CORDON-BLEU, A NOVEL MURINE DEVELOPMENTAL MARKER GENE ISOLATED BY A GENE TRAP APPROACH

by

Stéphan Emmanuel Gasca

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

CORDON-BLEU, A NOVEL MURINE DEVELOPMENTAL MARKER GENE ISOLATED BY A GENE TRAP APPROACH

by Stéphan Emmanuel Gasca

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy in the Graduate Department of Molecular and Medical Genetics at the University of Toronto. 1997

ABSTRACT

Gene trap (GT) vectors which can act as both a reporter and a mutagen have been previously used to isolate new genes that are essential for mouse development. I used gene trap GT vectors and embryonic stem (ES) cell chimeras to screen for insertions of the *lacZ* reporter gene into transcription units that are spatially and temporally regulated during early mouse embryogenesis. 308 $G418^R$ ES cell clones expressing the *lacZ* reporter gene were individually injected into blastocysts to screen for expression *in vivo*, in ES cell/embryo chimeras. The expression of the reporter gene in chimeric embryos was analyzed at embryonic day 8.5 (E8.5).

The characterization of one GT insertion permitted the cloning of a novel gene named *cordon-bleu* (*cobl*) because of its expression in E8.5 embryos. In embryos carrying the *GtC101* insertion, ß-Gal activity was first detected at E7.5 in the node region and at E8.5, expression was restricted to axial structures, namely gut endoderm, notochord and the floor plate of the neuro-epithelium. *Cobl*'s open reading frame shows no homology to any known protein or protein domain. No phenotype was detected in mice homozygous for the insertion and mapping of the mouse and human syntenic loci did not reveal any mutation for which *cobl* could be a candidate gene. The insertion may not have completely disrupted *cobl* function and *cobl's* potential role in vertebrate axis formation or patterning remains a conundrum.

Finally, I have used the viable *GtC101 lacZ* insertion into *cobl* as a molecular marker and studied ß-Gal expression in *Brachyury* (*T*) and *Danforth's short-tail* (*Sd*) mutants. Mutations at both loci are semi-dominant and affect the normal development of the notochord. I generated mouse lines carrying the *GtC101* marker and one of the *T*, T^{Wis} or *Sd* alleles. For the three alleles, *cobl* expression in axial mesoderm revealed the earliest defects to date in heterozygous and homozygous mutant embryos. Other domains of *cobl* expression also provide evidence for patterning defects in the somites and the tail bud resulting from abnormal organogenesis of the notochord.

ACKNOWLEDGMENTS

I would like to thank Pr. François Jacob, Pr. Bernard Dujon and Dr. Jean-François Nicolas who, indirectly, helped me getting the best posting I could imagine for my Military Service: at the Mount Sinai Hospital's Samuel Lunenfeld Research Institute in Toronto.

I would also like to thank Dr. Janet Rossant for welcoming the accomplishment of this duty in her laboratory and for inviting me to continue my Ph.D. studies under her guidance, in the excellent Department of Molecular and Medical Genetics at the University of Toronto.

I thank all past, present and everlasting members of the "Rossant Lab", for their Friendship, their help and suggestions at the bench and their useful comments on my "anglish" linguistic experiments.

I must also acknowledge the following people or organization who, at different times, shared their experience and material or gave financial support. I thank: Dr. William Dove for sending us mice carrying the T^{wts} allele; Kendrapasad Harpal for his help with histology; C.-c. Hui who shared his expertise with RNA in situ hybridization; Steve Scherer for the work on the mapping of the human *cobl* gene; Bill Skarnes for providing the pGT4.5a vector and primer #256; The Jackson Laboratory for providing the BSS interspecific backcross DNA panel.

This work was supported in part by a Government of Canada Award scholarship, a University of Toronto Open scholarship and NIH grant # HD-25334.

ATTRIBUTION OF THE DATA

The Chapter 2 contains the results from a pilot screen that helped us choosing the screening strategy followed in a second larger gene trap screen. The work presented in this pilot screen was achieved in collaboration with Dr. David Hill: D. Hill built the new PT-1 and PT-3 gene trap vectors and pre-screened *in vitro* half of the clones derived from the pGT4.5a and PT-1 vectors and all the clones generated with the PT-3 vector (Table 4). We also worked together for the dissection and analysis of chimeric embryos (Table 6). I did most of the injections of ES cell clones into blastocysts with the help of Janet Rossant, who injected a third of the clones tested *in vivo* (Table 5). During the pilot screen, I performed routinely on my own each step involved in the gene trap expression screen strategy presented here, from the electroporation of GT vectors into ES cells to the preparation of histological sections from chimeric embryos.

In the second half of this chapter, I added results from the large screen that we undertook after the pilot screen. Because it was a collaborative project between the two laboratories of Dr. Janet Rossant and Dr. Alexandra Joyner, I tried to present these published results (Wurst *et al.*, 1995) in a condensed form. My principal contributions to the large screen were to inject ES cell clones into blastocysts (I processed about a quarter of the clones analyzed during that screen) and to analyze and discuss the data before publication.

I am responsible for all the work involved in the characterization of the mouse *cordon-bleu* gene (Chapter 3) with the following exceptions: the section shown in Figure 9A and Figure 10 was prepared by Ken Harpal; the whole mounts *in situ* hybridizations in Figure 13 were prepared by Dr. John Klingensmith. Céline Champigny did some of the sub-cloning and sequencing reactions for clones covering the 5' and 3' ends of the *cordon-bleu* cDNA contig. The mapping of the human *cobl* gene was done by Dr. Steve Scherer in Dr. Lap Chee Tsui's laboratory at the Hospital for Sick Children of Toronto.

I did all the work presented in the last data chapter (Chapter 4).

ABBREVIATIONS

a.a.	Amino acids
AER	Apical ectodermal ridge
b p	Base pair
bHLH	Basic helix-loop-helix
ß-Gal	ß-Galactosidase enzyme
ß-gal	ß-galactosidase gene
ßgeo	Fusion of the *ß-galactosidase* and *neomycin* genes
ßgyg	Fusion of the *ß-galactosidase* gene with the *hygromycin* resistance gene
cDNA	Complementary DNA
cM	centiMorgans
CNH	Chordoneural hinge
CNS	Central nervous system
cobl	*cordon-bleu* gene
E. coli	*Escherichia coli*
E8.5 or 8.5 dpc	Day 8.5 post-coitum
En-2	*Engrailed-2* gene
ENU	Ethyl-nitrosourea
ES	Embryonic stem
ET	Enhancer trap
$G418^R$	Resistance to the G418 drug
GFP	Green fluorescent protein
GT	Gene trap
GtC101	Locus name for the insertion of pGT4.5a into the *cordon-bleu* gene
Hprt	*Hypoxanthine phosphoribosyltransferase* gene
IRES	Internal ribosomal entry site
k b	Kilo bases
lacZ	*ß-galactosidase* gene
neo	*Neomycin* gene
nls	Nuclear localization signal
nts	nucleotides
ORF	Open reading frame
pAT	Poly-A trap
PBS	Phosphate buffered saline
Pgk-1	*Phosphoglycerate kinase 1* gene
poly-A	Poly-adenylation signal
PT	Promoter trap
RACE-PCR	Rapid amplification of cDNA ends coupled to polymerase chain reaction
RFLPs	Restriction length polymorphisms
Sd	*Danforth's short tail* mutant allele
Shh	*Sonic hedgehog* gene
SSLPs	Simple sequence length polymorphisms
SV-40	Simian-virus 40
T	*Brachyury* large deletion allele
Tau	Signal peptide translocating proteins along axons
TK	*Thymidine kinase* gene
TM	Trans-membrane
T^{ins}	*Brachyury* T^{ins} insertion mutant allele
X-gal	5-bromo-4-chloro-3-indolyl-ß-D-galacto-pyranoside
ZPA	Zone of polarizing activity

TABLE OF CONTENTS

LIST OF TABLES

INTRODUCTION

CHAPTER 2

CHAPTER 3

CHAPTER 4

LIST OF FIGURES

INTRODUCTION

CHAPTER 2

CHAPTER 3

CHAPTER 4

DISCUSSION

- CHAPTER I -

INTRODUCTION

CHAPTER ONE : INTRODUCTION

APPROACHES TO DEVELOPMENTAL BIOLOGY

Genetics is a powerful approach to identify and analyze the function of the molecules controlling vertebrate development. In prokaryotes and unicellular eukaryotes, the identification of mutant variants and their genetic analysis has been extensively used to study the function of proteins and the regulation of gene expression. Molecular biology techniques subsequently allowed isolation of genes and identification of the molecular basis of the mutations. Molecular techniques have also been used to manipulate genes to engineer specific alterations whose effects can be tested *in vivo* by inverse genetics and *in vitro* by biochemical approaches. Similarly genetics has been used in metazoan model systems to identify and study developmental genes. The combination of genetic mutation screens with molecular characterization of novel genes has allowed both functional and structural analysis of proteins involved in developmental processes.

The laboratory mouse, *Mus musculus*, is undeniably a good experimental and comparative model system to study vertebrate and mammalian development. However, vertebrates, with the exception of the zebrafish *Brachydanio rerio*, are not readily amenable for large scale genetic mutation screens. As a result, most of our early knowledge on vertebrate development has come from classical experimental embryology and from indirect genetic approaches used to analyze molecules first identified in assays performed *in vitro*, or taking advantage of the structural and functional analogies with invertebrate model systems. However, direct and independent approaches allowing mutagenesis and cloning of developmental genes are needed to further investigations on vertebrate embryogenesis.

Classical embryology

Experimental embryologists generally investigate the consequences of mechanical alterations to the normal development of an embryo. These experiments involve single cells or tissues and generally consist of ablating, grafting or marking cells in embryos which are then allowed to develop further *in vivo* or *in vitro*. In vertebrates and invertebrates, experimental embryology thus uncovered many basic developmental processes. These include the appearance of

restricted cell lineages early during development: the existence of specific cell fate and extensive cell migration patterns occurring within the different embryonic and extra embryonic lineages: the observation that at specific spatial and temporal checkpoints, cells could become committed to a determined fate: the potential for some tissues to act as inducer or organizer centres managing the differentiation of neighbor cells. All these features result in the progressive patterning of the embryo into regions with restricted developmental potential (reviewed by Slack, 1991). They also pointed out that the most critical stages of development were occurring during the transformation of the egg into a patterned embryo. A few dexterous embryologists were able to apply these experimental procedures to the mouse where the same general properties were observed (Lawson *et al.*, 1986; Lawson *et al.*, 1991; Tam and Tan, 1992; Beddington, 1994). However, manipulations on the mouse embryo are limited by its small size and its development in the uterus. Furthermore, experimental embryology is limited to the study of cellular interactions and cannot give access to the underlying molecular mechanisms controlling the early patterning of the embryo.

Developmental genetics

Induced mutagenesis, first and foremost in the fruit fly *Drosophila melanogaster* (*D. melanogaster*), allowed isolation of mutations in genes that could control the fate of large regions or complete segments of the embryo. These original mutations provided candidate genes for the components of the genetic pathways regulating regional specification of the embryo. Subsequently, systematic mutation screens were initiated to isolate and characterize more developmental genes involved in the control of embryo patterning. The first were the large genetic screens for recessive embryonic lethal mutations performed in the invertebrates *D. melanogaster* and *Caenorhabditis elegans* (*C. elegans*) (Nusslein-Volhard and Wieschaus, 1980; Meneely and Herman, 1979; Schierenberg *et al.*, 1980). These screens were designed to detect mutations in pre-determined genomic regions, which affect specific processes during embryogenesis. They enabled researchers to identify novel genes that are essential for embryonic patterning. In *D. melanogaster*, most of the genes important for early anterior-posterior axial patterning of the embryo have thus been identified (Nusslein-Volhard *et al.*, 1987). Many different mutant alleles of the same gene were often generated, demonstrating that saturation had been reached for this particular strategy. Consequently, it was possible to design comprehensive models of the

genetic pathways controlling the mechanisms of early embryo patterning. These examples demonstrated that it was conceivable to decipher the genetic pathways regulating early development of the embryo when genes were systematically mutagenized and identified.

Vertebrate model systems such as the chick and the frog are not suitable for direct genetic experimentation: the chick is a slow breeder and the frog's genome is tetraploid. These features are incompatible with the realization of mutagenic screens. The mouse and the zebrafish are better systems for the genetic study of vertebrate development. The zebrafish is a particularly suitable organism for mutation screens: its genome is smaller ($1.7\ 10^9$ bp) and embryos are produced in large numbers. Early development is very fast and embryos can be directly observed under a dissecting microscope during the course of their genesis: furthermore, induced parthenogenic embryos develop normally during early embryogenesis, thus facilitating the identification of recessive mutations (Grunwald and Streisinger, 1992; Solnica-Krezel *et al.*, 1994). Systematic genetic screens to identify mutations altering embryogenesis are underway in the zebrafish (Mullins *et al.*, 1994; Haffter *et al.*, 1996; Driever *et al.*, 1996). However, it is also a recent model system for which good genetic or physical map are not available yet whereas such maps are well characterized in the mouse (reviewed by Rossant and Hopkins, 1992). Furthermore, zebrafish is not a complete model for mammalian development (e.g. if one is interested the role of maternal and extraembryonic tissues during development or in later aspects of embryogenesis such as the organogenesis of the limbs and the brain). In the mouse model system, embryonic stem (ES) cell technology and novel strategies combining both genetic and molecular approaches now permit direct disruption and identification of developmental genes. Therefore, for the analysis of mammalian development, the mouse is well known and very versatile vertebrate model to perform genetic studies.

THE MOUSE: A GENETIC MODEL FOR VERTEBRATE DEVELOPMENT

The laboratory mouse or house mouse is the best model system for the genetic analysis of vertebrate development for both historical and practical reasons. For centuries, mouse fanciers have collected spontaneous mouse variants carrying mutations that mostly affected coat color or behavior (Silvers, 1979). The history of the mouse as a genetic model system started in the 1900's with the collection of

animal strains and analysis of these naturally occurring mutations by research laboratories (Haldane *et al.*, 1915). The mouse was the first mammal used to verify the rediscovered Mendel's laws of inheritance (Cuénot, 1902; Castle and Allen, 1903). Since then, additional spontaneous variants have been accumulated within the mouse stocks of large animal colonies (Green, 1989). Wild-type mouse strains were also established and inbred in laboratories. They provide a variety of homogenous genetic backgrounds that are used as a reference to study mutations and DNA polymorphisms. Mutant alleles and DNA polymorphisms between strains have been used to generate a genetic map of the mouse genome (Bonhomme *et al.*, 1979; Leder *et al.*, 1981; Avner *et al.*, 1987). Mutations still represented the majority of the 1300 genetic loci reported by Green (1989). More recently, molecular biology techniques allowed the identification of a plethora of novel genes and DNA polymorphism markers such as restriction fragment length polymorphisms (RFLPs) and simple sequence length polymorphisms (SSLPs). Interspecific crosses between species such as *Mus musculus* and *Mus spretus* permit a rapid mapping of these new markers (Rowe *et al.*, 1994). As a result, the mouse genetic map now contains about 14,000 loci and the construction of a physical map is well underway (Dietrich *et al.*, 1995). Therefore a good scaffold of markers already exists to help further genetic characterization of mouse developmental genes.

Practically, mice are easy to breed and have a relatively short generation time of 2 to 3 months. A large array of embryological and genetic manipulations is achievable in mice. In particular, the recent development of two techniques, the mouse ES cells technology and the insertion of exogenous DNA by homologous recombination in mammalian cells notably enlarged the possibilities for genetic studies in the mouse. ES cells can be cultured (Evans and Kaufman, 1981; Martin, 1981) and genetically manipulated *in vitro* while retaining their potential to participate in normal embryonic development (Bradley *et al.*, 1984; Gossler *et al.*, 1986) or to support development of the entire embryo (Nagy *et al.*, 1993). Homologous recombination allows one to specifically target and introduce mutations into genetic loci (Capecchi, 1989; Koller and Smithies, 1992). Therefore, virtually any mutation, generated in ES cells that are able to colonize the mouse germline, can be analyzed *in vivo* (Capecchi, 1989). However, "inverse genetic by targeted mutagenesis" (a gene is cloned first, altered *in vitro* by homologous recombination and re-introduced *in vivo* to analyze the phenotype caused by the mutation) is restricted to genes already cloned. The advantage of "forward genetics" (a random mutation spontaneous or induced *in vivo* leads to the cloning

of a candidate gene) is to be a more random approach, based on the observation of a phenotype indicating that a gene's function has been altered. Several approaches to random mutagenesis are possible in the mouse. Mutations can be spontaneous, induced by X-ray or chemicals, or generated by insertional mutagenesis.

Random mutagenesis in the mouse

Spontaneous variants

The majority of spontaneous variants are dominant or semi-dominant mutations affecting obvious traits of morphology (e.g. coat color, visible skeletal structures) or behavior, life span and fertility. These mutations have been very important for the establishment of the mouse as a genetic model system and they still supply an important resource of mutants. However, they do not provide a sufficient source of genetic material for developmental studies. The fortuitous mode of identification of mutant characters results in a strong bias toward dominant mutations and consequently, very few recessive embryonic lethal mutations affecting early developmental genes were isolated. Furthermore, the molecular characterization of these mutants requires difficult cloning strategies and has only been possible for a few of them (Gubbay *et al.*, 1990; Herrmann *et al.*, 1990; Bultman *et al.*, 1992; Michaud *et al.*, 1993; Cordes and Barsh, 1994; Avraham *et al.*, 1995; Schumacher *et al.*, 1996). The first of these mutants to be characterized were identified by the candidate gene approach (Balling *et al.*, 1988; Geissler *et al.*, 1988; Epstein *et al.*, 1991; Hill *et al.*, 1991; Hui and Joyner, 1993). The strongest drawback however, comes from the scarcity of these mutation events which is incompatible with any mutation screening strategy. X-ray and chemical mutagenesis were consequently used to induce mutations at much higher rates.

