in situ hybridization of small single copy genes. Cytogenetics and Cell Genetics 59:311-312, 1992

Li C. Tropak MB. Gerlai R. Clapoff S. Abramow-Newerly W. Trapp B. Peterson A. Roder J. Myelination in the absence of myelin-associated glycoprotein. Nature. 369(6483):747-50, 1994.

Li S. Crenshaw EB 3d. Rawson EJ. Simmons DM. Swanson LW. Rosenfeld MG. Dwarf locus

mutants lacking three pituitary cell types result from mutations in the POU-domain gene pit-1. Nature. 347(6293):528-33, 1990.

Linder P. Lasko PF. Ashburner M. Leroy P. Nielsen PJ. Nishi K. Schnier J. Slonimski PP. Birth of the D-E-A-D box [published erratum appears in Nature 1989 Jul 20;340(6230):246]. Nature. 337(6203):121-2, 1989.

Lis JT., Simon JA., Sutton CA. New heat shock puffs and b-galactosidase activity resulting from transformation of Drosophila with an hsp70-lacZ hybrid gene. Cell 35:403-410, 1983.

Lu D. Willard D. Patel IR. Kadwell S. Overton L. Kost T. Luther M. Chen W. Woychik RP. Wilkison WO. et al. Agouti protein is an antagonist of the melanocyte-stimulating-hormone receptor. Nature. 371(6500):799-802, 1994.

Luetteke NC. Phillips HK. Qiu TH. Copeland NG. Earp HS. Jenkins NA. Lee DC. The mouse waved-2 phenotype results from a point mutation in the EGF receptor tyrosine kinase. Genes & Development. 8(4):399-413, 1994.

Luetteke NC. Qiu TH. Peiffer RL. Oliver P. Smithies O. Lee DC. TGF alpha deficiency results in hair follicle and eye abnormalities in targeted and waved-1 mice. Cell. 73(2):263-78, 1993.

Lundblad JR. Kwok RP. Laurance ME. Harter ML. Goodman RH. Adenoviral E1A-associated protein p300 as a functional homologue of the transcriptional co-activator CBP.Nature.374(6517):85-8, 1995.

Luskin MB. Parnavelas JG. Barfield JA. Neurons, astrocytes, and oligodendrocytes of the rat cerebral cortex originate from separate progenitor cells: an ultrastructural analysis of clonally related cells. Journal of Neuroscience. 13(4):1730-50, 1993.

Luskin MB. Pearlman AL. Sanes JR. Cell lineage in the cerebral cortex of the mouse studied in vivo and in vitro with a recombinant retrovirus. Neuron. 1(8):635-47, 1988.

Luskin MB. Restricted proliferation and migration of postnatally generated neurons derived from the forebrain subventricular zone. Neuron. 11(1):173-89, 1993.

Lygerou Z. Conesa C. Lesage P. Swanson RN. Ruet A. Carlson M. Sentenac A. Seraphin B. The yeast BDF1 gene encodes a transcription factor involved in the expression of a broad class of genes including snRNAs. Nucleic Acids Research. 22(24):5332-40, 1994.

Lyon MF. Genetic variants and strains of the laboratory mouse. edited by. Lyon MF., and Searle AG. 2nd ed. Oxford <England>: New York: Oxford University Press: Stuttgart: New York: G. Fischer Verlag, 1989.

Macleod D. Lovell-Badge R. Jones S. Jackson I. A promoter trap in embryonic stem (ES) cells selects for integration of DNA into CpG islands. Nucleic Acids Research. 19(1):17-23, 1991.

Magnuson T. Experimental approaches to mammalian embryonic development. Rossant J., Pedersen RA., eds, pp. 437-474, Cambridge University Press, 1986.

Maniatis T., Goodbourn S., Fischer JA. Regulation of inducible and tissue-specific gene expression. Science 236:1237-1244, 1987.