Radiation or chemical induced mutations

X-ray radiation was initially used to mutagenize the mouse genome and generated new mutations such as the *T* (*Brachyury*) allele (Dobrovolskaïa-Zavadskaïa, 1927). But in the absence of a systematic screen for recessive mutations, the bias toward dominant mutations already observed with spontaneous mutants was persistent in this approach. Radiation also creates large deletions and/or rearrangements at the mutated locus that may alter more than one gene at a time. Chemical mutagenic agents such as *N-ethyl-N-nitrosourea* (ENU) (Russell *et al.*, 1979) have also been used to mutagenize the mouse germline. The advantage of

these agents with respect to radiation is that they only create point mutations or small deletions affecting one gene at a time. Radiation and chemicals are very efficient mutagenic agents because they generate mutations in a random fashion and at a high rate: ENU can induce an average of 3 to 15 10^{-4} mutations/locus/gamete (Russell *et al.*, 1979; Pretsch and Charles, 1984; Hitotsumachi *et al.*, 1985), up to 12 times the frequency obtained with X-rays (Hitotsumachi *et al.*, 1985) and 300 times the spontaneous rate (Russell *et al.*, 1981). These two mutagens as well as other chemical agents have been use to perform mutation screens in the mouse (reviewed by Rinchik, 1991).

Mutation screens are difficult in the mouse because of both generation time, space requirements, small litter size and intra-uterine development of the embryos. Saturation mutation screens such as the ones performed in invertebrates have only been possible in the mouse when applied to very restricted and well characterized regions of the genome where known genetic markers allowed an easier screening for offspring carrying recessive lethal mutations. Two chemical mutagenesis screens were performed to select for mutations in loci linked to the *t* region on chromosome 17 (Shedlovsky *et al.*, 1988) and to the *albino* (*c*) locus on chromosome 7 (Rinchik *et al.*, 1990). The latter screening strategy also relied on a collection of deletions at the *c* locus previously induced by radiation mutagenesis (Russell *et al.*, 1982). Another screen detected deletions induced by radiation at the *brown tryp-1* locus on chromosome 4 (Rinchik *et al.*, 1994).

There are still drawbacks to these approaches. They require large breeding facilities for the isolation of the mutant strains. In addition, the identification of candidate genes for these newly generated mutants still requires tedious cloning strategies. Because of the limitations of targeted and classical mutagenesis, approaches allowing large scale mutation screens and facilitating the cloning of novel genes were also considered. Mutagenesis techniques using random insertion of exogenous DNA sequences clearly offer strong advantages to further characterize altered genes.

Insertional mutagenesis

Mutagenesis by insertion of endogenous (transposons) or exogenous (viruses) transposable genetic elements is a natural occurrence and is an important source of genetic diversity for many life forms: the random insertion of a DNA sequence in the genome may alter the structure or function of genes located in the vicinity of the insertion locus. Transposable elements were first identified in maize by B.

McClintock (McClintock, 1950). Since then, transposable elements have been identified in bacteria, yeast, invertebrates and vertebrates. Even the wrinkled-seed phenotype of the *rugosus* allele described by Mendel in 1865 has been shown to be caused by a transposon insertion (Bhattacharyya *et al.*, 1990). The great advantage of insertional mutagenesis is that when the inserted DNA sequence is known, it can serve as a tag to identify adjacent genomic region and thus greatly facilitates the cloning of the insertion locus.

In the mouse, retroviral transposons have been used to generate mutations into the genome. Two methods, the transposition of endogenous proviruses and insertion of naturally occurring ecotropic retroviruses, have been successfully used to create recessive lethal mutations at new loci (Soriano *et al.*, 1987). Alternately, exogenous DNA can be introduced into the cells carried by a recombinant retrovirus (Rubenstein *et al.*, 1984; Huszar *et al.*, 1985; Soriano *et al.*, 1986; Sanes *et al.*, 1986) or by a plasmid vector (Breitman *et al.*, 1987; Goring *et al.*, 1987; Jaenisch, 1988). Retroviruses can directly infect embryos or cultured cells, whereas plasmid vectors need to be mechanically forced into cells. Micro-injection into eggs or transfection and electroporation of cells are methods currently used to introduce plasmid DNA into the mouse genome. The majority of exogenous DNA sequences insertion events into the genome occurs at random. Exogenous DNA sequences and retroviruses can both act as mutagenic agents, either by the insertion resulting in the disruption of endogenous gene function or by the retroviral promoter activating ectopic expression of neighbor genes. Because transgenic mice are the most current method to analyze transcription regulation *in vivo*, many transgenic insertions, sometimes causing a phenotype, have been generated and thus also became the subjects of mutations' analyses.

In 1991, T. Gridley evaluated the progress made in the characterization of insertional mutations generated by retroviral insertion or DNA micro-injection and resulting in a visible phenotype (Gridley, 1991). Of 28 insertional mutations reported, 11 were recessive embryonic lethal, 3 post-natal lethal and 14 viable mutations. Eleven insertions occurred in known genes (*Hprt*, *Steel* and *alpha-I-collagen*) or in loci for which a mutant allele already existed. The exogenous DNA could be used as a tag to clone the corresponding genomic loci and identify candidate genes altered by these insertions. Five novel disrupted genes (*limb deformity*, *Mov-34*, *Mpv-17*, *dilute* and *Hβ58*) were already cloned in 1991. Since then, 3 more genes, *hairless* (Cachon-Gonzalez *et al.*, 1994), *dystonia musculorum* (Brown *et al.*, 1994) and the *413.d* retroviral insertion into *nodal* (Zhou *et al.*, 1993)

have been cloned from the direct analysis of the insertion site. The *Extra-toes* locus was cloned by other means (Hui and Joyner, 1993). The phenotype of the line *myk-103* was due to ectopic expression of the *thymidine kinase* (TK) transgene rather than to the disruption of a transcription unit (Wilkie *et al.*, 1991). For the other insertions, no particular transcript could be identified, although candidate cDNAs were analyzed for 2 loci (*fused* and *legless*).

These results permit comparison between proviral and transgenic insertion. Six out of 7 mutations resulting from insertion of a provirus led to the molecular characterization of the endogenous loci. When proviruses integrate into the genome they only generate minor rearrangements: the disrupted genomic sequences are directly flanking proviral DNA and can easily be cloned. On the contrary, the microinjection technique often generates major chromosomal rearrangements upon insertion of exogenous DNA: for example, the *RSV-CAT strain2* contains a translocation at the insertion locus (Mahon *et al.*, 1988); deletions of approximately 20 kb, 50 kb and 2 cM (about 1000 kb) have been reported (Constantini *et al.*, 1989; Brown *et al.*, 1994). These rearrangements cause two problems: they may well be the primary cause for some of the phenotypes observed (rather than the insertion event itself) and they complicate the identification of the disrupted gene.

Overall, for the two approaches, the frequency of insertions causing a recessive lethal phenotype is between 5% and 10% (Soriano *et al.*, 1987; Jaenisch, 1988). Since transgenic strains are continuously being generated in many laboratories, novel mutants also continue to appear. Thus, the transgenic insertion A4 into the *motor endplate disease* (*med*) locus (Kohrman *et al.*, 1995), rapidly led to the cloning of a novel gene, *snc8a*, altered in *medTgNA4Bs* and *med* mutants (Burgess *et al.*, 1995a).

However, no screening strategy is available for these approaches. The identification of transgenic insertions with a recessive phenotype is generally fortuitous. Establishing and breeding heterozygous strains to screen large number of insertions would require a lot of time and space. The tagging by micro-injected DNA is not necessarily accurate or informative enough to rapidly clone a candidate gene.

Other particularities fortuitously observed upon insertion of exogenous DNA offered the possibility to design novel strategies and to improve the efficiency of mutagenesis by insertion. Early, in the making of transgenic mice, it was observed that the activity of exogenous promoters could be influenced by endogenous

sequences at the site of insertion. It was suggested then that transgenic mice could be used for the detection of endogenous regulatory sequences (Soriano *et al.*, 1986; Jaenisch and Soriano, 1986; Jaenisch, 1988). These regulatory elements were shown to be the enhancer sequences of endogenous transcription units (Hamada, 1986). It was therefore conceivable to design insertion vectors containing a reporter gene that would detect genomic regulatory elements at the site of insertion. The reporter gene expression would then reproduce the pattern of expression of endogenous genetic loci.

New types of insertion vectors, the trapping vectors, were thus designed to identify developmentally regulated genes and facilitate their molecular and functional characterization. These vectors conserved the advantages of mutagenesis by DNA insertion (i.e. random integration, tagging of the mutant allele and creation of novel genetic markers). More importantly, they also permitted the detection of genes by their expression pattern, regardless of the phenotype created by their insertion.

THE TRAPPING VECTORS

General features of trapping vectors

All trapping vectors depend on the detection of genomic cis-regulatory elements that regulate expression of a reporter gene after its insertion into a host genome (see Fig. 1 for the different modes of activation). In addition, the vectors always contain a selection marker allowing the selection of stable insertion events. Trapping vectors are "activated" when expression of their selection and/or reporter genes is induced after insertion by endogenous sequences. The reporter gene is generally the E. coli *β-galactosidase* (*lacZ*) gene which encodes the readily detectable ß-Galactosidase (ß-Gal) enzyme. Active ß-Gal can be produced in most organisms without altering normal cell functions. Simply, trapping vectors can be further classified in three categories, according to the mode of activation of the reporter gene: they are the promoter trap (PT), the enhancer trap (ET) and the gene trap (GT) vectors; in addition, another type of gene trapping vector, the polyA trap (pAT), that does not initially require the activation of a reporter gene has also been designed (Fig. 1).

Promoter trap vectors

The first trapping vectors were the promoter trap vectors used in prokaryotes

(Casadaban and Cohen, 1979). A promoterless reporter gene became activated when it inserted within an operon sequence. Promoter trap vectors were also designed for the mouse (Kerr *et al.*, 1989; von Melchner *et al.*, 1990; Reddy *et al.*, 1991). The vector is activated when it inserted downstream of a promoter able to drive expression of the reporter gene. Usually, such insertion must have occurred into a coding exon of a gene (Fig. 1 A).

Enhancer trap vectors

In enhancer trap vectors, the *lacZ* reporter gene expression is driven by a minimum promoter which is activated when endogenous regulatory sequences are present in cis. Insertions occurring within genes or intergenic regions may confer on the reporter gene the specificity of expression of endogenous genes (Fig. 1B). The first eukaryotic enhancer trap vectors were designed for *D. melanogaster* (O'Kane and Gehring, 1987). Enhancer trap vectors were also designed for the mouse (Allen *et al.*, 1988; Gossler *et al.*, 1989) and could efficiently detect regulatory elements.

Gene trap vectors

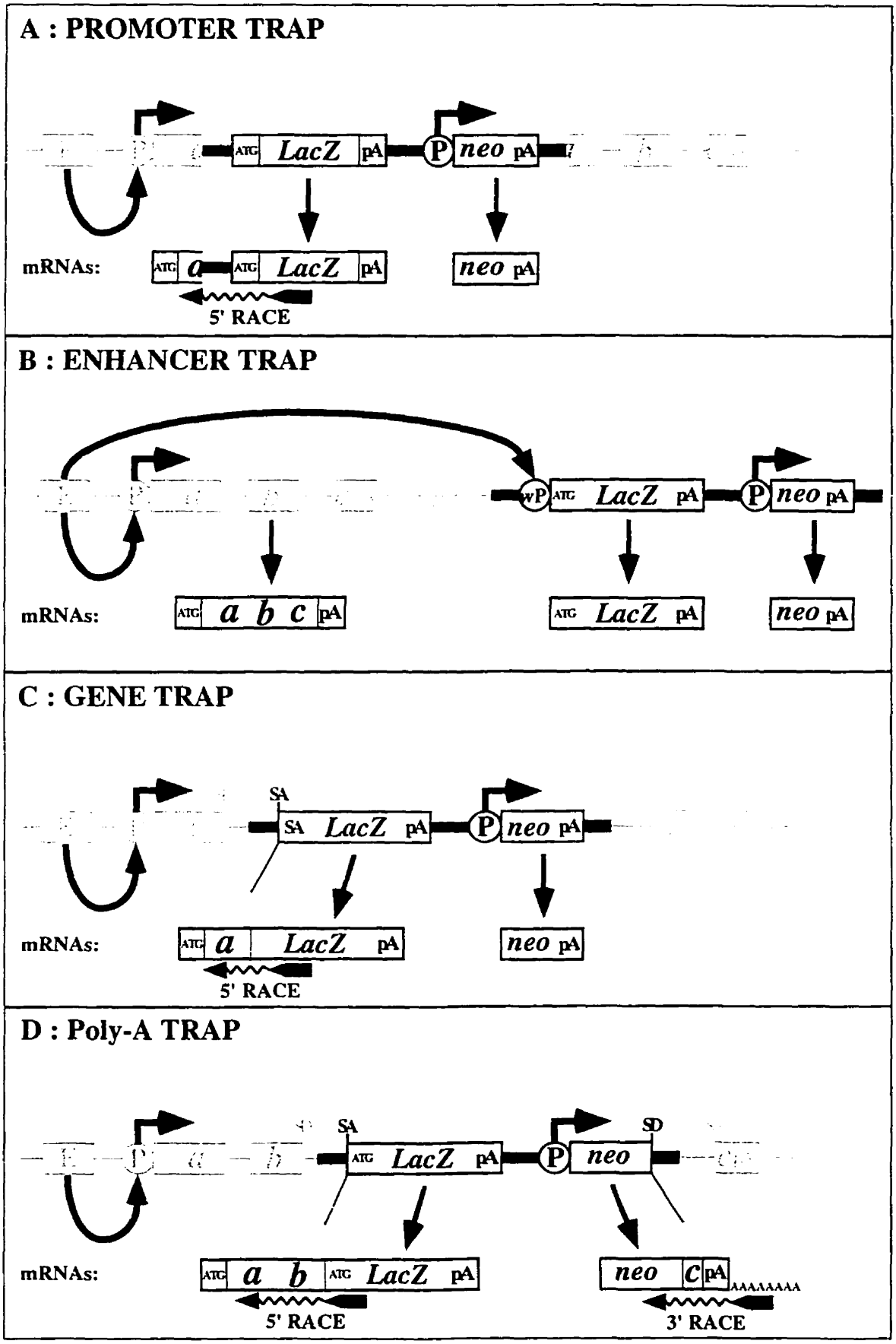
The gene trap vectors were designed to select for insertion events occurring within transcription units and generating a fusion transcript between the endogenous gene and the reporter gene (Fig. 1C). It had been shown that splicing between donor (S.D.) and acceptor (S.A.) sites from different genes or synthetically derived could be used by mammalian cells to form functional chimeric transcription units (Chu and Sharp, 1981). In gene trap vectors, the reporter gene contains a S.A. site at its 5' end and thus when insertion occurred into a gene's intron, it behaves like an artificial exon (Gossler *et al.*, 1989; Brenner *et al.*, 1989; Friedrich and Soriano, 1991).

Poly A trap vectors

Most trapping vectors rely on insertion into genes active in undifferentiated or differentiated ES cells that are able to activate the expression of the reporter gene, the selection gene or both. The poly-A trap (p-AT, poly-A is short for poly-adenylation signal) vectors were designed to be activated when inserting into any transcription unit whether the latter is active or inactive at the time of

Fig. 1. - Modes of activation for different types of trapping vectors.

In A. B. C & D. the trapping vectors (in black) are activated after insertion in specific regions of a locus (in gray) encoding the *abc* gene. (A) In the promoter trap, the *ß-galactosidase* (*lacZ*) reporter gene is only transcribed when it inserted into an exon (here exon a). (B) The enhancer trap's reporter gene is activated when the endogenous enhancer (E) activates the weak promoter (wP) driving its expression. (C) When it inserts into an intron, the gene trap vector's *lacZ* reporter behaves like an exon and is activated when it becomes fused to the upstream exons of the *abc* gene via its splice acceptor site (SA). (D) In poly-A trap vectors, the selection gene (*neo*) is constitutively expressed but lacks any poly-adenylation (pA) signal. Thus, stable selection transcripts are only made when *neo* becomes fused via its splice donor site (SD) to an endogenous poly-A signal. The *lacZ* reporter gene is activated as in a gene trap vector.


A : PROMOTER TRAP
ATG
LacZ
pA
P
neo pA
mRNAs:
a
5' RACE
B : ENHANCER TRAP
wP
a b c
C : GENE TRAP
SA
D : Poly-A TRAP
SD
neo
c
AAAAAAAA
3' RACE

selection (Niwa *et al.*, 1993; Yoshida *et al.*, 1995; Imai *et al.*, 1995). In a p-AT vector, the selection gene is expressed from a constitutive promoter and it is ending with a splice donor site instead of a poly-A signal. Therefore, $G418^R$ clones must contain insertion events putting the *neo* gene upstream of an exon containing a poly-A signal (Fig. 1D); these sites alone will allow production of stable neo transcripts, whether the endogenous gene at the locus is active or not during selection. The sites of insertion of the selected $G418^R$ clones can then be analyzed for the presence of a gene and for the expression pattern of the reporter gene.

The pros and cons of trapping vectors

The trapping vectors have three main applications: i) they can reveal, via their reporter gene, the expression patterns of the genes located at the insertion sites; ii) the known vector sequence facilitates the cloning of endogenous sequence at the insertion sites; iii) the insertion may alter the function of endogenous genes. The different types of trapping vectors carry specific advantages regarding these applications.

The PT vectors are good mutagens. However, a promoterless reporter gene requires an insertion in an exon sequence (Fig. 1A; insertions into introns will be excised during endogenous RNA maturation) and in frame with the endogenous protein if it does not carry its own translation start codon (ATG). In eukaryote genomes the frequency for such insertion events is very low because of the relatively large proportion of untranslated sequences. Therefore the use of enhancer and gene trap vectors have been largely preferred in eukaryotes.

The frequency of activation of the reporter gene is much higher for enhancer trap than for promoter trap insertions because of the difference in size of the genomic regions able to activate the vector: ET vectors are able to detect regulatory elements at large distances. On the other hand, an ET inserted into an intergenic region may have little mutagenic effects (see endogenous transcript in Fig. 1B) and the cloning of the insertion locus may not necessarily give a direct access to the gene of interest.

The GT vectors are the most balanced in terms of possible applications: like PT, they only detect insertions within genes and at a higher rate than with a PT vector because "activating" insertion sites include the introns which are usually larger than exons. Activation of the reporter gene is also less dependent on the structure of the genomic fusion: the reporter gene becomes fused with the endogene via its S.A. sequence. The fusion transcript mimics very closely the expression pattern of

the endogene, and produces a truncated endogenous protein that is more likely to have altered activity. Furthermore, the direct fusion of exon sequence with vector sequence in the fusion transcripts greatly facilitates direct cloning of the endogene. Such cloning involves the 5' Rapid Amplification of cDNA Ends-Polymerase Chain Reaction technique (RACE-PCR): with a primer specific to vector sequence, the fusion transcript can be used as a template by the reverse transcriptase to generate a first strand of cDNA containing endogenous sequence (Fig. 1 C, D): this cDNA can then be amplified by PCR, sub-cloned and sequenced (see Fig. 11, Chapter III, p. 79).