Mann GB. Fowler KJ. Gabriel A. Nice EC. Williams RL. Dunn AR. Mice with a null mutation of

the TGF alpha gene have abnormal skin architecture, wavy hair, and curly whiskers and often develop corneal inflammation. Cell. 73(2):249-61, 1993.

Marcus GA. Silverman N. Berger SL. Horiuchi J. Guarente L. Functional similarity and physical association between GCN5 and ADA2: putative transcriptional adaptors. EMBO Journal. 13(20):4807-15, 1994.

Martin G. 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 United States of America 78:7634-7638, 1981.

Mclaren A. Mammalian chimeras Cambridge, <Eng.>: New York: Cambridge University Press, 1976.

McLean I. W., Nakane P. K. Periodate-Lysine-paraformaldehyde fixative. A new fixative for immunoelectron microscopy. Journal of Histochemistry and Cytochemistry 22: 1077-1083, 1974.

McMahon AP. Bradley A. The Wnt-1 (int-1) proto-oncogene is required for development of a large region of the mouse brain. Cell. 62(6):1073-85, 1990.

McNeish JD. Scott WJ Jr. Potter SS. Legless, a novel mutation found in PHT1-1 transgenic mice. Science. 241(4867):837-9, 1988.

Meisler MH. Insertional mutation of 'classical' and novel genes in transgenic mice. Trends in Genetics. 8(10):341-4, 1992.

Mercer JA. Seperack PK. Strobel MC. Copeland NG. Jenkins NA. Novel myosin heavy chain encoded by murine dilute coat colour locus [published erratum appears in Nature 1991 Aug 8;352(6335):547]. Nature. 349(6311):709-13, 1991.

Miller MW. Duhl DM. Vrieling H. Cordes SP. Ollmann MM. Winkes BM. Barsh GS. Cloning of the mouse agouti gene predicts a secreted protein ubiquitously expressed in mice carrying the lethal yellow mutation. Genes & Development. 7(3):454-67, 1993.

Mitchell PJ., Tjian R. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science 245:371-378, 1989.

Molineaux SM. Engh H. de Ferra F. Hudson L. Lazzarini RA. Recombination within the myelin basic protein gene created the dysmyelinating shiverer mouse mutation. Proceedings of the National Academy of Sciences of the United States of America. 83(19):7542-6, 1986.

Monuki ES. Kuhn R. Weinmaster G. Trapp BD. Lemke G. Expression and activity of the POU transcription factor SCIP. Science. 249(4974):1300-3, 1990.

Monuki ES. Weinmaster G. Kuhn R. Lemke G. SCIP: a glial POU domain gene regulated by cyclic AMP. Neuron. 3(6):783-93, 1989.

Moran E. Interaction of adenoviral proteins with pRB and p53. FASEB Journal. 7(10):880-5, 1993.

Moreau JF., Donaldson DD., Bennet F., Witek-Giannotti J., Clark SC., Wong GG. Leukaemia inhibitory factor is identical to the myeloid growth factor human interleukin for DA cells. Nature 336:690-692, 1988.

Moyer JH. Lee-Tischler MJ. Kwon HY. Schrick JJ. Avner ED. Sweeney WE. Godfrey VL. Cacheiro NL. Wilkinson JE. Woychik RP. Candidate gene associated with a mutation causing recessive polycystic kidney disease in mice. Science. 264(5163):1329-33, 1994.

Muchardt C. Yaniv M. A human homologue of Saccharomyces cerevisiae SNF2/SWI2 and Drosophila brm genes potentiates transcriptional activation by the glucocorticoid receptor. EMBO Journal. 12(11):4279-90, 1993.

Murphy P. Davidson DR. Hill RE. Segment-specific expression of a homoeobox-containing gene in the mouse hindbrain. Nature. 341(6238):156-9, 1989.

Musci T., Mullen R. Cell mixing in the spinal cords of mouse chimeras. Developmental Biology 152:133-144, 1992.