PT, ET and GT vectors require some gene expression to be activated and insertions into gene that are not expressed go undetected. The p-AT vectors are good for detection of genes that are not expressed in ES cells. In p-AT vectors insertions, it is also possible to clone downstream endogenous sequence with the 3' RACE-PCR technique, which amplifies sequence between the vector and the poly-A tail (Fig. 1D). However, they are also activated by insertion into the 3' untranslated region of genes which may affect their potential to disrupt normal gene function. Furthermore, this strategy does not exclude insertions into A/T rich intergenic regions which are likely to contain cryptic poly-A signal sequences.

Elements used in trapping vectors

Apart from the features characterizing the type of vectors cited above, other elements have also been added to the basic structure of trapping vectors, mostly to GT vectors. These various elements are listed, with their properties, in Table 1. These improvements to the vectors can be made at different levels: they mostly affect the frequency or the specificity of insertion events able to activate the reporter gene, or can provide additional help for the cloning of the insertion locus.

Different vectors have been used to introduce trapping constructs into mouse cells. Recombinant retroviruses are efficient vectors because they insert with no further rearrangement than a 4 to 6 bp duplication of genomic sequence. Plasmid vectors introduced into cells by electroporation are also used to insert the trapping elements into the genome. They can also integrate without major DNA rearrangement. With both techniques it is possible to control the vector copy number in the selected clones. Multiple copies of a trapping vector inserted in tandem are a problem because they could produce false positive clones (Yoshida *et al.*, 1995). For example, if transcription initiates in a vector copy and runs through

an adjacent copy, the promoter driving the selection gene expression could also direct the expression of the reporter gene located downstream. Multiple copies may also create artificial regulatory elements and therefore mis-leading expression patterns.

Different types of selection genes have been used. The *histidinol dehydrogenase* (*hisD*) gene, confers resistance to L-histidinol and was used in some early promoter trap experiments but it may have interfered with germline transmission (von Melchner *et al.*, 1992). The *E. coli neomycin* gene (*neo*) confers resistance to the G418 drug in mammalian cells and does not alter the potential of cells for germline transmission. It is commonly used as a selection gene in the mouse.

The *lacZ* reporter gene is the most popular cell marker used in eukaryotes: detection of its expression is easy and it is developmentally neutral (Sanes *et al.*, 1986). The Green Fluorescent Protein (GFP) reporter gene is an efficient reporter in bacteria, *C. elegans* (Chalfie *et al.*, 1994), *D. melanogaster* (Yeh *et al.*, 1995) and in the mouse (Zernicka-Goetz *et al.*, 1997). The fusion between the *lacZ* and *neo* genes (*βgeo* gene) allowed the combination of selection with detection of activated insertions. The sensitivity of detection for genes expressed at low levels was also increased: levels of expression sufficient to confer $G418^R$ do not necessarily produce detectable levels of ß-Gal (Friedrich and Soriano, 1991). Another *lacZ* fusion gene, *βgyg*, that can act both as a selection and as a reporter, contains the gene providing resistance to hygromycin (Natarajan and Boulter, 1995). This new selection/reporter gene is an interesting alternative and would allow double selection in experiments involving cells that may already contain the *neo* selection gene.

Splice donor sites, internal ribosome entry sites (IRES), stop and ATG start codons have all been added in various combinations to improve the efficiency of reporter gene or selection gene activation by endogenous sequences (see Table 1). Sub-cellular localization signals such as the nuclear localization signal (nls) (Bonnerot *et al.*, 1987) and the Tau (Callahan and Thomas, 1994) tags are independent of the insertion site and may improve the detection of the cells expressing the reporter gene. For example, the axonal transport of the reporter gene product by the Tau signal may help determining the morphology of neurons when a subset of neural cells in a tissue is expressing the reporter gene.

Other modifications affect the selectivity of reporter gene activation. For example, in "secreted protein specific" gene trap vectors (ST vectors, Skarnes *et al.*,

1995), the addition of a transmembrane domain (TM) in a GT vector selects for insertion into genes containing a peptide secretion leading signal: proper cytoplasmic or membrane-bound ß-Gal enzymatic activity is only possible when both an endogenous leading signal and the vector TM domain are present in the fusion protein. Otherwise, the ß-Gal fusion protein is first sequestered in secretory vesicles where the pH neutralizes its activity and is then secreted by the cells.

The cloning of genomic sequence from the insertion locus can be facilitated by including in the vector, helper elements for plasmid rescue. They facilitate the cloning of flanking genomic sequence when 5' or 3' RACE-PCR were not possible or when analysis of the genomic structure at the insertion site is needed.

Overall, the trapping vectors present many advantages for the detection, cloning and mutagenesis of novel genetic loci. All vectors insert apparently randomly into the genome: the bulk of selection, expression analysis and the cloning of trapped genes can be done *in vitro* in tissue culture systems such as ES cells in the mouse. Trapping vectors have thus been employed in a number of genetic screens following different strategies designed to identify novel developmental genes.

Screening strategies

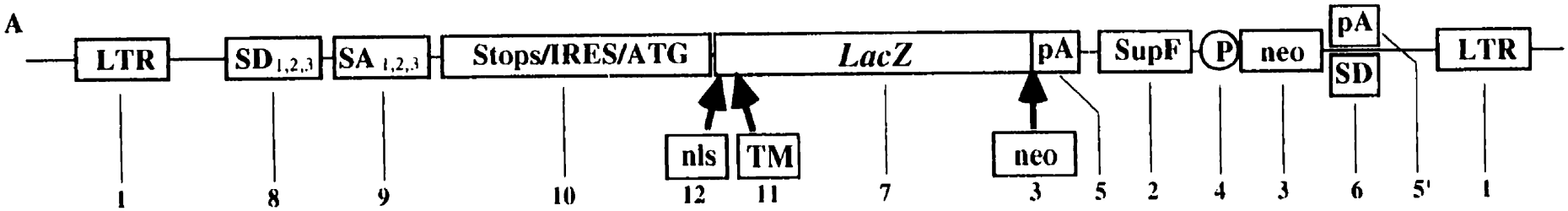
Although all trapping screens initiate with the identification or the selection for insertions into potential transcriptional units, different routes can be followed to further select for the loci that will be studied.

Defining developmental genes

Developmental genes can be defined according to three criteria: their function, structure and expression. These measures allow different approaches to determine whether the product of a gene is a good candidate as a regulatory protein involved in the patterning of the embryo. The function of a gene product can be analyzed through the developmental defects resulting from its alteration in mutant alleles. The structure of a gene or parts of it may be homologous to that of other known developmental genes. Furthermore, structural analysis allows one to attribute possible functions to a gene. Finally, because the expression of developmental genes is often spatially and temporally restricted during embryogenesis, the expression pattern of a gene may also suggest that it plays an important role during development.

Table 1

A



B

#	Element	Vectors[a]	Description
			Vector insertion and genomic locus cloning:
1	LTRs	Opt.	The LTRs of retroviral vectors integrate efficiently into the genome without major rearrangement. Thus, they facilitate the cloning of the genomic locus which is immediately adjacent.
2	Cloning aids	Opt.	The genomic sequence fused to the vector can be cloned by plasmid rescue into *SupF*$^-$ or *Amp*^{s} bacterial cells using the *SupF* or the *Ampicillin* selection markers.
			Selection for stable insertion:
3	Selection gene	All	The *neomycin* (*neo*) gene is the most frequent marker. Alternates *hygromycin* β-*galactosidase* (*lacZ*) fusion with *neo* (*βgeo*) or *hygromycin* (*βgyg*).
4	Constitutive promoter	All	All stable insertions (inside and outside genes) are selected. Only used when the selection gene is not fused to the reporter gene.
5,5'	poly-A	All	(5) Poly-adenylation signals ensure the stability of the transcripts. Furthermore, when the endogenous gene becomes fused to the reporter gene, its transcripts are truncated at this site. (5') Absent in p-AT vectors.
6	P-neo-SD	pAT	The neo gene is only activated when it inserts upstream of both a SA site and a polyadenylation (pA) signal. Selects for insertion into genes whether they are expressed or not.
			Reporter gene activation:
7	Reporter gene	All	The *β-galactosidase* (*lacZ*) gene is the most popular. The selection gene alone has also been used as a reporter. The *β-glucoronidase* (*GUS*) gene is used instead of *lacZ* in plant cells.
8	Splice Donor$_{1,2,3}$	GT (plants)	Used in plants where introns are smaller & insertions are likely to occur into exon downstream of a SA site.
9	Splice Acceptor$_{1,2,3}$	GT pAT	At least one SA is required. Adding SA in the 3 reading frames increase the chances of producing active B-Gal when it does not have its own ATG.
10	Stops/ IRES/ATG	PT, ET (GT pAT)	Eliminates the requirement to respect the reading frame between the endogene and the reporter gene. The reporter gene is translated on its own. The ATG alone is necessary in ET with a weak promoter. In pAT where insertion may occur in 3' untranslated regions, a Kozak ATG may increase the rate of reporter activation.
11	Trans-membrane Domain	ST.	In the secretory trap vector, the reporter gene is only activated when it inserts downstream of a sequence coding for a signal peptide (i.e. into genes coding for transmembrane or secreted proteins).
			Reporter detection:
12	Nuclear localization signal	Opt.	When B-Gal is fused to a protein, the pattern of expression reflects the sub cellular localization of the targeted protein. The nls sequence translocates the reporter protein to the nucleus. Positive cells are more easily identifiable and the signal is very distinct from background B-Gal activity.

[a] Vectors requiring these elements are indicated; vectors using these elements as an option are in parentheses; Opt.: an option for all vectors. *Abbrev.*: IRES: internal ribosomal entry site; LTR: long terminal repeat; nls: nuclear localization signal; SA: splice acceptor site; SD: splice donor site; ST: secretory gene trap; TM: protein transmembrane domain.

Screening criteria

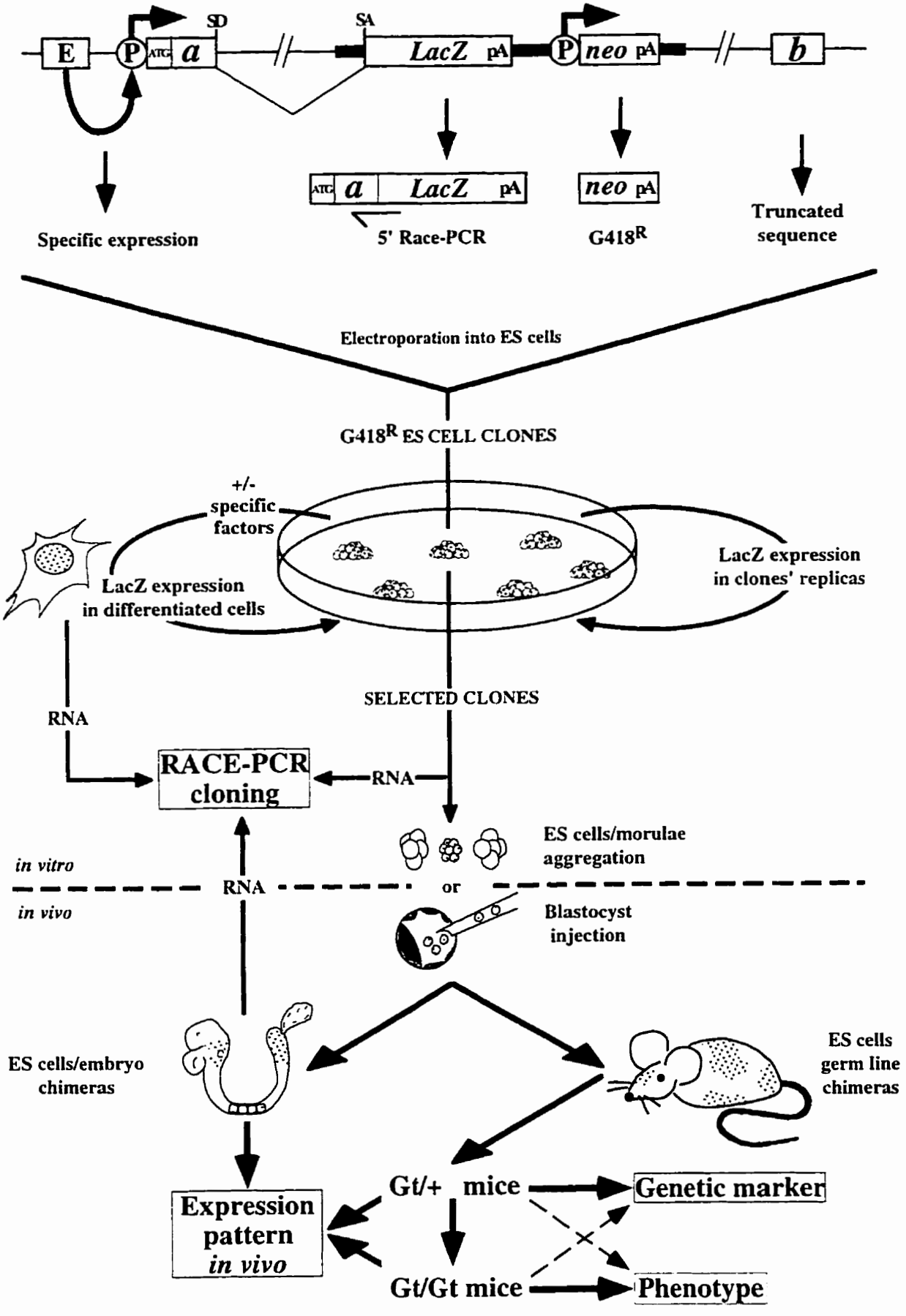
Therefore three approaches can be used independently to characterize trapped loci (see an overview of screening strategies in Fig. 2): the type of expression displayed by the reporter gene *in vitro* or *in vivo*; the identification of a phenotype in mice homozygous for the locus disrupted by the trapping vector; the direct cloning of the locus (using the 5' or 3' RACE-PCR techniques, the plasmid rescue of flanking genomic sequences or mini-libraries derived from the mutated genome. Each type of trapping vector favors some approaches more than others. For example, since enhancer traps do not necessarily insert within transcription units, they are less likely to disrupt a gene function and do not always facilitate the cloning of genes. However, they detect regulatory elements at very high frequencies. Gene trap vectors have been used in the majority of screens and they were used following the three possible approaches. However, although a screen primarily involving the cloning of insertion sites (Yoshida *et al.*, 1995) and a large screen for gene trap insertions causing a phenotype in mice (Friedrich and Soriano, 1991) were reported, the screening for the expression pattern of the reporter gene *in vivo* has been used in most studies. First, insertions showing a restricted pattern of expression are likely to be linked to a gene involved in embryo patterning; second, it is necessary when the selection for stable vector integration does not eliminate insertions into silent loci (with vectors using constitutive expression of the selection gene or with p-AT vectors).

Target cells

The choice of target cell has great consequences on the possible outcomes of a trapping screen. In the mouse, ES cells and multipotent cells have been used because they can differentiate *in vitro* and can be screened for potentially interesting insertion events. It is important to consider the target cell's pre-screening potential. Although trapping approaches on multipotential cell lines (for which inducible differentiation pathways are well defined) led to the cloning of novel genes (Okazaki *et al.*, 1994; Imai *et al.*, 1995), the functional analysis of the latter *in vivo* would require *de novo* mutagenesis by homologous recombination in ES cells. ES cells provide the most flexible experimental system from the screening steps *in vitro* to the genetic and expression analyses *in vivo*, and all the strategies used in the analysis of trapping events are applicable to ES cells. ES cells can give rise to a wide range of differentiated cell types *in vitro* and it is possible to screen

Fig. 2. - Diagram representing potential screening strategies.

An array of analytical criteria is available for the selection of gene trap vector insertions into mouse ES cells.


E
P
ATG
a
SD
SA
LacZ
pA
P
neo
pA
b
Specific expression
ATG
a
LacZ
pA
5' Race-PCR
neo pA
G418R
Truncated sequence
Electroporation into ES cells
G418R ES CELL CLONES
+/-
specific
factors
LacZ expression
in differentiated cells
LacZ expression
in clones' replicas
SELECTED CLONES
RNA
RACE-PCR
cloning
RNA
ES cells/morulae
aggregation
in vitro
RNA
or
in vivo
Blastocyst
injection
ES cells/embryo
chimeras
ES cells
germ line
chimeras
Expression
pattern
in vivo
Gt/+ mice
Genetic marker
Gt/Gt mice
Phenotype

specific factors for the induction or repression of reporter gene expression (e.g. retinoic acid in Forrester *et al.*, 1996). With a good ES cell line like R1 (Nagy *et al.*, 1993) chimeric embryos and germline transmission can be obtained by aggregation with wild-type embryos. Embryos totally derived from ES cells can be made by aggregation with tetraploid embryos. The latter can be very useful for the study of dominant embryonic lethal phenotypes (Carmeliet *et al.*, 1996) or to confirm expression patterns observed in chimeric embryos (Wurst *et al.*, 1995). ES cells can also be selected via Fluorescence Activated Cell Sorting (FACS) analysis without loosing their developmental potential (Reddy *et al.*, 1992).

Using multipotent cells and in particular ES cells, it is possible to pre-screen insertion events *in vitro*, thus making large scale screens possible in the mouse system.

Trapping screens in the mouse

The screens listed in Table 2 represent examples of the different strategies that have been used to detect novel genes important for development. The enhancer trap screens performed in *D. melanogaster* were not included in this table: since the first screens (O'Kane and Gehring, 1987; Bier *et al.*, 1989; Bellen *et al.*, 1989), trapping vectors carried by the transposable P-element have been used in multiple screens to identify patterns of expression, create genetic markers and induce mutations; presently, these vectors are being used in large scale projects to saturate the *D. melanogaster* genome with insertion mutations and to facilitate its mapping and sequencing (Spradling *et al.*, 1995).