Musci TS. Mullen RJ. All-or-none craniorachischisis in Loop-tail mutant mouse chimeras. Development. 110(1):229-37, 1990.

Nagy A. Gocza E. Diaz EM. Prideaux VR. Ivanyi E. Markkula M. Rossant J. Embryonic stem cells alone are able to support fetal development in the mouse. Development. 110(3):815-21, 1990.

Nagy A. Rossant J. Nagy R. Abramow-Newerly W. Roder JC. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. Proceedings of the National Academy of Sciences of the United States of America. 90(18):8424-8, 1993.

Natsoulis G. Thomas W. Roghmann MC. Winston F. Boeke JD. Ty1 transposition in Saccharomyces cerevisiae is nonrandom. Genetics. 123(2):269-79, 1989.

Nave KA. Bloom FE. Milner RJ. A single nucleotide difference in the gene for myelin proteolipid protein defines the jimpy mutation in mouse. Journal of Neurochemistry. 49(6):1873-7, 1987.

Nehls M. Pfeifer D. Schorpp M. Hedrich H. Boehm T. New member of the winged-helix protein family disrupted in mouse and rat nude mutations. Nature. 372(6501):103-7, 1994.

Niwa H. Araki K. Kimura S. Taniguchi S. Wakasugi S. Yamamura K. An efficient gene-trap method using poly A trap vectors and characterization of gene-trap events. Journal of Biochemistry. 113(3):343-9, 1993.

O'Connor LT. Lauterborn JC. Gall CM. Smith MA. Localization and alternative splicing of agrin mRNA in adult rat brain: transcripts encoding isoforms that aggregate acetylcholine receptors are not restricted to cholinergic regions. Journal of Neuroscience. 14(3 Pt 1):1141-52, 1994.

O'Kane CJ., Gehring WJ. Detection in situ of genomic regulatory elements in Drosophila. Proceedings of the National Academy of Sciences of the United States of America 84: 9123-9127, 1987.

O'Leary DD. Do cortical areas emerge from a protocortex?. Trends in Neurosciences. 12(10):400-6, 1989.

O'Rourke NA. Dailey ME. Smith SJ. McConnell SK. Diverse migratory pathways in the developing cerebral cortex. Science. 258(5080):299-302, 1992.

Okabe I. Bailey LC. Attree O. Srinivasan S. Perkel JM. Laurent BC. Carlson M. Nelson DL. Nussbaum RL. Cloning of human and bovine homologs of SNF2/SWI2: a global activator of transcription in yeast S. cerevisiae. Nucleic Acids Research. 20(17):4649-55, 1992.

Olson EN. MyoD family: a paradigm for development? Genes & Development 4:1454-1461, 1990.

Padgett RA., Grabowski PJ., Konarska MM., Seiler S., Sharp P. Splicing of messenger RNA precusors. Annual Review of Biochemistry 55:1119-1150, 1986.

Palmiter R., Binster RL. Germ-line transformation of mice. Annual Review of Genetics 20:465-499, 1986.

Peckham I. Sobel S. Comer J. Jaenisch R. Barklis E. Retrovirus activation in embryonal carcinoma cells by cellular promoters. Genes & Development. 3(12B):2062-71, 1989.

Peterson CL. Herskowitz I. Characterization of the yeast SWI1, SWI2, and SWI3 genes, which encode a global activator of transcription. Cell. 68(3):573-83, 1992.

Pratt MA. Kralova J. McBurney MW. A dominant negative mutation of the alpha retinoic acid receptor gene in a retinoic acid-nonresponsive embryonal carcinoma cell. Molecular & Cellular Biology. 10(12):6445-53, 1990.

Price J. Neurobiology. Organizing the cerebrum. Nature. 362(6421):590-1, 1993.

Price J. Thurlow L. Cell lineage in the rat cerebral cortex: a study using retroviral-mediated gene transfer. Development. 104(3):473-82, 1988.

Rakic P. Guidance of neurons migrating to the fetal monkey neocortex. Brain Research. 33(2):471-6, 1971.

Rakic P. Mode of cell migration to the superficial layers of fetal monkey neocortex. Journal of Comparative Neurology. 145(1):61-83, 1972.

Rakic P. Principles of neural cell migration. Experientia. 46(9):882-91, 1990.

Rakic P. Specification of cerebral cortical areas. Science. 241(4862):170-6, 1988.

Randazzo FM. Khavari P. Crabtree G. Tamkun J. Rossant J. brg1: a putative murine homologue of the Drosophila brahma gene, a homeotic gene regulator. Developmental Biology. 161(1):229-42, 1994.

Reddy S. DeGregori JV. von Melchner H. Ruley HE. Retrovirus promoter-trap vector to induce lacZ gene fusions in mammalian cells. Journal of Virology. 65(3):1507-15, 1991.

Roach A. Takahashi N. Pravtcheva D. Ruddle F. Hood L. Chromosomal mapping of mouse myelin basic protein gene and structure and transcription of the partially deleted gene in shiverer mutant mice. Cell. 42(1):149-55, 1985.

Robbins J. Dilworth SM. Laskey RA. Dingwall C. Two interdependent basic domains in nucleoplasmin nuclear targeting sequence: identification of a class of bipartite nuclear targeting sequence. Cell. 64(3):615-23, 1991.

Robertson EJ. Pluripotential stem cell lines as a route into the mouse germ line. Trends in Genetics

2:9-13, 1986.

Robertson EJ. Teratocarcinomas and embryonic cells: a practical approach. IRL Press, 1987.

Robertson EJ., Bradley A., Kuehn M., Evans M. Germ-line transmission of genes introduced into cultured pluripotential cells by retroviral vector. Nature 323:445-448, 1986.

Robertson EJ., Evans MJ., Kaufman MH. X-chromosome instability in pluripotentential stem cell lines derived from parthenogenetic embryos. Journal of Embryology and Experimental Morphology 74:297-309, 1983.

Robinson HL. Gagnon GC. Patterns of proviral insertion and deletion in avian leukosis virus-induced lymphomas. Journal of Virology. 57(1):28-36, 1986.

Rohdewohld H. Weiher H. Reik W. Jaenisch R. Breindl M. Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. Journal of Virology. 61(2):336-43, 1987.

Rossant J. Experimental approaches to mammalian embryonic development. Rossant J., Pedersen RA., eds , pp. 97-120, Cambridge University Press, 1986.

Rossant J. Interspecific cell markers and lineage in mammals. Philosophical Transactions of the Royal Society of London - Series B: Biological Sciences. 312(1153):91-100, 1985.

Rossant J., Joyner A. Towards a molecular genetic analysis of mammalian development. Trends in Genetics 5:277-283, 1989.

Roth S. Stein D. Nusslein-Volhard C. A gradient of nuclear localization of the dorsal protein determines dorsoventral pattern in the Drosophila embryo. Cell. 59(6):1189-202, 1989.

Ruppert S. Wang EH. Tjian R. Cloning and expression of human TAFII250: a TBP-associated factor implicated in cell-cycle regulation. Nature. 362(6416):175-9, 1993.

Rushlow CA. Han K. Manley JL. Levine M. The graded distribution of the dorsal morphogen is initiated by selective nuclear transport in Drosophila. Cell. 59(6):1165-77, 1989.

Russell ES. Hereditary anemias of the mouse: a review for geneticists. Advances in Genetics. 20:357-459, 1979.

Ruvkun G. Finney M. Regulation of transcription and cell identity by POU domain proteins. Cell. 64(3):475-8, 1991.

Saiki RK., Scharf S., Faloona F., Mullis KB., Horn GT., Erlich HA., Arnheim N. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sick cell anemia. Science 230:1350-1354, 1985.