The first trapping screen in the mouse used an enhancer trap vector and the transgenic mice technology (Allen *et al.*, 1988). Among the 52 (out of 200 injected) embryos that integrated the *HSV-TK-lacZ* enhancer trap, 11 embryos expressed ß-Gal in a spatially restricted pattern. However, this experiment also suffered the drawbacks associated with the micro-injection technique: Some embryos had up to 500 copies of the vector; they were all sacrificed to analyze the reporter gene expression, thus preventing any further study on these insertions *in vivo*; finally the cloning of disrupted genes would be impaired by the chromosomal rearrangements caused by the insertion. In all the other screens performed in the mouse, the trapping vectors were introduced by electroporation or with a recombinant retrovirus into a variety of ES cells or multipotent cells (Table 2).

In the mouse, most screening strategies were based on the identification of insertions giving a restricted expression pattern to the reporter gene. Screens

Table 2 Trapping Screens

Screening strategy				Results			Ref.
Target genome[a]	Vector type[b]	No. of insertions selected *in vitro*[c]	Target genes	No. of restricted expression pattern *in vivo*[d]	GI.	Phenotype[e]	
Screens performed in the mouse							
Zygotic egg	ET	52 Transgenic embryos 20 Transgenic strains	None	11 (F0 Transgenic embryos) 5 (F1 Transgenic embryos)	n.a. All	n.a. n.d.	(1)
D3	ET	59 G418^R	None	27 (46%)	2	viable	(2,3)
	GT	7 G418^R & LacZ+	Expressed in ES cells	1	3	2 r.e.l.	(4)
D3	PT	7 His^R	Expressed in ES cells	n.a. (*hisD* is the reporter gene)	]c	1 r.e.l.	(5)
	PT	3 G418^R		n.a. (*neo* is the reporter gene)	3	1 r.e.l.	
AB1 AB2.1	GT	42 ßgeo+ transmitted to the germ line	Expressed in ES cells	15 (55%, 15/27 reported in ref. 6) Assayed in F1 embryos	42	18 r.e.l. 2 n.s.	(6, 7)
D3	GT	279 G418^R & LacZ+	Expressed in ES cells	36 (13%)	n.d.	n.d.	(8)
CGR8	GT-ST	11 ßgeo+ & LacZ+	Secreted proteins expressed in ES cells	6 (54%)	3	1 r.e.l.	(9)
AB-1	GT	11 ßgeo+ transmitted to the germ line	Expressed in ES cells	1 (10%) Assayed in F1 embryos	11	1 r.e.l.	(10)
R1	GT	20 ßgeo+ & RA responsive	Expressed in ES cells and responsive to R.A.	19 (95%)	5	1 r.e.l.	(11)
MS-1	GT-pAT	26 G418^R & LacZ+	Expressed in ES cells	6 (23%)	0	n.a.	(12)
D3		12 "	"	n.d.	9	n.d.	
TT2		6 G418^R & randomly cloned	poly-A signal	*1 (on Northern blot; others n.d.)*	1	viable	(13)
E14	GT	19 G418^R & LacZ+	Expressed in ES cells	1 (5%)	1	1 r.e.l.	(14)
				Total (PT & GT for which phenotype was tested)[e]:	70	26 r.e.l.	
P19	GT-pAT	10 G418^R & LacZ+	R.A. responsive gene	*5 (up or down regulated by R.A.)*	n.a.	n.a.	(15)
Screens in other experimental model systems							
C. elegans	PT	96 recombinant DNA pools	Expressed in vivo	10 (n.d.)	10	n.a.	(16)
A. thaliana	ET	214 selected in vivo		40 (18%)	All	8	(17)
"	GT	246 "		29 (12%)	All	10	"
"	ET	87 "		53 (60%)	All	1	(18)

(Notes on next page)

[a] ES cell lines are indicated in bold type.

[b] E-, G-, P-, pA, S-T: enhancer-, gene-, promoter-, polyA- and secretory trap respectively; the polyA and secretory trap vectors are forms of gene trap vectors.

[c] Only the lines analyzed further are shown here; Bgeo+ lines are all $G418^R$ and around 95% of them are LacZ+.

[d] Analyzed between E8.5 and E12.5 in ES cell/embryo chimeras unless specified otherwise (*in vitro* analyses in *italics*). The percentage represents the efficiency of pre-selection: it refers to the number of insertions pre-selected in vitro and showing a restricted expression pattern in vivo.

[e] For all germ lines (i.e. including some cell lines that did not show a restricted expression pattern of the reporter gene in vivo).

Abbreviations:

GL: No. of germ line transmission; m.s.: male sterile; n.a.: not applicable; n.d.: not determined; r.e.l.: recessive embryonic lethal.

References:

(**1**)Allen, *et al.*, (1988); (**2**)Gossler, *et al.*, (1989); (**3**)Korn, *et al.*, (1992); (**4**)Skarnes, *et al.*, (1992); (**5**)von Melchner, *et al.*, (1992); (**6**)Friedrich and Soriano, (1991); (**7**)Chen, *et al.*, (1994); (**8**)Wurst, *et al.*, (1995); (**9**)Skarnes, *et al.*, (1995); (**10**)Deng and Behringer, (1995); (**11**)Forrester, *et al.*, (1996); (**12**)Niwa, *et al.*, (1993); (**13**)Yoshida, *et al.*, (1995); (**14**)Takeuchi, *et al.*, (1995); (**15**)Imai, *et al.*, (1995); (**16**)Hope, (1991); (**17**)Sundaresan, *et al.*, (1995); (**18**)Klimyuk, *et al.*, (1995).

usually involve two rounds of selection: a pre-screen for insertions that activate the reporter (Wurst *et al.*, 1995) or selection gene (pAT vectors, Niwa *et al.*, 1993; Yoshida *et al.*, 1995; Imai *et al.*, 1995) or both (ßgeo fusion in Friedrich and Soriano, 1991; Skarnes *et al.*, 1995; Takeuchi *et al.*, 1995) in cultured cells; a second screen analyzes expression *in vivo* using ES cell chimeras (Korn *et al.*, 1992; Skarnes *et al.*, 1992; Wurst *et al.*, 1995; Forrester *et al.*, 1996) or germline chimeras (Friedrich and Soriano, 1991). Pre-screening *in vitro* selects for insertions activating the selection and reporter gene in undifferentiated ES (Friedrich and Soriano, 1991; Wurst *et al.*, 1995; Skarnes *et al.*, 1995) and regulating the reporter gene expression after induction of differentiation *in vitro* (Forrester *et al.*, 1996; Imai *et al.*, 1995).

These screens yielded various kinds of information. The screens focusing on the pattern of expression *in vivo* illustrated the types of expression patterns that could be detected and their frequencies. Interestingly, the frequency of genes that are regulated during development ranged from 10% to 50% in different screens (Table 2). These variations could be due to the vectors' structure and the number of developmental stages analyzed. The screen for genes containing a signal sequence produced two independent insertions at a single locus and preferentially detected large transmembrane proteins, suggesting that this approach (or the vector used) may be biased toward a limited number of target genes (Skarnes *et al.*, 1995). The screen for insertion into genes that are induced or partially repressed by retinoic acid *in vitro* clearly enriched for insertions giving a restricted pattern of expression *in vivo* (95% of them, Forrester *et al.*, 1996). Screens that also involved the transmission of most or some of the insertion to the germline provided more functional data. In total, 70 insertions (42 are from the screen of Friedrich and Soriano, 1991) were transmitted to the germline and about 40% resulted in recessive embryonic lethal mutations (n=26) or male sterility (n=2) (Table 2). However, the majority of mutations result from insertions into genes that are ubiquitously expressed. Such genes are less likely to encode proteins with patterning functions. Generally, insertions have only been partially characterized. Some insertions detecting an "interesting" pattern of expression were not transmitted to a mouse transgenic strain and/or were not cloned. Likewise, not all the insertions analyzed *in vivo* in transgenic mice have been cloned. Therefore, it is difficult to estimate the number of insertions that trapped developmental genes. Further characterization of these insertions and results from other trapping screens in progress in different laboratories will generate more comprehensive data.

Table 3 Genes cloned by trapping approaches

Insertion	Gene[a]	Cloning	Remarks	Chr.	Phenotype	References
Enhancer trap in mouse ES cells						
6028	***Etl-1***	genomic	Some homology to the fly Brahma and yeast SNF2 SW12 transcription factors.		no GL	(Soininen, 92)
3015	***Etl-2***	"	Novel type I cytokin receptor.	4	no GL	(Neuhaus, 94)
Etl-4	n.d.	"	Linked to *Danforth's short tail*	2	viable	(Koseki, 93)
Promoter trap in mouse ES cells						
1A4	*REX-1*	genomic	Expressed in the testis.		m. s.[b]	(vonMelchner, 92)
1B6	*fug-1*	genomic 5' RACE	Yeast *RNA1*	15	r. e. l.	(DeGregori, 94)
J3A3	*eck*	genomic 5' RACE	*Eph* related receptor tyrosine kinase.		viable	(Chen, 96)
Gene trap in mouse ES cells						
	GT10	5' RACE			no GL	(Skarnes, 92)
	GT4-1	5' RACE			pn. death	"
	GT4-2	5' RACE	Zinc-finger protein		def. growth	"
ROSA5	*TEF-1*	genomic	Yeast transcription factor family.		r. e. l.	(Chen, 94)
GtC101	***cobl***	5' RACE	Human *cobl* ESTs identified.	11	viable	(Gasca, 95)
ROSA 3-7	*BTF3*	5' RACE	Human *BTF3*		r. e. l.	(Deng, 95)
B6	***jumonji***	5' RACE	Some homology to *RBP2*		r. e. l.	(Takeuchi, 95)
514	*netrin*	5' RACE	Homologue of chicken *netrin1*		N.A.	(Skarnes, 95)
534	*LAR*	"	Two independant insertion in the same gene.		viable	
484	"	"			n.d.	
497	***sek***	"			n.d.	
519	***nCadherin***	"	New member of nCadherin receptors family.		n.d.	
531	*PTPk*	"			viable	
	AYU-(1,2,3,5,6)	3' RACE genomic	ES cell lines were not capable of germ line colonization.		n.a.	(Niwa, 93)
	I.114	5' RACE	I. & R. refer to induction or repression by R.A.. All show restricted expression patterns during embryogenesis.		viable	(Forrester, 96)
	I.163				"	
	I.193				"	
	R.140				r. e. l.	
I.214	*MARsM4*	"	Rat *muscarinic acetylcholine receptor* [*subtype M4*			
R.24	*c-fyn*					
	pat-12	3' RACE		10	viable	(Yoshida, 95)
PAT-8.2	*Etn3* LTR	"			n.d.	
pat-(5, 7, 8.1)		"			n.d.	
Gene trapping insertions in mouse multipotent cell lines						
G19	***ZFP-57***	genomic	Zinc finger protein		n.a.	(Okazaki, 94)
	GT1	genomic 5' RACE	Neuron specific, induced by R.A.		n.a.	(Imai, 95)
c	***gas5***	genomic 5' RACE	Family of growth arrest specific genes.		n.a.	(Bonnerot, 92)
Other model systems						
24C7 *C. elegans*	***pes-1***	genomic	Some homology with a domain of *D. melanogaster forkhead*.	n.a.	n.a.	(Hope, 94)
GT148 *A.thaliana*	***Prolifera***	5'RACE	Some homology with the yeast *MCM2-3-5* gene	n.a.	r. e. l.	(Springer, 95)

a Novel genes are in bold type; genes already isolated in other species are underlined; others were known mouse genes.

b Sterility could result from disrupted expression of *REX-1* normally expressed in the testis or could be a toxic effect due to the expression of *hisD* in the germ line as seen with the *HSV-tk* gene (Wilkie *et al.*, 1991).

c This unintentional gene trap retroviral vector also inserted into this locus in 21 independent clones.

Abbreviations.: Chr.: chromosomal location; def.: defective; GL: germ line transmission; m.s.: male sterile; n.a.: not applicable; n.d.: not determined; pn.: perinatal; r.e.l.: recessive embryonic lethal;

Large screens performed in the worm *C. elegans* (Hope, 1991) and the plant *A. thaliana* (Klimyuk *et al.*, 1995; Sundaresan *et al.*, 1995) demonstrate that this approach is also very efficient in these model systems to generate developmental markers (Hope, 1991; Young and Hope, 1993) and induce mutations (Table 2 and Klimyuk *et al.*, 1995; Sundaresan *et al.*, 1995).

Molecular characterization of trapped genes

Thirty-six mouse genes (listed in Table 3) were cloned following integration of a trapping vector. Two genes were cloned by the enhancer trap approach and 3 after promoter trap insertions (see Table 3 for references). The majority of genes were cloned following gene trap approaches. In most cases, endogenous sequence was cloned using 5' RACE-PCR techniques. The 3' RACE approach was also used for the poly-A trap vector insertions. For all enhancer trap and some gene trap insertions, the cloning of genomic sequence flanking the insertion locus was necessary. In these cases, cloning was made by plasmid rescue or through the screening of genomic libraries made from the selected clones. Nevertheless, when an insertion occurred in a big intron or at a large distance from the trapped gene, this cloning approach may still require additional cloning steps, involving chromosomal walking. Overall the cloning of genes after insertion of a trapping vector was easier when it was possible to use the 5' RACE-PCR technique.

Five of the genes were already known in the mouse. Eight of the newly cloned genes are novel in the mouse but represent homologues of genes cloned in other species or belong to gene families, suggesting that they could have been cloned by other means. Yet, the majority of genes (23) are completely novel (or present only weak homology with known sequences), thus validating the trapping approach to directly access new genetic loci.

Some of these insertions also disrupted the normal function of the trapped genes (*fug-1*, *TEF-1*, *BTF3*, *R.140* and *jumonji*) and created recessive embryonic lethal mutations. These insertions provide good examples of the polyvalence of the gene trapping vectors: Single insertion events allowed researchers to simultaneously detect and mutate potential developmental genes and to generate transgenic mouse strains that are essential to analyze the function of these genes *in vivo*.

THESIS WORK

The development of the mouse embryonic axis

Thus, trapping vectors and in particular GT vectors provide us with good tools to identify and analyze novel mouse genes and to investigate their role during development.

I was particularly interested in the characterization of novel genes involved in axis formation and patterning. During early mouse embryogenesis, embryonic axes are established at the time of gastrulation. As the different germ layers of the embryo are formed by the gastrulation process, they are organized around the anterior-posterior (A/P) and dorsal-ventral axes. These tissues are simultaneously divided into territories, prefiguring the future body plan, whose identity is initially controlled by the expression of patterning genes. The signals controlling A/P axis formation and patterning originate from an organizer tissue whose function has been demonstrated *in vivo* (Spemann and Mangold, 1924; and reviewed by Streit *et al.*, 1993). This inducing activity could also be divided into different classes of signaling potential. First, organizer tissue from older developmental stages was only capable of inducing more posterior structures, thus suggesting the existence of distinct head, trunk or tail organizer activities. Second, the organizer showed both mesoderm and neural inducing activities (reviewed in Gilbert, 1994). Expression cloning and analysis of genes expressed in the organizer region has allowed investigators to identify molecules that are also able, with varying potential, to induce axial structures (reviewed by Lemaire and Kodjabachian, 1996). Other candidate genes, such as *Brachyury* (*T*) (Herrmann *et al.*, 1990), *nodal* (Zhou *et al.*, 1993; Conlon *et al.*, 1994) and *embryonic ectoderm development* (eed, Niswander *et al.*, 1988; Faust *et al.*, 1995; Schumacher *et al.*, 1996) were identified in the mouse as genetic variants altering the proper formation of the axis during gastrulation.

However, the full spectrum of genes involved in neural and axial specification and their modes of interaction are not yet clear. In the mouse and other vertebrates, many of these genes are expressed in the node (or equivalent organizer structures in other vertebrates) and its descendant, the notochord. The notochord plays a particular role during development as the anatomical embryonic axis and as a source of patterning signals (reviewed by Placzeck, 1995). Therefore, genes expressed in the node and the notochord are good candidate genes that are likely to play a role in axis formation or patterning.

Further screens directed at gene expression during mouse gastrulation and axis formation, at 8.5 dpc, could prove fruitful in identifying novel mammalian genes involved in these processes. The completion of a gene trap screen at this stage of development should also provide a large number of candidate genes involved in other patterning processes. Furthermore, the identification of novel "axis specific" genes controlling the expression of the *lacZ* reporter gene would provide useful markers to investigate the phenotype of known mutations affecting axis development.

Embryos carrying mutations in either of two loci, *Brachyury* (*T*) or *Danforth's short tail* (*Sd*), are good and complementary experimental models to analyze the development of the notochord and its role in embryo patterning. The *T* mutation alters early development of the trunk axis and only notochord precursors are formed. The *Sd* mutation affects the survival of the notochord and thus has more effects on the patterning role of the notochord. Moreover, these 2 mutations are semi-dominant, thus allowing the study of axis formation in both heterozygous and homozygous embryos. These mutant backgrounds can be used to further characterize novel genes involved in the development of the mouse axis.

Thesis work

In the following chapters, I present experiments investigating some of the potential of the gene trap approach for the genetic and molecular study of vertebrate embryonic development.

The GT vector, pGT4.5a, containing the *lacZ* reporter gene and the *neo* selection marker was developed in the laboratory (Gossler *et al.*, 1989). This GT vector was introduced into ES cells and chimeric embryos were generated to screen for insertions of the *lacZ* reporter gene into transcription units that are spatially and temporally regulated during early mouse embryogenesis (Gossler *et al.*, 1989). This GT vector could act as a reporter, a mutagen and a molecular tag to clone new mouse genes (Skarnes *et al.*, 1992). Three modified versions of this GT construct, the vectors pGT4.5a, PT-1 and PT-3 were designed to improve gene trapping efficiency (W. Skarnes; D. P. Hill, unpublished results).

My Thesis work started with the testing of these vectors in order to use them in a large screen for GT insertions in mouse ES cells. We carried out a pilot screen and a large-scale screen for insertions into genes that were regulated during early embryogenesis. I characterized a GT insertion into a novel gene, *cordon-bleu*, specifically expressed in the axial structures of early mouse embryos. I used this GT

insertion as a marker to analyze axis development in *Brachyury* and *Danforth's short-tail* mutant embryos. The large screen and the analysis of one particular insertion illustrate several important aspects of the potential uses of GT insertions for developmental biology studies.

- CHAPTER II -

A LARGE SCALE GENE TRAP SCREEN

The results of the "large screen" presented in this chapter appeared in the following publication:

A Large Scale Gene-Trap Screen For Insertional Mutations In Developmentally Regulated Genes In Mice.

* Wolfgang Wurst, Janet Rossant, Valerie Prideaux, Malgosia Kownacka, Alexandra Joyner, David P. Hill, François Guillemot, Stéphan Gasca, Dragana Cado, Anna Auerbach and Siew-Lan Ang. Genetics (1995), **139**, 889-899.

**All authors contributed equally to this project.*

CHAPTER TWO : A LARGE SCALE GENE TRAP SCREEN

INTRODUCTION

In *D. melanogaster* and *C. elegans*, the ability to carry out large scale screens for developmental mutations has proven essential in unraveling the molecules and the genetic programs controlling early embryogenesis. In the mouse, this strategy is made difficult by the development of the embryos *in utero*.

The establishment of mouse embryonic stem (ES) cell lines in culture now permits investigators to manipulate the mouse genome and to select clones carrying the desired mutations *in vitro*. The foremost genetic approach using ES cells involves targeted mutagenesis of genes via homologous recombination (Capecchi, 1989; Koller *et al.*, 1989; Koller and Smithies, 1992). Thus, by inverse genetics the function of cloned genes, that are predicted to be important for embryogenesis, can be altered *in vitro* and subsequently analyzed *in vivo*. Most often, such genes are identified by homology either to genes that have been shown to be developmentally important in other species, or to genes that contain functionally conserved protein domains of interest. Although this method of selecting candidate genes has proven very successful in identifying important developmental genes, inverse genetic approaches remain limited to the analysis of cloned genes.

ES cells also provide a good system for random mutagenesis strategies. Efficient means of identifying and mutating novel genes by the introduction into ES cells of trapping vectors that drive *lacZ* reporter gene expression from endogenous cellular promoters have been reported (see Chapter I and the reviews by Gossler and Zachgo, 1993; Hill and Wurst, 1993). In the screens presented here, we used GT vectors that serve as an artificial exon after insertion into an endogenous transcription unit (Gossler *et al.*, 1989; Friedrich and Soriano, 1991). When cell lines containing this vectors are used to make chimeric embryos, the localization of ß-Gal activity generated by the fusion protein is very similar to the expression pattern of the endogenous gene found at the site of insertion (Skarnes *et al.*, 1992). Generation of fusion transcripts also makes it possible to directly clone the transcribed region of the disrupted host gene using the rapid amplification of cDNA ends by PCR protocol (RACE-PCR) (Frohman *et al.*, 1988; Skarnes *et al.*, 1992). Finally, GT vectors are mutagenic agents since they sometimes alter the function of trapped genes (Friedrich and Soriano, 1991; Skarnes *et al.*, 1992).

A large number of genes that are critical for vertebrate embryos patterning during development displays specific spatial and temporal regulation at early stages of embryogenesis (e.g. Hox genes, reviewed in Krumlauf, 1994). Thus, following the reverse rationale, we conducted a large screen for insertional mutations into mouse genes that are developmentally regulated during embryogenesis, expecting that these genes may code for important patterning molecules. The results of such a "expression pattern" screen would reveal information on the fundamental domains of gene expression during mouse development and also provide a large number of candidate mutations affecting the development of the tissues expressing the reporter gene.

In this chapter, I report the results of a pilot screen and a large scale screen conducted to identify and mutate genes that are expressed in the mouse embryo around the time of the establishment of the basic body plan at 8.5 day post coitum (8.5 dpc or E8.5). At this stage, gastrulation is still underway in the mouse and neurulation, somitogenesis and the organogenesis of many organs (optic and otic vesicles, heart, notochord, definitive endoderm, liver, blood islands, endothelium, allantois and germ cells) has begun. Patterning processes are also preponderant in these newly formed tissues, as embryonic axes are established and cells committed to specific developmental fate become successively restricted to embryonic compartments and sub-compartments. Indeed, many patterning genes are expressed at this stage and display specific expression domains restricted to various territories, compartments or organs in the vertebrate embryo (reviewed in Faust and Magnuson, 1993; De Robertis *et al.*, 1994; Boncinelli and Mallamaci, 1995).

We have characterized *in vivo* the expression patterns of 303 different genes that activated expression of the reporter gene in embryonic stem cells.

MATERIALS AND METHODS

Vectors

The gene trap vectors used in the screens were pGT4.5a. PT-1 and PT-3. The three vectors were tested during the pilot screen. All the cell lines from the large screen however. were generated with the PT-1 GT vector (Fig. 3A). The vector **pGT4.5a** (obtained from W. Skarnes) is a modification of the gene trap vector pGT4.5 previously described (Gossler *et al.*, 1989). in which the *En-2* poly adenylation signal sequence after the *neomycin* resistance gene was replaced by the SV40 signal. **PT-1** (constructed by D. P. Hill) is a modification of the pGT4.5a vector: the *β-actin* promoter driving expression of the *neomycin* resistance gene was replaced with the *Pgk-1* promoter (Boer *et al.*, 1990). This modification resulted in a 5-fold increase in the number of G418^R colonies per electroporation without affecting the proportion of ß-galactosidase expressing colonies among the resistant colonies (see Table 4). The original vector and PT-1 contain the splice acceptor sequence from the *En-2* gene upstream of the E. coli *β-galactosidase* gene (*lacZ*). lacking its own ATG. The **PT-3** (constructed by D. P. Hill) vector was derived from PT-1 by the addition of an ATG codon at the beginning of the reporter gene coding sequence. In this vector. the reporter gene is translated regardless of the trapped gene reading frame. thus resulting in a 3-fold increase of G418^R colonies expressing ß-Gal (see Table 4).

Electroporation of ES cells and in vitro screening

Before introduction into cells. the vector DNA was linearized by digestion with *HindIII*. The reaction mixture was heated to 90°C for 15 min. and the linearized DNA was ethanol precipitated. DNA was resuspended in phosphate buffered saline (PBS) at a concentration of 1mg/ml for electroporation. Electroporation and screening of ES cells were performed as described previously (Hill and Wurst. 1993; Wurst and Joyner. 1993). After 8-10 days of selection. when G418^R colonies were readily apparent. most of the colonies were individually picked (pilot screen) or replica plated for the large screen (as described in Gossler *et al.*, 1989; Gossler and Zachgo. 1993). In the large screen. colonies that showed any *lacZ* expression on the replica. either scattered or throughout the colony. were picked. expanded. tested again for ß-Gal activity. and frozen away for later analysis in chimeras.

Pilot screen

Single G418^R colonies were picked and transferred to gelatinized 96-wells microtiter plates in DMEM medium (15% fetal calf serum) supplemented with LIF. The differentiation screen was performed entirely in 96-wells plates. After 2 days, cells were resuspended with trypsin, split into 2 plates and cultured under the same conditions. On day 3, 1 plate was stained for detection of ß-Gal activity in undifferentiated ES cells. On day 4, 15% of ES cells were passed onto new plates as an ES cell stock. The remaining cells were cultured in DMEM medium (15% fetal calf serum) in the absence of feeder cells or LIF factor: 75% were passed into non-tissue plates to induce their differentiation into embryoid bodies and 10% were plated onto tissue culture plates to promote their differentiation into a monolayer of fibroblast-like cells. Both types of differentiated cells were stained for ß-Gal activity after 6 days of culture under these conditions. Clones expressing the *lacZ* reporter in ES or differentiated cells were immediately passed onto 15 mm plates for expansion and freezing. Selected clones were tested again under the same conditions before injection into blastocysts.

Production and analysis of chimeras

Beta-Gal positive clones were thawed, grown for one week and then individually injected into blastocysts obtained from outbred CD1 mice (Charles River Laboratories, Quebec). In the pilot screen, all 47 cell lines tested were tested for ß-gal expression at 8.5 and 12.5 dpc and for chimerism at 12.5 dpc (ES cell contribution to the embryo can then be scored by the presence of eye pigmentation derived from the agouti D3 ES cell line in the albino CD1 host embryos; Nagy *et al.*, 1990). In the large screen, for 139 out of 290 clones, 30-40 blastocysts were injected with 12-15 cells each and transferred into the uteri of three recipient females on the third day of pseudo-pregnancy. Two recipients from each clone were sacrificed at 8.5 dpc, when embryos were at early somite stages (4-15 somites). Dissected embryos and their extraembryonic membranes were fixed and stained for ß-Galactosidase activity as previously described (Gossler and Zachgo, 1993; Hill and Wurst, 1993). The numbers of embryos that expressed ß-galactosidase and the pattern of the staining were recorded, and all embryos displaying developmental regulation of ß-galactosidase expression were photographed. The third recipient was allowed to continue the pregnancy until 12.5 dpc when chimerism could be estimated. Most of the 12.5 dpc embryos were also stained for ß-galactosidase activity. The remaining 151 clones were used to

generate 20-30 embryos in two recipients which were sacrificed and analyzed for reporter gene expression on 8.5 dpc. Three chimeras showing identical expression patterns were considered an acceptable minimum because patterns were reproducible from embryo to embryo despite varying degrees of mosaicism (Gossler *et al.*, 1989). In cases where there was doubt concerning the pattern of reporter gene expression, the injections were repeated until three chimeras showing identical patterns were obtained. Information on the majority of clones was derived from more than three chimeras, and data based on two chimeras were reported in a few cases where expression was clearly ubiquitous.

Production of ES-tetraploid chimeras

A few clones that displayed interesting patterns were subjected to analysis after aggregation with tetraploid host embryos (Nagy *et al.*, 1990; Nagy and Rossant, 1993). In such chimeric embryos, the ES cells out compete the compromised tetraploid host cells during development, resulting in embryonic tissues that are derived solely from ES cell descendants (Nagy *et al.*, 1990). The absence of mosaicism allows for more precise confirmation of the ß-Galactosidase staining pattern.

RESULTS

Pilot screen

Prior to the large GT screen performed with the PT-1 GT vector (Wurst *et al.*, 1995), the pGT4.5a, PT-3 and PT-1 vectors (see and Material & Methods for their description) were electroporated into D3 ES cells and different screening strategies were tested in pilot experiments.

Pre-screening in vitro in undifferentiated and differentiated ES cells

As the selection for G418^R clones only selects for cell lines carrying a stable insertion of the GT vector, the first stage of the screen is to detect events that are able to activate the reporter gene and for which insertions probably occurred into genes. The obvious candidate ES cell clones are those that already express *lacZ* at the time of selection. However, the final aim of this screen is to detect genes showing spatial and temporal regulation of expression *in vivo*, during early embryogenesis because such genes are likely to be involved in the establishment of the embryo's basic body plan. Another advantage of ES cells is their capacity to spontaneously differentiate *in vitro* into a variety of early embryonic cell types when cultured under certain conditions (Doetschman *et al.*, 1985). Since the differentiation of these cell types *in vitro* tends to parallel the order of appearance of tissues formed *in vivo*, we assumed that the differentiation of ES cells *in vitro* could be used as an additional pre-screening criterion: our hypothesis was that insertions whose reporter gene expression was regulated during ES cells differentiation *in vitro*, would be more likely to have occurred into genes with restricted patterns of expression during early embryo development.

About 10,000 G418^R ES cell clones (carrying the different GT vectors, see Table 4) were assayed for ß-Gal expression. Eighty-four clones expressed the reporter gene in undifferentiated ES cells (see Table 4 below for numbers & percentages). About 0.5% of the G418^R clones expressed the reporter gene in the absence of an ATG codon before *lacZ* (pGT4.5a and PT-1). As predicted, when the reporter gene was provided with its own ATG start codon (PT-3), the number of ß-Gal positive G418^R ES cell clones increased by about 3-fold. All positive ES cell clones and about 6000 negative G418^R clones were split and let to differentiate as a monolayer or as embryoid bodies for a short period of 4 to 6 days (see Material & Methods) before being assayed again for ß-Gal expression. Twelve clones that expressed ß-Gal in undifferentiated ES cells, down-regulated or completely turned off expression of

TABLE 4

Summary of electroporations and *in vitro* differentiation screening for different gene trap vectors

vector	Number of G418^R colonies[a]	*In vitro* screening ES cells-differentiated cells blue-blue	blue-white	white-blue[b]	Frequency blue/G418^R
pGT4.5a	2328 (15:10 to 400)	9	3	14 (1278)	0.51%
PT-1	6135 (4:600 to 2500)	29	7	5 (3420)	0.58%
PT-3	2100 (2:500 to 1600)	34	2	8 (2064)	1.71%
Total	10.563	72	12	27 (6762)	

a: *in parenthesis* (No. of electroporations : range of No. of G418^R colonies/electroporation).
b: *in parenthesis* (No. of white ES cell colonies differentiated *in vitro*).

the reporter gene upon differentiation (in all cells or in sub-population of cells). Among the clones that did not express the reporter gene in ES cells, 26 showed some ß-Gal activity after differentiation. However, the majority of these clones only showed very little ß-Gal staining (present in all cells at very low levels or only detectable in a few cells). Nevertheless, these clones were likely to correspond to insertions into genes involved in early developmental differentiation steps and a subset of them was further analyzed for expression of the reporter gene *in vivo*.

TABLE 5.

No., expression and contribution of cell lines analyzed *in vivo*

Vector	ß-Galactosidase expression at 8.5 dpc Restricted	Widespread	None	No ES cells contribution	Total
pGT4.5a	2	2	8	1	13
PT-1	1	7	4	3	15
PT-3	0	0	1	18	19
Total:	**3**	**9**	**13**	**22**	**47**

Expression screen in vivo

Forty-seven ES cell lines from the different electroporations selected for their pattern of expression *in vitro* (as undifferentiated or differentiated ES cells) were injected into CD1 blastocysts and screened *in vivo* in ES cell chimeric

embryos, by histochemical staining for ß-Gal activity (Table 5). Reporter expression was analyzed in 8.5 dpc embryos and ES cells contribution was estimated by the presence of eye pigmentation in 12.5 dpc embryos when no ß-Gal staining could be observed. The clones derived from electroporations of the PT-3 GT vector contributed very poorly to chimeras (1/19 clone produced chimeric embryos and no ß-Gal expression could ever be observed). It probably reflects the use of an unhealthy batch of ES cells for the electroporation rather than an intrinsic toxic property of the PT-3 GT vector. Thus, these clones were not included further in the results. Among the remaining 28 clones, 3 (11%) had restricted pattern of reporter gene expression at 8.5 dpc, 9 (32%) were ubiquitously expressed, 12 (43%) showed no expression and 4 (14%) showed no contribution of the ES cells to embryonic tissues (Table 6A).

Expression in chimeric embryos made with the three GT insertions displaying restricted expression pattern of the *lacZ* reporter are described in more details (Table 6B). No strong correlation could be found between the expression patterns observed *in vitro* and *in vivo* (Table 6A). Two of the GT insertions with restricted expression patterns also expressed the reporter in ES cells. Moreover, at selection time, cultures of the H201 clone contained a proportion of differentiated cells large enough to have it mistaken for a clone expressing *lacZ* in ES cells. Therefore pre-screening ES cells after differentiation *in vitro* did not enrich significantly the screen for clones with restricted patterns of reporter gene expression *in vivo* at 8.5 dpc. Differentiating small population of ES cells in 96 micro-titer dish is a long procedure during which many clones could not be completely tested because they did not grow or were contaminated by bacterial or yeast infections. The number of clones showing regulated expression during the pre-screen was also low, compared to the clones expressing the reporter in ES cells immediately after G418 selection. It was then decided that a more efficient and more productive screen would simply involve the selection of all the $G418^R$ clones showing some expression of the reporter gene in undifferentiated ES cell clones.

Large screen

Isolation of ß-galactosidase-expressing ES cell clones

The vector PT-1 (Fig. 3A and Materials and Methods) was used to generate all ß-Gal expressing ES cell clones. Clones containing vector DNA were isolated by selection

TABLE 6.

A. Expression in 8.5 dpc chimeric embryos for clones tested *in vitro* and *in vivo* [a]

Expression *in vivo*	*In vitro* screening classification [b] blue-blue (9)[c]	blue-white (7)[c]	white-blue (2)[c]	white-few blue (10)[c]
Restricted expression (3)[c]	1 (5DC6)	1 (C101)	1 (H201)	0
Ubiquitous expression (9)[c]	6	3	0	0
No expression (12)[c]	2	2	1	7
No contribution (4)[c]	0	1	0	3

B. Description of restricted expression patterns in 8.5 dpc chimeric embryos

Clones	E/eE	Expression in embryos' whole mounts and paraffin sections
5DC6 (B-B)[d]	eE	-7 chimeras- PT-1 vector Expression is restricted to the yolk sack; the tree-like distribution of ß-Gal positive cells is characteristic of the endoderm layer lineage. This interpretation was confirmed by analysis of histological sections (data not shown).
C101 (B-W)[d]	E	-17 chimeras- pGT4.5a vector Expression marks the entire axis of the embryo from the head to the tail bud. On sections, ß-Gal is found in the roof of the gut, notochord and floor plate.
H201 (W-B)[d]	E+eE T [e]	-17 chimeras- PT-1 vector Expression marks strongly the head ectoderm and seems ubiquitous after overnight staining; ß-Gal was also detected in the allantois and yolk sac. On sections, ß-Gal is only found in the ectoderm and mesoderm layers.

a 19 cell lines from PT-3 were not included (they gave an abnormally low contribution of ES cells clones to chimeras).

b 13 from pGT4.5a and 15 from PT-1; (total No. tested *in vivo*).

c (Total No.).