Sanes JR. Rubenstein JL. Nicolas JF. Use of a recombinant retrovirus to study post-implantation cell lineage in mouse embryos. EMBO Journal. 5(12):3133-42, 1986.

Sanes JR., Rubenstein JLR., Nicolas JF. Use of a recombinant retrovirus to study post-implantation cell lineage in mouse embryos. EMBO 5:3133-3142, 1986.

Sanger F., Nicklen S., Coulson AR. DNA sequencing with chain-termination inhibitors.

Proceedings of the National Academy of Sciences of the United States of America 74:5463-5467, 1977.

Schaeffer L. Roy R. Humbert S. Moncollin V. Vermeulen W. Hoeijmakers JH. Chambon P. Egly JM. DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. Science. 260(5104):58-63, 1993.

Schnieke A. Harbers K. Jaenisch R. Embryonic lethal mutation in mice induced by retrovirus insertion into the alpha 1(I) collagen gene. Nature. 304(5924):315-20, 1983.

Scholer HR. Balling R. Hatzopoulos AK. Suzuki N. Gruss P. Octamer binding proteins confer transcriptional activity in early mouse embryogenesis. EMBO Journal. 8(9):2551-7, 1989a.

Scholer HR. Dressler GR. Balling R. Rohdewohld H. Gruss P. Oct-4: a germline-specific transcription factor mapping to the mouse t-complex. EMBO Journal. 9(7):2185-95, 1990a.

Scholer HR. Hatzopoulos AK. Balling R. Suzuki N. Gruss P. A family of octamer-specific proteins present during mouse embryogenesis: evidence for germline-specific expression of an Oct factor. EMBO Journal. 8(9):2543-50, 1989b.

Scholer HR. Ruppert S. Suzuki N. Chowdhury K. Gruss P. New type of POU domain in germ line-specific protein Oct-4. Nature. 344(6265):435-9, 1990b.

Schreiner CM. Scott WJ Jr. Supp DM. Potter SS. Correlation of forelimb malformation asymmetries with visceral organ situs in the transgenic mouse insertional mutation, legless. Developmental Biology. 158(2):560-2, 1993.

Schwartzberg PL. Goff SP. Robertson EJ. Germ-line transmission of a c-abl mutation produced by targeted gene disruption in ES cells. Science. 246(4931):799-803, 1989.

Sekiguchi T. Nohiro Y. Nakamura Y. Hisamoto N. Nishimoto T. The human CCG1 gene, essential for progression of the G1 phase, encodes a 210-kilodalton nuclear DNA-binding protein. Molecular & Cellular Biology. 11(6):3317-25, 1991.

Serfling E., Jasin M., Schaffner W. Enhancers and eukaryotic gene transcription. Trends in Genetics 1:224-230, 1985.

Sicinski P. Geng Y. Ryder-Cook AS. Barnard EA. Darlison MG. Barnard PJ. The molecular basis of muscular dystrophy in the mdx mouse: a point mutation. Science. 244(4912):1578-80, 1989.

Silhavy TJ., Casadaban MJ., Shuman HA., Beckwith JR. Conversion of β-galactosidase to a membrane-bound state by gene fusion. Proceedings of the National Academy of Sciences of the United States of America 73:10:3423-3427, 1976.

Simeone A. Acampora D. Nigro V. Faiella A. D'Esposito M. Stornaiuolo A. Mavilio F. Boncinelli E. Differential regulation by retinoic acid of the homeobox genes of the four HOX loci in human embryonal carcinoma cells. Mechanisms of Development. 33(3):215-27, 1991.

Sinclair AH. Berta P. Palmer MS. Hawkins JR. Griffiths BL. Smith MJ. Foster JW. Frischauf AM. Lovell-Badge R. Goodfellow PN. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature.

346(6281):240-4, 1990.

Singh G. Supp DM. Schreiner C. McNeish J. Merker HJ. Copeland NG. Jenkins NA. Potter SS. Scott W. legless insertional mutation: morphological, molecular, and genetic characterization. Genes & Development. 5(12A):2245-55, 1991.