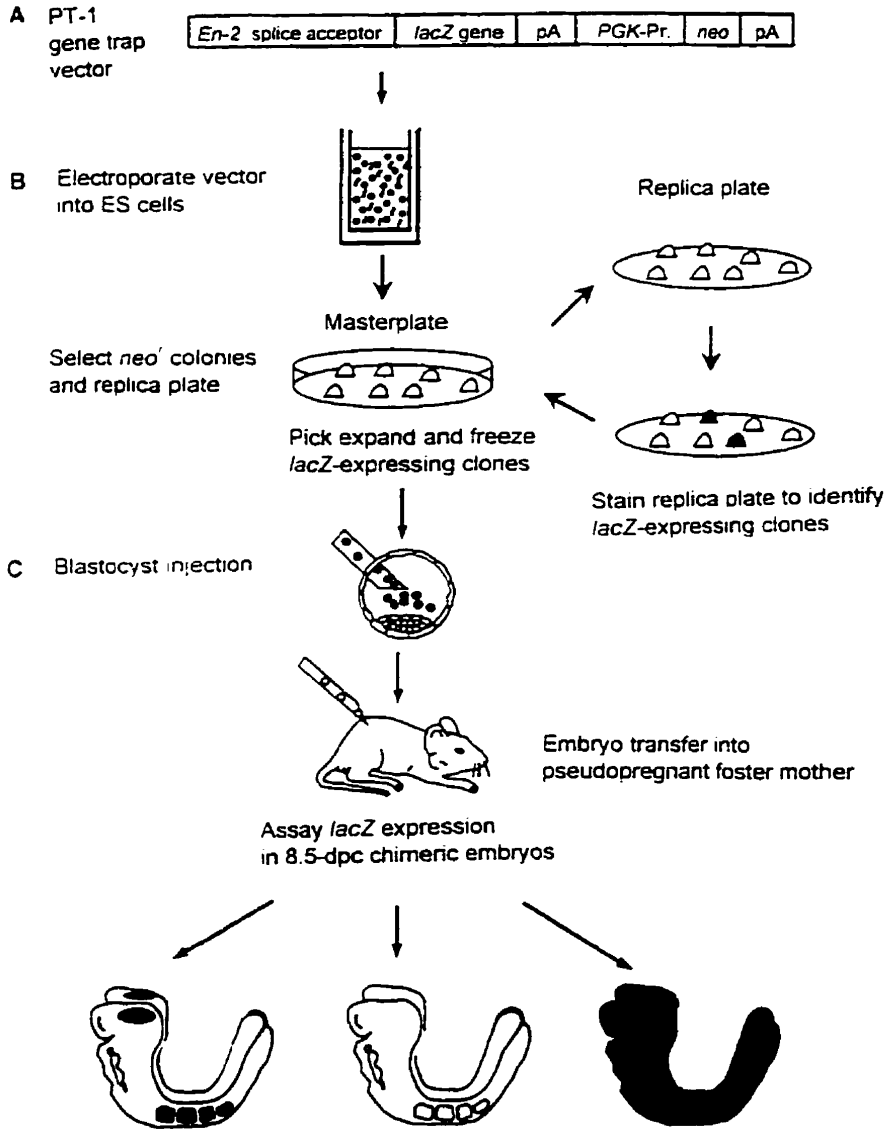
d Expression *in vitro* before, and after differentiation (B=blue; W=white).

e Also analyzed in tetraploid chimeric embryos.

Abbreviations: B: ß-Gal expression (blue); E: embryonic; eE: extra-embryonic expression; W: no ß-Gal expression (white).

for expression of the *neomycin* resistance gene, driven by the *Pgk-1* promoter ($G418^R$ clones). These $G418^R$ colonies were then replica plated and assayed for ß-Gal activity (Fig. 3 & Gossler *et al.*, 1989). From 38,730 $G418^R$ clones, 393 (~1%) ß-Gal-expressing clones were identified, from these 393 clones, we were able to establish 300 cell lines which were expanded and kept as frozen stocks.

Fig. 3. - Schematic representation of the PT-1 gene trap vector and the screening strategy followed to analyze *ß-galactosidase* expression patterns in 8.5 dpc ES cell chimeras.


A PT-1 gene trap vector
En-2 splice acceptor
lacZ gene
pA
PGK-Pr.
neo
pA
B Electroporate vector into ES cells
Replica plate
Masterplate
Select neo' colonies and replica plate
Pick expand and freeze lacZ-expressing clones
Stain replica plate to identify lacZ-expressing clones
C Blastocyst injection
Embryo transfer into pseudopregnant foster mother
Assay lacZ expression in 8.5-dpc chimeric embryos

Since colonies that were expressing *β-gal* usually contained a mixture of undifferentiated and partially differentiated ES cells. ß-Gal staining was observed in either or both cell types. When 208 selected clones were examined more closely for *β-gal* expression patterns within the partially differentiated colonies. the expression in 13 clones appeared restricted to partially differentiated cells. Only two clones showed *β-gal* expression restricted to undifferentiated ES cells. The remainder of the clones showed ß-Gal activity throughout the colony.

Reporter gene expression in 8.5 dpc chimeras

Two hundred and ninety ES clones expressing the reporter gene were injected into CD1 blastocysts. and chimeric embryos were assayed for *β-gal* expression at the 4-15 somite stage of embryogenesis. Of the 290 clones injected. 279 lines provided enough chimeric embryos to satisfy our criteria for inclusion in this study (see Materials and Methods). The patterns of *β-gal* expression observed at 8.5 dpc could be divided in three classes: restricted. ubiquitous and not detectable (Table 7).

TABLE 7

Table 7: Summary of integrations tested in 8.5 dpc embryos

	Number of ES cell clones			
	Total	Restricted expression	Widespread expression	No expression
Large screen	279	36	88	155
Pilot screen	24	3	9	12
	303	39 (13%)	97 (32%)	167 (55%)

After examining further the chimeras from the 36 lines that showed spatially restricted patterns of expression in the 8.5 dpc embryos. we classified their expression patterns into four categories. A description of the expression pattern for each clone is provided in Table 8.

i) Seventeen clones showed highly restricted tissue-specific expression (Fig. 4). Three clones were only expressing *lacZ* in the embryo proper and 6 in extra-embryonic tissues only (allantois and yolk sac mesoderm). The remaining clones expressed *lacZ* in both. Tissues showing specific expression were the gut. the neural tube. the head mesenchyme or neural crest cells and in particular. the node and notochord precursors in the clone PT-1-19 (Fig. 4A).

TABLE 8

Gene trap *β-galactosidase* expression patterns in chimeric embryos at 8.5 dpc

Clones	Notes[1]	E / eE[2]	Description[3]
Highly restricted tissue specific expression			
PT-1-19	*Fig.4A*	E	Node-specific at early somite stages, and midline cells anterior to node (possibly notochord precursors).
6-13-1		E / eE	Few scattered cells in NT's headfolds. *Allantois and yolk sac.*
6-15-2	*Fig.4B* T	E / eE	Specific staining in portion of ventral pharynx, mid- and hindgut and posterior NT. *Very weak staining in yolk sac.*
9-10	*Fig.4C*	E / eE	Specific longitudinal stripe along the dorsal posterior NT and scattered cells in the head. *Strong staining in yolk sac.*
12-27		E / eE	Strong staining in groups of cells lateral to the hindbrain (possibly neural crest); scattered cells staining at posterior and along NT. *Weak staining in yolk sac.*
12-50		E	Bilateral stream of cells between metencephalon and otic vesicle (possibly neural crest or paraxial mesoderm derived cells).
12-52	T	E / eE	Strong in the posterior NT; scattered cells in heart and head. *Yolk sac.*
14-49		E	Expression in scattered cells in the heart, around the otic vesicles and in the branchial arches (possibly neural crest).
14-50	*Fig.4D*	E / eE	Specific dorsal hindgut staining. *Allantois and yolk sac.*
PT-1-1		eE	*Few cells in allantois and yolk sac*
13-76		E / eE	Weak widespread staining in embryo. *Strong in allantois and yolk sac*
PT-1-14	*Fig.4E*	eE	*Yolk sac mesoderm (probably in blood islands)*
6-15-1	*Fig.4F*	eE	*Strong staining in allantois; also in yolk sac mesoderm.*
7-5-2		eE	*Yolk sac only.*
13-11		eE	*Staining at base of allantois and in yolk sac mesoderm.*
14-57		eE	*Yolk sac mesoderm.*
6-9-1		E / eE	Few scattered cells in embryo. *Groups of cells in allantois and yolk sac*
Highly Restricted Region-Specific			
6-16-3	*Fig.5A*	E / eE	Stronger expression at anterior and posterior ends of the embryo with graded reduction towards the middle. *Strong staining in yolk sac endoderm.*
8-7-1	*Fig.5B* T	E / eE	Stronger staining in the dorsal anterior and posterior regions of the embryo with graded reduction towards middle. *Weak staining in allantois and yolk sac.*
9-3	*Fig.5C*	E / eE	Widespread low level expression with stronger staining in anterior neural folds and in somites. *Weak staining in allantois.*

(*continued* -)

TABLE 8 (*continued*)

Gene trap *β-galactosidase* expression patterns in chimeric embryos at 8.5 dpc

Clones	Notes[1]	E / eE[2]	Description[3]
Tissue-Specific Plus Widespread Low-Level			
PT-1-13	δ	E eE	Stronger staining in heart, skin and along the midline of the NT and hindbrain; scattered cells in the mandible. *Yolk sac*
7-9-3	*Fig.6A*	E eE	At early stages, widespread staining probably in mesoderm; later, strongest staining in the somites and the midline of NT. *Weak staining in the yolk sac*
9-7	*Fig.6B*	E eE	Strongest staining in the somites, heart and hindbrain. *Weaker staining in the yolk sac*
9-9		E eE	Strongest staining in the somites and dorsal NT. *Yolk sac*
9-12		E	Strongest in the ventral CNS
11-33		E eE	Strongest staining in the heart. *Strong staining in the yolk sac*
12-2		E eE	Stronger in the posterior end of embryo and in a segment (rhombomere?) of the hindbrain. *Weak yolk sac staining*
13-15	δ	E eE	Widespread expression with the exception of the NT (may be mesoderm specific). *Yolk sac*
13-17		E	Stronger anteriorly in head and heart with weak widespread staining.
13-48		E	Strongest in the head (mainly CNS).
14-59		E eE	Strongest in the anterior end of embryo. *Low in yolk sac*
Region-Specific Patterns Plus Widespread Low-Level			
PT-1-7		E eE	Weak widespread staining with stronger marking of four longitudinal stripes along the entire NT. *Low in yolk sac*
5-8-1	*Fig.7C* T	E	Weak widespread expression with strongest staining in mid- and hindbrain and a sharp segment of unstained cells across the hindbrain.
9-4	*Fig.7B*	E	Widespread early weak staining but later stages show stronger marking of two segments across the hindbrain.
13-28		E eE	Weak widespread staining with stronger staining in spinal cord and hindbrain; expression is the strongest in segment at the anterior boundary of expression in the hindbrain. *Weak allantois marking.*
13-31	*Fig.7A* δ	E eE	Broad weak staining with stronger staining in the midline of the mid- and forebrain and in the heart; strongest specific staining in a segment of cells in the hindbrain (at the level of the otic vesicles)with a weaker anterior segment also detectable. *Weak yolk sac.*

[1] T: tested by aggregation with tetraploid embryos; δ: only differentiated ES cells expressed β-Gal *in vitro*. Reference to figure number is indicated when applicable.

[2] Embryonic (E), extraembryonic (eE) or both.

[3] Extraembryonic expression is indicated in *italics*. NT = neural tube; CNS = central nervous system.

Fig. 4. - E 8.5 ES cell chimeric embryos showing tissue specific expression.

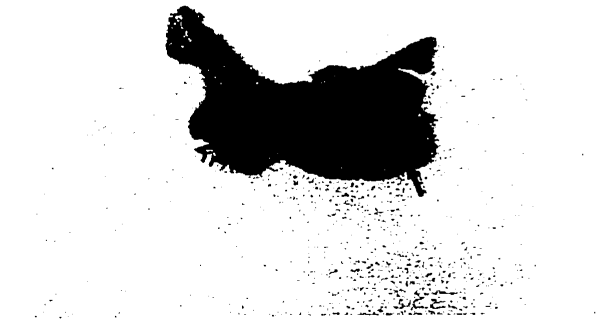
Reporter gene expression was detected in the embryo proper (A-D) or extraembryonic tissues (E. F). (A) An approximately 8 somite-stage chimeric embryo from ES clone PT-1-19 showing node-specific (arrow) expression. (B) A late 8.5 dpc totally ES cell derived embryo from ES clone 6-15-2 showing pharynx (arrowhead). gut (arrow) and posterior spinal cord (white arrow) expression. (C) A late 8.5 dpc chimeric embryo from clone 9-10 exhibiting longitudinal stripes of ß-galactosidase-expressing cells along the posterior neural tube (arrow). in scattered cells in the head (arrowhead) and in the yolk sac (white arrow). (D) A 9.5 dpc chimeric embryo from ES clone 14-50 showing specific expression in the dorsal hindgut and cloaca (arrow). (E) An early 8.5 dpc chimeric embryo from ES clone PT-1-14 showing expression in the mesoderm layer of the yolk sac. probably in blood islands. (F) A 8.5 dpc embryo from ES clone 6-15-1 showing strong expression in the mesoderm layer of the yolk sac (arrow) and the allantois (white arrow).


B
C
D
E
F

Fig. 5. - E 8.5 ES cell chimeric embryos showing reporter gene expression in graded patterns along the anterior/posterior axis.

ES cell chimeric embryos from lines 6-16-3 (A) and 8-7-1 (B) show strong reporter gene expression at the posterior (white arrow) and anterior (arrow) ends. The embryo in B represents a totally ES cell-derived embryo. (C) An ES cell chimeric embryo from line 9-3 showing stronger staining in the anterior neural folds (white arrow) and in the somites (arrow).

A



B

C

Fig. 6. - E 8.5 ES cell chimeric embryos showing widespread expression with stronger tissue-specific reporter gene expression.

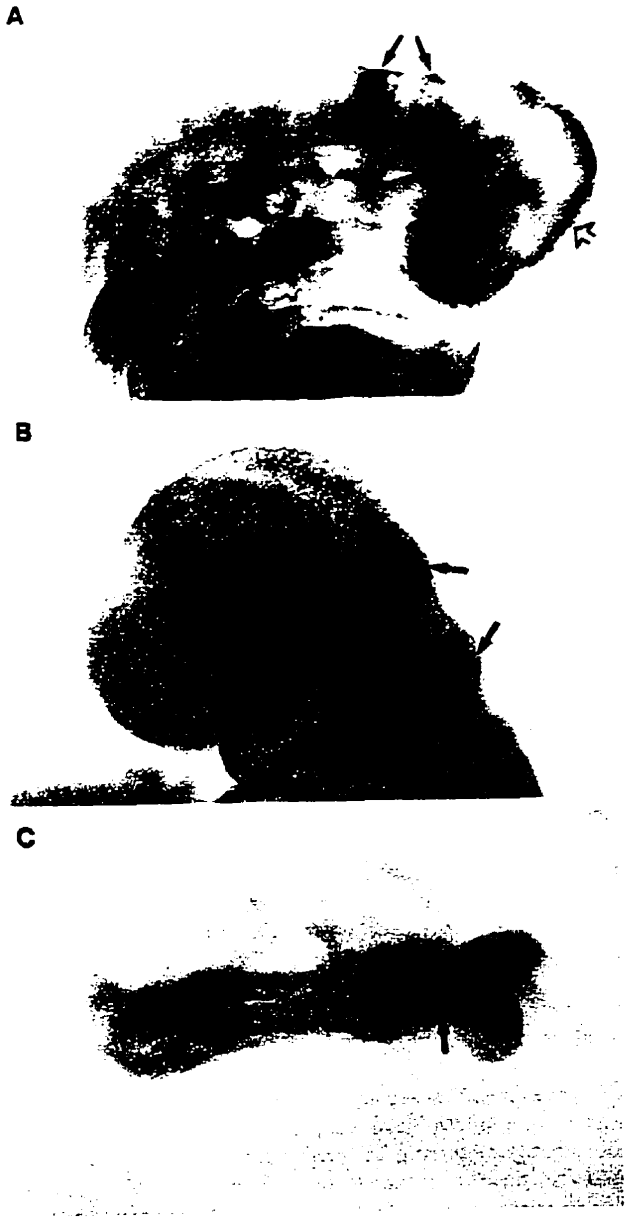
(A) An ES cell chimeric embryo from line 7-9-3 showing widespread staining probably in the mesoderm with stronger staining in the somites (white arrow) and in the dorsal midline of the neural tube (arrow). (B) An ES cell chimeric embryo from line 9-7 showing stronger staining in the somites (arrow), heart (white arrow) and hindbrain.

A

B

Fig. 7. - E 8.5 ES cell chimeric embryos showing widespread low expression with stronger reporter gene expression in spatially defined domains.

(A) An ES cell chimeric embryo from line 13-31 showing stronger staining in the midline of the mid- and forebrain (white arrow) and strong staining in two bands in the hindbrain (arrows). (B) An ES cell chimeric embryo from line 9-4 showing stronger staining in two stripes across the hindbrain (arrows). (C) A totally ES cell-derived chimeric embryo from ES clone 5-8-1 showing stronger staining in the hindbrain (arrow) and the midbrain (to the right).


A
B
C

ii) Three clones showed highly restricted region-specific patterns along the anterior-posterior axis, without any obvious tissue-specificity (Fig. 5). In particular, 2 clones showed strong expression at the anterior and posterior ends of the embryo with a reduction of staining towards the middle of the embryo.

iii) Eleven clones showed tissue-specific patterns plus widespread low-level expression (Fig. 6). The patterns of expression ranged across a large variety of tissue types, including heart (3 lines), somites (3 lines) and CNS (6 lines).

iv) Five clones showed region-specific patterns plus widespread low-level expression (Fig. 7). Specific expression was marking some boundaries in the region of developing rhombomeres (4 clones) or the trunk neural tube (1 clone).

Reporter gene expression in ES-tetraploid chimeras

Ten clones were used to generate aggregation chimeras using tetraploid host embryos. Such chimeras are almost entirely ES-cell-derived (Nagy *et al.*, 1990; Nagy and Rossant, 1993). These clones were selected because the patterns of reporter-gene expression seen in the screen were not clear and needed confirmation. Four clones could generate chimeric embryos. These embryos confirmed the patterns of *lacZ* gene expression that had been observed in the previously described diploid chimeras.

Analysis of 12.5 dpc embryos

A large proportion (55%) of the *lacZ* expressing ES cell clones failed to show reporter gene activity in 8.5 dpc embryos. It is important to determine if these clones did not express the reporter gene at this stage or if they were unable to produce chimeras. for a subset of clones, ES cell contribution to the embryos was analyzed by the presence of eye pigmentation in 12.5 dpc chimeras (Table 9). Ninety-three (78%) clones produced chimeric embryos at this stage. This suggests that a similar proportion 8.5 dpc embryos showing no expression were chimeras but did not express detectable reporter gene activity.

Table 9

ES cells contribution analyzed in 12.5 dpc chimeric Embryos

ß-Galactosidase expression at 8.5 dpc	Total no. of lines analyzed	ES cells contribution[a]	No ES cells contribution[a]	N.d. [b]
+	70	53	10	7
-	69	40	17	12
Total	139	93	27	19

[a] Contribution was scored by the ß-Gal staining and the presence of eye pigments at 12.5 dpc.

[b] No embryos were recovered at 12.5 dpc.

Another important question is to know how many clones would display a distinct expression pattern if tested at another stage. We also assayed reporter-gene expression for 70 clones at 12.5 dpc (Table 10). Thirty clones had a completely different expression pattern at this stage. In particular, clones turned on the expression of the reporter gene between 8.5 and 12.5 dpc (10 of 26 clones) or restricted its expression to fewer tissues (5 of 28 clones).

Table 10

ß-Galactosidase expression in 12.5 dpc chimeric embryos

ß-Galactosidase expression at 8.5 dpc	No. of embryos examined	Expression at 12.5 dpc: Widespread expression	Restricted expression	No expression
Widespread	28	21	5	2
Embryonic Pattern	14	8	3	3
Extraembryonic only	2	2	0	0
No expression	26	8	2	0

DISCUSSION

We assayed 303 clones, containing potentially mutagenic integrations of a gene-trap vector, for expression of the *lacZ* reporter gene in 8.5 dpc chimeric embryos.

Twenty-eight clones were analyzed in a pilot screen. Integrations with 3 slightly different gene trap vectors (see Material & methods) were analyzed. Clones selected for insertions of the PT-3 vector did not form chimeras efficiently and this vector was used further. It is not clear whether this was due to the vector or to the batch of ES cells used for these 2 electroporations. The PT-1 and pGT4.5a vectors produced similar results. However, the highest number of $G418^R$ clones was obtained with the PT-1 vector (where the *Pgk-1* promoter drives *neo* expression), and this GT vector was selected for future experiments. The pilot screen was also designed to test a simple strategy for efficiently pre-selecting *in vitro*, ES cell clones expected to display restricted expression pattern *in vivo* in the gastrulating embryo. ES cells are able to differentiate *in vitro* into a variety of cell types that are similar to cell types found *in vivo* during early embryo development (Doetschman *et al.*, 1985; and reviewed in Baker and Lyons, 1996). We scored expression of the reporter gene in $G418^R$ ES cell clones after spontaneous differentiation into fibroblast-like cells or into embryoid bodies. Twenty-eight clones showing regulated expression of *lacZ in vitro* were assayed for expression in 8.5 dpc chimeric embryos.

About half of the insertions did not give detectable levels of reporter gene expression at 8.5 or 12.5 dpc. One third of the trapped genes were expressed ubiquitously in the embryos. Three genes displaying restricted expression in extra-embryonic tissues or in the embryo proper were identified. However, the pre-screen was not efficient and not informative enough. Most of the insertions selected *in vitro* already expressed lacZ in undifferentiated ES cells (Table 4). Only 1 clone (out of 12 tested) that did not express the reporter gene in undifferentiated ES cells was expressed *in vivo*. Furthermore, the type of expression observed *in vitro* did not allow us to predict which cell lines would display a restricted expression pattern *in vivo*. This strategy was not followed further and it was decided to test, in a larger screen, insertions into genes that are expressed in ES cells. In a larger screen, we analyzed *in vivo* the expression pattern of 279 insertions. The larger screen yielded similar results (ubiquitous, restricted or no expression *in vivo*), in the absence of clone pre-selection *in vitro*.

In total, we identified 146 endogenous cellular transcription units expressed at 8.5 or 12.5 dpc (Table 7 & 10). The lack of expression in half of the embryos was not due to inability of the ES cells to contribute to chimeras, since for most of these clones the presence of ES-derived cells could be scored by the presence of eye pigmentation when the embryos were allowed to develop to 12.5 days (Table 9). We can estimate that only approximately 10% of the ES cell clones that underwent the screening procedure were not able to contribute to chimeric embryos.

A wide range of developmental patterns was observed (Table 6 B & 8). Since this screen was limited to genes that are expressed in ES cells, the question of whether the frequency of these classes of gene expression patterns reflects the frequency of the types of genes that are developmentally regulated in the embryo remains unresolved. It is clear that some developmentally restricted genes such as *En-2* and *Hox 1.3* are expressed in undifferentiated ES cells, while others such as *wnt-1* and *En-1* are not (Joyner *et al.*, 1985; Joyner and Martin, 1987; McMahon and Bradley, 1990; Jeannotte *et al.*, 1991). A more laborious but less biased screen would involve the analysis of all clones that had incorporated the vector into their genome. A large-scale screen of this type would be very difficult because many integrations would fall outside of transcription units and, due to the nature of the vector, even those that were within genes could only be expressed if integrated in the correct orientation and reading frame. Other vectors may be better designed for such a systematic approach (see Discussion in Chapter V).

Many of the reporter-gene expression patterns that we observe are consistent with the behavior of ES cells in culture. Spontaneous differentiation of the D3 ES cell line has been shown to give rise to a number of identifiable cell types, including yolk sac-like structures that contain blood islands and primitive blood vessels as well as cardiac muscle cells (Doetschman *et al.*, 1985). Of the 39 clones that displayed restricted expression, 26 clones showed expression in the yolk sac (Table 6 B & 8); in two of the lines expression may be restricted to blood islands (PT-1-14 and 6-9-1). Expression in the developing heart was observed in 7 clones. Since in our pre-screen colonies were allowed to partially differentiate, we may have a bias toward genes that are expressed in tissues that are generated during *in vitro* differentiation. The three clones that were expressed exclusively in differentiated cells (H201, 13-31 and PT-1-13) and displayed regulated expression of the reporter gene showed expression in both the yolk sac and the heart.

Central nervous system expression was predominant in 19 out of 39 lines. This

result is also not surprising since at 8.5 dpc the CNS is undergoing active growth and organization with respect to dorsal/ventral and anterior/posterior polarity as well as with respect to segmental identity. A large number of genes responsible for these events might be expected to be active in the early embryo and ES cells and thus be identified in our screen.

The results of this gene-trap screen demonstrate that it is possible to identify a wide assortment of genes, showing tissue-specific and spatially-restricted expression during development by expression in chimeras, even when the analysis was limited to genes expressed in ES cells and during one developmental stage, at 8.5 dpc. This time was chosen since it is the time when the basic body plan of the embryo is being established. However, other time windows could be added. Limited analysis at 12.5 dpc indicates that if the time window of expression analysis could be broadened, more genes with restricted expression patterns could be identified. However, for each stage analyzed it would be necessary to generate new chimeric embryos.

Insertions with the most interesting developmental expression patterns can then be selected for further analysis at the molecular level and can be transmitted through the germline for phenotypic analysis of the mutation. One ES cell line (C101) isolated during the pilot screen and displaying a very specific pattern of expression *in vivo* was selected for further analysis and will be described in the next chapter.

- CHAPTER III -

CORDON-BLEU, A NOVEL MURINE DEVELOPMENTAL GENE.

This chapter includes results previously published in the following paper:

Characterization Of A Gene Trap Insertion Into A Novel Gene, *Cordon-Bleu*, Expressed In Axial Structures Of The Gastrulating Mouse Embryo.

Stéphan Gasca, David P. Hill, John Klingensmith and Janet Rossant.

Developmental Genetics (1995), **117**, 141-154.

CHAPTER THREE : *CORDON-BLEU*, A NOVEL MURINE DEVELOPMENTAL GENE.

INTRODUCTION

Embryonic axes are fully established in vertebrate embryos at gastrulation. As the definitive germ layers that will form the embryo proper are emerging, they are simultaneously organized around the anterior-posterior and dorsal-ventral axes. Analysis of the genes specifically expressed in axial structures around the time of gastrulation is beginning to identify genes that are involved in axis formation. However, the full spectrum of genes involved in axial specification and their modes of interaction are not yet clear.

Most information has come from study of genes first identified as important by their molecular nature and pattern of expression (reviewed by Beddington and Smith, 1993; Faust and Magnuson, 1993; De Robertis *et al.*, 1994; Boncinelli and Mallamaci, 1995; Lemaire and Kodjabachian, 1996). However, the direct genetic approach of identifying mutations that affect gastrulation has also provided several important candidates such as *Brachyury* (*T*) (Herrmann *et al.*, 1990), *nodal* (Zhou *et al.*, 1993; Conlon *et al.*, 1994) and *embryonic ectoderm development* (eed, Niswander *et al.*, 1988; Faust *et al.*, 1995; Schumacher *et al.*, 1996). Although large-scale mutation screens are practically difficult in the mouse, the application of the different trapping strategies to the ES cells system provided developmental biologists with novel approaches to perform genetic screens aimed at the random identification and mutagenesis of novel genes.

The gene trap screening strategy based on the expression of the *lacZ* reporter gene in ES cells *in vitro* and in ES cell/embryo chimeras *in vivo*, allowed us to screen through a large collection of ES cell clones carrying GT insertions, for genes showing spatially and temporally restricted patterns of expression in the late gastrulating embryo at E8.5 (Chapter II & Wurst *et al.*, 1995). From this screen we identified a number of genes with restricted expression suggesting that they could be playing a role in embryonic patterning. Genes expressed in segments in the hindbrain region (clones 5.8.1, 9-4 & 13-31) reflect the establishment of specific compartments along the antero-posterior axis; in others, stronger CNS expression in the ventral region (clone 9-12) or in longitudinal stripes (clones 9-10 & PT-1-7) could indicate patterning events along the dorso-ventral and medio-lateral axes. In particular, two genes (clones C101 & PT-1-19) showed an expression

pattern restricted to those structures thought to play a role in axial patterning, namely the node and the notochord. Genes known to be important for axis formation or patterning such as *HNF3-β* and *Shh* are also expressed in these tissues during the early stages of their development (Ang *et al.*, 1993; Ruiz i Altaba *et al.*, 1993b; Sasaki and Hogan, 1993; Echelard *et al.*, 1993; Marti *et al.*, 1995). The insertion PT-1-19 was only weakly expressed in the node and anterior notochord precursors at early somite stages, whereas the C101 insertion was strongly expressed in axial tissues from the 3 germ layers, with a pattern reminiscent of that of *HNF3-β* and *Shh*.

In the following chapter, I will describe the characterization of the clone C101, identified during the pilot screen (Chapter II), which carries a GT insertion into a novel gene expressed in the node, notochord, floor plate and the roof of the gut. The identification of this gene, *cordon-bleu*, provides another possible component of the genetic hierarchy of genes involved in axial patterning during early mouse embryogenesis.

MATERIALS & METHODS

Production of the C101 ES cell clone

Electroporation of the linearized pGT4.5a GT vector and selection of G418 resistant ($G418^R$) ES cell lines was performed as described earlier (Gossler *et al.*, 1989). 100 µg of vector linearized at the unique *HindIII* site were directly (without precipitation step) electroporated into $5x10^7$ male D3 ES cells (Doetschman *et al.*, 1985) at a final concentration of 2 mg/ml. $G418^R$ ES cell lines were then individually picked, expanded as single clones and tested for *lacZ* expression *in vitro* and *in vivo*.

Production of ES cell chimeras and breeding

About 30 CD1 blastocysts were injected with 15 to 30 C101 ES cells each and reintroduced into pseudo-pregnant CD1 females. Females were dissected at embryonic day 8.5 (E8.5) to assay for ß-Gal activity in chimeric embryos. To produce a germline chimera, C57BL/6 blastocysts were injected with C101 ES cells, reintroduced into pseudo-pregnant CD1 females, and embryos were allowed to develop to term. Males with strong contribution from ES cells as judged by coat color were tested for contribution to the germline by crossing with C57BL/6 females. Inbred mice heterozygous for the insertion were generated by crossing with 129/Sv females. Offspring carrying the insertion could be identified by *lacZ* staining of tissue samples from ear punches used to number the mice. Their genotype was confirmed by Southern blot analysis. Mice homozygous for the insertion were produced by intercrosses between heterozygous parents. Their genotype was determined by Southern blot analysis on genomic DNA digested with *BamHI*, *BglII* and *PvuII* and hybridized with the *En-2* sequence present on the vector (see below), which allows comparison of the relative intensity of the bands between the endogenous and the insertion loci. The genotype of males was also confirmed by breeding with wild-type CD1 mice, and staining embryos for ß-galactosidase activity.

Ho23 ES cell line

Three 129/Sv *GtC101/+* superovulated females crossed with three 129/Sv *GtC101/+* males gave 43 blastocysts from which we recovered three new ES cell lines (according to morphology, seeNagy *et al.*, 1993). Two were heterozygous for

the insertion. One, ho-23, was homozygous at the insertion locus, as judged by relative intensity of the En-2 bands on Southern blots (see below). Ho-23 was also tested by injection into blastocysts and was able to participate in normal embryonic development.

ß-Gal staining of whole-mount embryos

Embryos were staged as embryonic day 0.5 (E0.5) on the morning the vaginal plug was detected. Embryos were dissected from the uterus in PBS and then fixed in 0.2% gluteraldehyde (Fisher), 2 mM $MgCl_2$, 5 mM EGTA, 0.1 M sodium phosphate at room temperature for 15 - 60 minutes and then washed at least twice in 0.1 M sodium phosphate, 2 mM $MgCl_2$, at room temperature. They were then stained in the dark in 1 mg/ml X-gal (Vector Biosystems), 5 mM $K_3Fe(CN)_6$, 5 mM $K_4Fe(CN)_6$, 2 mM $MgCl_2$ and 0.1M sodium phosphate at 37°C overnight. Before processing and paraffin embedding for sectioning, embryos were post-fixed in fresh 4% paraformaldehyde, 0.2% gluteraldehyde in PBS overnight at 4°C.

Cloning and analysis of endogenous cDNAs

Total RNA for RT-PCR was prepared from ES cells and from embryos via the method of Chomczynski (1987). Poly A+ RNA was purified by extraction from total RNA with oligo (dT) cellulose (Ausubel *et al.*, 1989). The RACE strategy (Frohman *et al.*, 1988) was adapted to the following procedure: the Copy Kit TM (Invitrogen) was used, following the manufacturer's instructions, to synthesize blunt ended double stranded cDNA from 0.5µg of polyA+ RNA from ho-23 ES cells annealed with 10 nanograms (ng) of the *lacZ* primer #1 (5'-GCAAGGCGATTAAGTTGGGT-3') in the presence of MeHgOH. The synthesis yielded 70 ng of double stranded cDNA. 3.5 ng were incubated with 2 pmoles of UNI-Amp Eco RI adapters (Clontech) and T4 ligase (Boehringer Mannheim). The ligation products were size selected on a SP400 column (Pharmacia) and 1/10 were then subjected to 40 cycles of PCR (94°C for 60 sec; 60°C for 45 sec; 72°C for 2 min.) and a final extension at 72°C for 10 min., using 2.5 units of Taq polymerase in 1X Taq buffer (Perkin Elmer Cetus) in the presence of 100 ng of UNI-Amp Eco RI primers (Clontech) and 500 ng of primer #256 (Fig. 11 and Skarnes *et al.*, 1992). The size range of the amplified cDNAs was visualized by Southern blot analysis with the *En-2* probe (see below). The PCR products were size-selected by gel electrophoresis and fragments larger than 300 bp were digested by Eco RI and Kpn I, sub-cloned into the pBluescript II KS(+) vector (Stratagene) and sequenced with the Sequenase kit (U.S. Biochemicals). The

sequence data were managed with the help of the Program Manual for the Wisconsin Package, Version 8, September 1994, Genetics Computer Group. The cDNAs were compared to the database sequences with the FASTA (Pearson and Lipman, 1988) and the BLAST (Altschul *et al.*, 1990) comparison algorithms. The 279 bp C101/7 fragment (lacking *En-2* sequences), was amplified by 40 cycles of PCR (94°C for 60 sec; 62°C for 60 sec; 72°C for 60 sec) and a final extension at 72°C for 2 min., using 0.1 unit of Taq polymerase in 1X Taq buffer (Perkin Elmer Cetus) in the presence of 250 ng of C101 primers #25 (5'-GCGAAGTCAGCATCTGGAGG-3') and C101 primer #283 (5'-TGGTGGTGGCTGCTGTGGTG-3'), and sub-cloned into the vector pSL301 (Invitrogen).

DNA and RNA probes

The *lacZ* probe is the entire 3.1 kb Bam HI fragment purified from the pGT4.5a vector. The *Engrailed-2* (*En-2*) probe is a 240 bp Hind III-Sst I fragment from the *En-2* genomic DNA clone S15 (kindly provided by A. Joyner) which includes 82 bp of the *En-2* intron 1 and 156 bp of the *En-2* exon 2 which are fused to *lacZ* in the GT vector. This probe can bind equally to the endogenous *En-2* genomic sequence and to the same sequence in the inserted vector allowing comparison of the allele number at the two loci (when total genomic DNA is digested with Bam HI + Bgl II + Pvu II, the two bands are of comparable size: 1.5 and 2.0 Kb). The C101/7 PCR fragment subcloned into pSL301 vector Eco RV site, was excised by digestion with Eco RI and Xba I. The b3 probe is the entire 1.2 kb b3 cDNA subcloned into pBluescript II KS(+) (Stratagene) at the EcoRI site. All the DNA probes were purified from the vectors by electrophoresis and were radiolabelled with (32-P)-dCTP by the random priming method of Feinberg and Vogelstein (1983). Sense and antisense C101/7 and b3 riboprobes were synthesized from the clones linearized respectively at the Bgl II and Xba I sites for C101/7 and at the Eco RV and Xba I sites for b3. The transcription was performed from 1 µg of template DNA. The C101/7 probe for sectioned in situ hybridization was synthesized in the presence of 10 µCi of (35-S)-UTP as described previously (Hui and Joyner, 1993). The transcription of the b3 probe for whole mount in situ hybridization was performed in the presence of digoxigenin-labeled UTP (Boehringer Mannheim) as described previously (Conlon and Rossant, 1992).

Southern and Northern blot analysis

Genomic DNA was purified from proteinase K-digested ES cell pellets and

mouse tail biopsies by phenol-chloroform extraction. Total RNA was isolated from ES cells or embryos by the method of Chomczynski (1987). Electrophoresis was carried out in 1X TAE for genomic DNA and in 1X MOPS for RNA (Sambrook *et al.*, 1989). Both Southern and Northern blots used GeneScreen filters (Dupont). Hybridization with radiolabelled probes, for 18 hours at 63°C, was in 0.5 M Na_2HPO_4/NaH_2PO_4, pH 7.4, 7% SDS, 1 mM EDTA for Southern blots and in 5X SSC, 5X Denhart, 0.1% SDS, 100 mg/ml sheared salmon sperm DNA, 5% dextran sulfate for Northern blots. Filters were exposed against reflection ™NEF-film (Dupont-NEN).