Singh P. Coe J. Hong WJ. A role for retinoblastoma protein in potentiating transcriptional activation by the glucocorticoid receptor Nature. 374(6522):562-565, 1995.

Skarnes WC, Auerbach BA., Joyner AL.A gene trap approach in mouse embryonic stem cells: the lacZ reporter is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice. Genes & Development 6:903-918, 1992.

Smith A., Heath JK., Donaldson DD., Wong GG., Moreau J., Stahl M., Rogers D. Inhibition of pluripotential embryonic stem cell differentiation by purified polypeptides. Nature 336:688-690, 1988.

Smith CW. Patton JG. Nadal-Ginard B. Alternative splicing in the control of gene expression. Annual Review of Genetics. 23:527-77, 1989.

Smith MA. O'Dowd DK. Cell-specific regulation of agrin RNA splicing in the chick ciliary ganglion. Neuron. 12(4):795-804, 1994.

Soininen R. Schoor M. Henseling U. Tepe C. Kisters-Woike B. Rossant J. Gossler A. The mouse Enhancer trap locus 1 (Etl-1): a novel mammalian gene related to Drosophila and yeast transcriptional regulator genes. Mechanisms of Development. 39(1-2):111-23, 1992.

Soriano P. Gridley T. Jaenisch R. Retroviruses and insertional mutagenesis in mice: proviral integration at the Mov 34 locus leads to early embryonic death. Genes & Development. 1(4):366-75, 1987.

Soriano P., Jaenisch R. Retrovirus as probes for mammalian development: allocation of cells to the somatic and germ cell lineages. Cell 46:19-29, 1986.

Southern PJ., Berg P. Transformation of Mammalian cells to antobiotic resistance with a bacterial gene under control of the SV40 early region promoter. Journal of Molecular and Applied Genetics 1:327-341, 1982.

Spence SE. Gilbert DJ. Swing DA. Copeland NG. Jenkins NA. Spontaneous germ line virus infection and retroviral insertional mutagenesis in eighteen transgenic Srev lines of mice. Molecular & Cellular Biology. 9(1):177-84, 1989.

Stephenson DA. Mercola M. Anderson E. Wang CY. Stiles CD. Bowen-Pope DF. Chapman VM. Platelet-derived growth factor receptor alpha-subunit gene (Pdgfra) is deleted in the mouse patch (Ph) mutation. Proceedings of the National Academy of Sciences of the United States of America. 88(1):6-10, 1991

Steward R. Relocalization of the dorsal protein from the cytoplasm to the nucleus correlates with its function. Cell. 59(6):1179-88, 1989.

Stoye JP. Coffin JM. Polymorphism of murine endogenous proviruses revealed by using virus class-specific oligonucleotide probes [published erratum appears in J Virol 1988 Jul;62(7):2530]. Journal of Virology. 62(1):168-75, 1988.

Sudol M. Expression of protoc0oncogenes in neural tissues. Brain Research Reviews 13:391-403, 1988.

Suter U. Moskow JJ. Welcher AA. Snipes GJ. Kosaras B. Sidman RL. Buchberg AM. Shooter EM. A leucine-to-proline mutation in the putative first transmembrane domain of the 22-kDa peripheral myelin protein in the trembler-J mouse. Proceedings of the National Academy of Sciences of the United States of America. 89(10):4382-6, 1992.

Suzuki N. Rohdewohld H. Neuman T. Gruss P. Scholer HR. Oct-6: a POU transcription factor expressed in embryonal stem cells and in the developing brain. EMBO Journal. 9(11):3723-32, 1990.

Tamkun JW. Deuring R. Scott MP. Kissinger M. Pattatucci AM. Kaufman TC. Kennison JA. brahma: a regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. Cell. 68(3):561-72, 1992.

Tan SS. Breen S. Radial mosaicism and tangential cell dispersion both contribute to mouse neocortical development. Nature. 362(6421):638-40, 1993.

Tan SS. Faulknerjones B. Breen SJ. Walsh M. Bertram JF. Reese BE. cell dispersion patterns in different cortical regions studied with an x-inactivated transgenic marker. Development. 121(4):1029-1039, 1995.

Tan SS., Breen S. Radial mosacixm and tangential cell dispersion both contribute to mouse neocortical development. Nature 362:638-640, 1993.

Troelstra C. van Gool A. de Wit J. Vermeulen W. Bootsma D. Hoeijmakers JH. ERCC6, a member of a subfamily of putative helicases, is involved in Cockayne's syndrome and preferential repair of active genes. Cell. 71(6):939-53, 1992.

Veres G. Gibbs RA. Scherer SE. Caskey CT. The molecular basis of the sparse fur mouse mutation. Science. 237(4813):415-7, 1987.

Vergnaud G. Page DC. Simmler MC. Brown L. Rouyer F. Noel B. Botstein D. de la Chapelle A. Weissenbach J. A deletion map of the human Y chromosome based on DNA hybridization. American Journal of Human Genetics. 38(2):109-24, 1986.

Vidal SM. Malo D. Vogan K. Skamene E. Gros P. Natural resistance to infection with intracellular parasites: isolation of a candidate for Bcg. Cell. 73(3):469-85, 1993.

Vincent S. Marty L. Fort P. S26 ribosomal protein RNA: an invariant control for gene regulation experiments in eucaryotic cells and tissues. Nucleic Acids Research. 21(6):1498, 1993.

von Melchner H. DeGregori JV. Rayburn H. Reddy S. Friedel C. Ruley HE. Selective disruption of genes expressed in totipotent embryonal stem cells. Genes & Development. 6(6):919-27, 1992.

von Melchner H. Reddy S. Ruley HE. Isolation of cellular promoters by using a retrovirus promoter trap. Proceedings of the National Academy of Sciences of the United States of America. 87(10):3733-7, 1990.

von Melchner H. Ruley HE. Identification of cellular promoters by using a retrovirus promoter trap. Journal of Virology. 63(8):3227-33, 1989.

von Melchner H. Ruley HE. Retroviruses as genetic tools to isolate transcriptionally active chromosomal regions. Environmental Health Perspectives. 88:141-8, 1990.

Walsh C. Cepko CL. Clonal dispersion in proliferative layers of developing cerebral cortex. Nature. 362(6421):632-5, 1993.

Walsh C. Cepko CL. Clonally related cortical cells show several migration patterns. Science. 241(4871):1342-5, 1988.

Walsh C. Cepko CL. Widespread dispersion of neuronal clones across functional regions of the cerebral cortex. Science. 255(5043):434-40, 1992.

Walsh C., Cepko C. Clonal dispersion in proliferative layers of developing cerebral cortex. Nature 362:632-635, 1993.

Walther C. Gruss P. Pax-6, a murine paired box gene, is expressed in the developing CNS. Development. 113(4):1435-49, 1991.

Walther C. Guenet JL. Simon D. Deutsch U. Jostes B. Goulding MD. Plachov D. Balling R. Gruss P. Pax: a murine multigene family of paired box-containing genes. Genomics. 11(2):424-34, 1991.

Wang EH. Tjian R. Promoter-selective transcriptional defect in cell cycle mutant ts13 rescued by hTAFII250. Science. 263(5148):811-4, 1994.

Wegner M. Drolet DW. Rosenfeld MG. POU-domain proteins: structure and function of developmental regulators. Current Opinion in Cell Biology. 5(3):488-98, 1993.

Weiher H. Noda T. Gray DA. Sharpe AH. Jaenisch R. Transgenic mouse model of kidney disease: insertional inactivation of ubiquitously expressed gene leads to nephrotic syndrome. Cell. 62(3):425-34, 1990.

Weintraub H., Davis R., Tapscott S., Thayer M., Krause M., Benezra R., Blackwell TK., Turner D., Rupp R., Hollenberg S., Zhuang Y., Lassar A. The MyoD gene family: nodal point during specification of the muscle cell lineage. Science 251:761-766, 1991.

White TJ., Arnheim N., Erlich HA. The polymerase chain reaction. Trends in Genetics 5:185-189, 1989.

Wilkinson DG. Bhatt S. Cook M. Boncinelli E. Krumlauf R. Segmental expression of Hox-2 homoeobox-containing genes in the developing mouse hindbrain. Nature. 341(6241):405-9, 1989.

Wilkinson KD., Bhatt S., Cook M., Boncinelli E., Krumlauf R. Segmental expression of Hox-2 homeobox-containing genes in the developing mouse hind brain. Nature 341:405-464, 1989.

Williams DE. Eisenman J. Baird A. Rauch C. Van Ness K. March CJ. Park LS. Martin U. Mochizuki DY. Boswell HS. et al. Identification of a ligand for the c-kit proto-oncogene. Cell. 63(1):167-74, 1990.

Williams RL., Hilton DJ., Pease S., Willson TA., Stewart CL., Gearing DP., Wagner EF., Metcalf D., Nicola NA., Gough NM Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. Nature 336:684-687, 1988.

Wood SA. Allen ND. Rossant J. Auerbach A. Nagy A. Non-injection methods for the production of embryonic stem cell-embryo chimaeras. Nature. 365(6441):87-9, 1993.

Woychik RP. Maas RL. Zeller R. Vogt TF. Leder P. 'Formins': proteins deduced from the alternative transcripts of the limb deformity gene. Nature. 346(6287):850-3, 1990.

Woychik RP., Stewart TA., Davis LG., D'Eustachio P, Leder P. An inherited limb deformity created by insertional mutagenesis in a transgenic mouse. Nature 318:36-40, 1985.

Wu H. Fassler R. Schnieke A. Barker D. Lee KH. Chapman V. Francke U. Jaenisch R. An X-linked human collagen transgene escapes X inactivation in a subset of cells. Development. 116(3):687-95, 1992.

Yamamori T., Fukada K., Aebersold R., Korsching S., Fann MJ., Patterson PH. The cholinergic neuronal differentiation factor from heart cells is identical to leukemia inhibitory factor. Science 246:1412-141 6, 1989.

Yarden Y. Kuang WJ. Yang-Feng T. Coussens L. Munemitsu S. Dull TJ. Chen E. Schlessinger J. Francke U. Ullrich A. Human proto-oncogene c-kit: a new cell surface receptor tyrosine kinase for an unidentified ligand. EMBO Journal. 6(11):3341-51, 1987.

Yokoyama T. Copeland NG. Jenkins NA. Montgomery CA. Elder FF. Overbeek PA. Reversal of left-right asymmetry: a situs inversus mutation. Science. 260(5108):679-82, 1993.

Yoshida H. Hayashi S. Kunisada T. Ogawa M. Nishikawa S. Okamura H. Sudo T. Shultz LD. Nishikawa S. The murine mutation osteopetrosis is in the coding region of the macrophage colony stimulating factor gene. Nature. 345(6274):442-4, 1990.

Yoshinaga SK. Peterson CL. Herskowitz I. Yamamoto KR. Roles of SWI1, SWI2, and SWI3 proteins for transcriptional enhancement by steroid receptors. Science. 258(5088):1598-604, 1992.

Zdarsky E. Favor J. Jackson IJ. The molecular basis of brown, an old mouse mutation, and of an induced revertant to wild type. Genetics. 126(2):443-9, 1990.

Zhou X. Sasaki H. Lowe L. Hogan BL. Kuehn MR. Nodal is a novel TGF-beta-like gene expressed in the mouse node during gastrulation. Nature. 361(6412):543-7, 1993.