Mapping

The BSS DNA panel from the Jackson Laboratory community resource [(C57BL/6JEi X SPRET/Ei) F1 female X SPRET/Ei male] interspecific backcross (Rowe *et al.*, 1994) was used to map the trapped locus with the C101/7 probe. Genomic DNA from C57BL/6 and *Mus spretus* was digested with various enzymes and the probe C101/7 was used on Southern blots to detect restriction fragment length polymorphisms (RFLPs) between the two strains. When DNA was digested with *XbaI* the C101/7 probe detected a RFLP. Genomic DNA from the BSS panel (from 94 N2 backcrosses and the 2 parental strains) was digested by *XbaI*, separated by electrophoresis and transferred on nylon GeneScreen membranes (Dupont). Filters were hybridized with the probe C101/7 and exposed against film. Each sample was scored for the presence of the *Mus spretus* allele and the results were transmitted back to the database of the Jackson Laboratory for comparison with the existing map. The new locus carrying this GT insertion was named *GtC101*.

cDNA library screen

An unamplified λgt11 random- and oligo(dT)-primed CD1 adult brain cDNA library (Clontech; clones bx and Ex) was first screened with the C101/7 probe. 1.5×10^6 phage plaques were transferred to nitrocellulose filters (Schleicher and Schuell) and hybridized under conditions of medium stringency in 5X SSC, 5X Denhart, 0.1% SDS, 100 mg/ml sheared salmon sperm DNA, 10% dextran sulfate, for 48 hours at 56°C with 7×10^5 cpm/ml of radiolabelled C101/7 probe. Filters were washed at 20°C: twice for ten minutes in 2X SSC and one hour in 1X SSC containing 0.1% SDS. After autoradiography, five different clones were plaque purified and their inserts were sub-cloned into the EcoRI site of the pBluescript KS + (Stratagene) plasmid vector. Double-stranded DNA sequencing was performed following the A.L.F. Sequencing protocol (Pharmacia). These cDNAs were then used

to isolate overlapping additional clones covering the 5' and 3' end of the full length cDNA. The 5' most clone was obtained from a λ–ZAPII random- and oligo(dT)-primed CD1 newborn brain cDNA library (Stratagene; clones NB5 & NB*x*)

In situ hybridization

CD1 embryos (Charles River, Montreal) were dissected from the uterus at E7.5, E8.5 and E9.5. The whole mount in situ hybridization with sense and anti-sense b3 riboprobes was performed as described (Conlon and Rossant, 1992; Conlon and Herrmann, 1993). The in situ hybridization of sections was performed as described (Hui and Joyner, 1993), and exposed for 3 and 4 weeks before development.

RESULTS

The C101 gene trap ES cell line

The C101 ES cell clone carrying the pGT4.5a vector was isolated from the pilot screen (described in Chapter II) and presented a restricted pattern of expression both *in vitro* and *in vivo*. C101 ES cells in culture strongly expressed the *lacZ* reporter gene, detectable after 1.5 h of incubation at 37°C (many ES cell clones only showed detectable expression after an overnight incubation period). Most cells differentiating *in vitro* did not express the fusion protein, suggesting that the promoter may be regulated during early development of the embryo. Southern blot analysis, using the *lacZ* and *En-2* probes, showed that this cell line contained a single copy insertion of the vector (data not shown).

Generation of chimeras

Upon dissection at embryonic day 8.5 (E8.5), 11/30 C101-ES/CD1 chimeric embryos displayed identical restricted patterns of expression after staining for ß-Gal activity: expression was limited to axial structures of the embryo from the tail bud to the foregut and midbrain. Histological sections from these embryos revealed that expression was in the roof of the gut, the notochord and the floor plate of the neural tube.

Table 11

Frequencies of transmission of the C101 insertion to the F2 generation of inbred 129/Sv heterozygous mice

Genotype	No.	%
Wild-type	21	24.7
Heterozygous	44	51.8
Homozygous	20	23.5
Total	85	100.0

The insertion was transmitted to the germline in order to analyze in more detail its expression in transgenic embryos and to identify any possible mutant phenotype. One chimeric male was obtained that transmitted the ES cell genotype to 60% of its offspring as determined by coat color transmission. Inbred mice heterozygous for the insertion were generated by crossing this male with 129/Sv

females. No dominant phenotype has been observed in heterozygous mice and they were recovered with the expected frequency for a single gene transmission (Table 11).

Expression of the lacZ reporter gene in heterozygous embryos

129/Sv *GtC101/+* males were crossed to CD1 females to assay for *lacZ* expression in heterozygous embryos. The pattern of expression was consistent with that previously observed in chimeric embryos. No expression was detected prior to E7.5, including at preimplantation stages, although the gene was expressed in ES cells. Expression was first detected at E7.5 at the late streak stage, when the mesoderm layer is complete, in a small group of cells at the distal tip of the embryo in the presumptive node region (Fig. 8 A, B). Expression in the node was clearly restricted to mesodermal cells initially (Fig. 9 A) but extended to underlying axial endoderm as the node develops (data not shown). Expression of the transgene extended anteriorly in the midline (Fig. 8 C) as the head-process progresses. Some positive cells were also present just posterior to the node and in more lateral regions (Fig. 8 C). At the head-fold stage, before the head-process reaches its most anterior position, positive cells began to appear along the axis posterior to the node and into the hindgut. At early somite stages, anterior expression marked the developing foregut and notochordal plate and extended posteriorly to the regressing node (Fig. 8 D). At the 5 to 10 somite stages the notochord was ß-Gal positive along the entire axis from its anterior limit below the midbrain to the base of the allantois. The most anterior region of the primitive streak also remained ß-Gal positive. Expression was clearly seen in the dorsal roof of the developing gut, in both foregut and hindgut extensions (Fig. 8 D). As the floor plate of the neural tube forms at the somite stage, it too became ß-Gal positive (Fig. 9 B). The anterior limit of expression in the floor plate corresponded to the anterior limit of contact between the neural plate and the foregut. Expression was confined to the midline cells of the midbrain, and did not extend rostrally or laterally into the forebrain as seen at this stage for other floor plate markers like *HNF-3ß* and *Shh* (Ang *et al.*, 1993; Ruiz i Altaba *et al.*, 1993b; Sasaki and Hogan, 1993; Echelard *et al.*, 1993; Roelink *et al.*, 1994). No expression was observed in prechordal plate mesoderm, although expression was detected later from E10.0 in the head mesenchyme in the forebrain

Fig. 8. - Detection of the ß-Gal fusion protein from E7.5 to E10.5 stages of development in embryos heterozygous for the *GtC101* insertion.

In A, B, C, D posterior is to the right. Proximal is at the top in (A, D) and dorsal at the bottom in (E, F). (A) E7.5, late primitive streak, stage of first expression. (B) and (C) ventral views of two older E7.5 embryos. (B) Expression is still localized to the node region (arrow). (C) Positive cells are now present along the axis (arrow) in the head process. A few positive cells are also present in paraxial positions and immediately posterior to the node (white arrows). (D) Pattern of expression in five embryos at consecutive stages of development between E7.5 and E9.5 from left to right. Expression is restricted to the node (n, arrowhead) at E7.5 (neural plate, presomite stage), extends from the anterior of the primitive streak to the foregut at E8.0 (late headfold stage, 3 somites), marks the entire axis from the anterior ventral midbrain to the base of the allantois at E8.5 (before turning, 8 somites), remains restricted to the midline structures from the head to the tail bud at E8.5 (beginning of turning stage, 10 somites) and expands to the liver anlage and the branchial clefts at E9.5 (end of turning stage, 13 somites). (E) Expression in a 25 somite E9.5 embryo. The anterior limit of expression in mesenchyme cells is in the olfactory placode region (black arrowhead), the midbrain (outlined arrowhead) for floor plate cells and the anterior end of the notochord for endodermal cells. Expression is seen in the endoderm layer of the branchial clefts, in the liver and in the somites (arrow). (F) Expression in a 30 somite E10.5 embryo. The expression in the somites is restricted to the center (arrow) of each somitic segment (separated by dashes).

Abbreviations: amniotic cavity (ac), branchial arches (b), exocoelum (Ex), gut (g), head process (hp), heart (he), liver (L), node (N), primitive streak (ps), tail bud (t), ventral midbrain (mb).

Scale bar = 50 µm (A), 100 µm (B,C,E), 150 µm (F) and 200 µm (D).

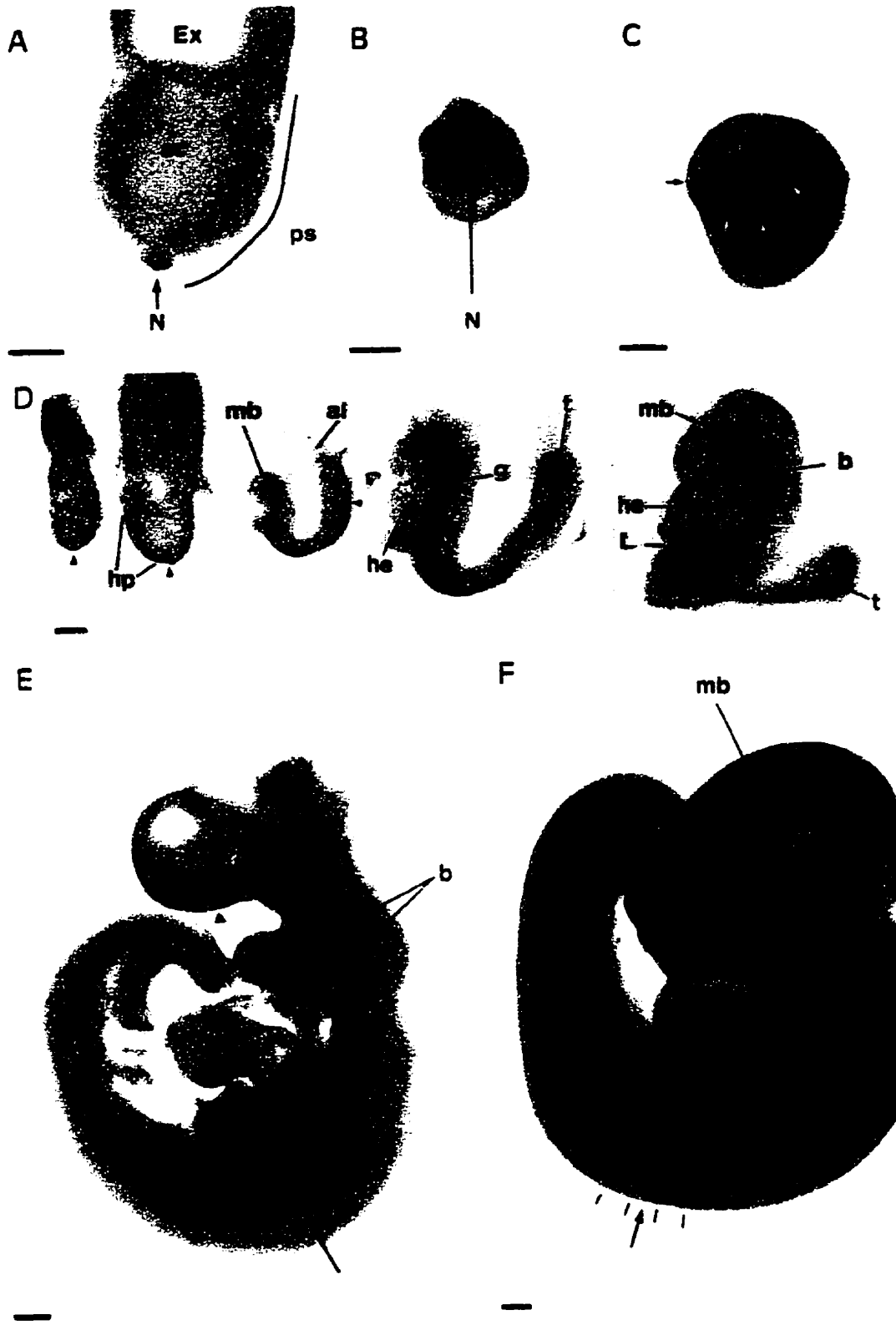
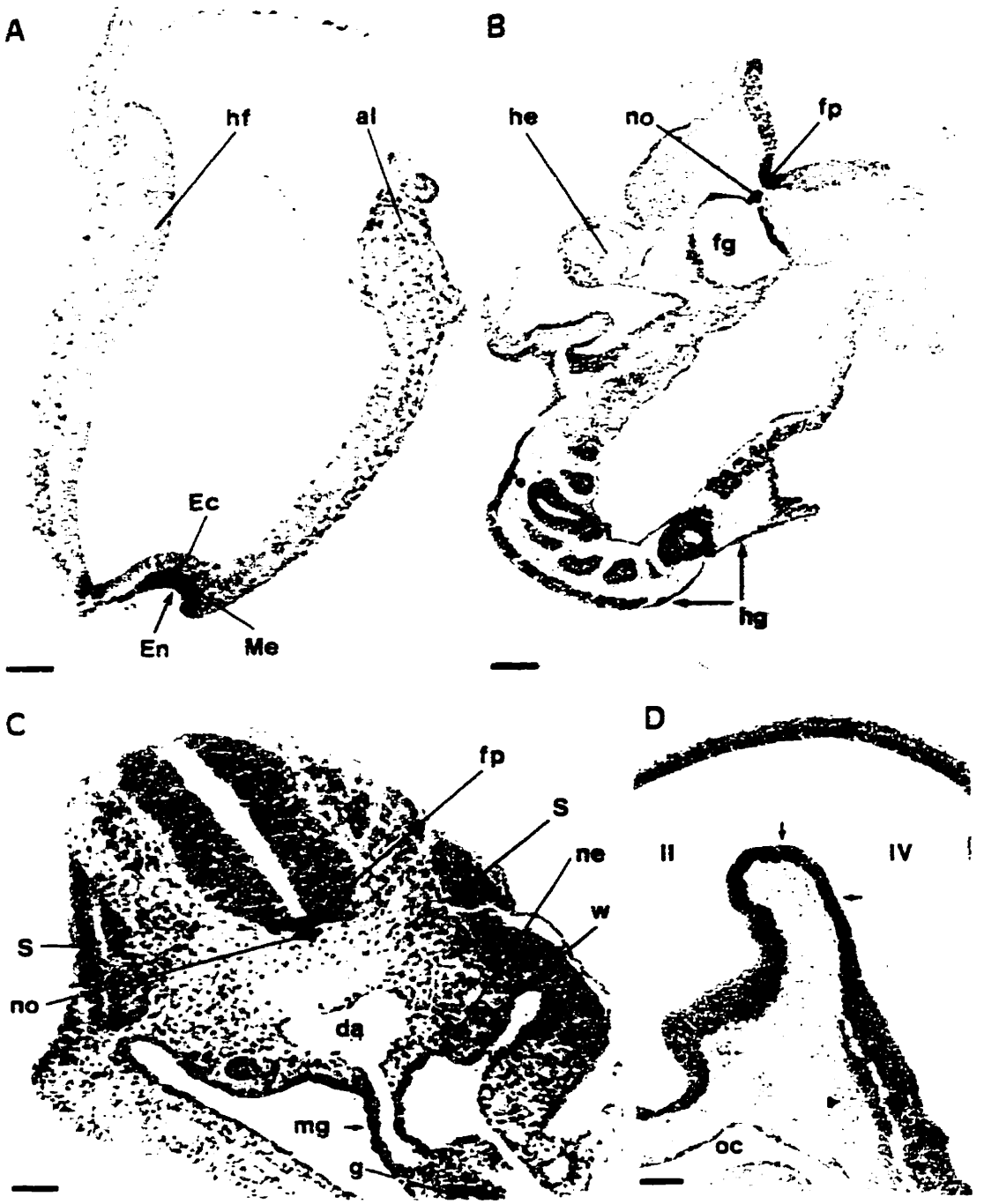

A
Ex
ps
N
B
N
C
D
mb
al
hp
he
mb
b
he
t
E
b
F
mb

Fig. 9. - Histological sections of embryos heterozygous for *GtC101*, stained for ß-Galactosidase activity of the fusion protein.

Sections were stained with eosin. (A) Para-sagittal section of E7.5 embryo showing that positive cells are restricted to the mesoderm layer of the node. Anterior is to the left. (B) The plane of section crosses the axis of an E8.5 embryo three times, through the head (top) and twice through the trunk. Expression is restricted to the floor plate of the neural tube, the notochord and the roof of the foregut and hindgut. (C) section posterior to the forelimb through the middle of the fifteenth somite. Expression is in the floor plate, notochord and surrounding mesenchyme, the gut, the mesogastrium, part of the somites and very weakly in the Wolffian ducts. (D) detail of a mid-saggittal section through the head. Anterior is to the left. Expression is restricted in the posterior midbrain floor plate (between arrows). Some cells of the cephalic mesenchyme are also positive (arrowhead).

Abbreviations: allantois (al), dorsal aorta (da), ectoderm (Ec), endoderm (En), floor plate (fp), foregut (fg), headfold (hf), heart (he), hindgut (hg), mesoderm (Me), mesogastrium (mg), midbrain (mb), nephrogenic cord (ne), notochord (no), oral cavity (oc), somite (S), Wolffian ducts (W), second ventricle (II) and fourth ventricle (IV).

Scale bar = 50 µm (A, C), 100 µm (B, D).


A
hf
al
Ec
En
Me
B
he
no
fp
fg
hg
C
fp
S
ne
w
S
no
da
mg
g
D
II
IV
oc

region. In E9.5 embryos, at the stage of turning, expression was still seen in the floor plate, notochord and the roof of the gut but also occurred in the pharyngeal endoderm, the liver anlage and the tail bud (Fig. 8 D, E). In E10.0 embryos (25 somites) positive cells started to appear in axial head mesenchyme and in trunk sclerotome tissues. Expression was first detected in the somites at this stage (Fig. 8 E) and became strong by E11.5. Positive cells were located in a medial subset of somitic cells, between the sclerotome and myotome compartments (Fig. 8 F and 9 C). However, expression was not observed, at later stages, in muscle precursors or tissues but was clearly detected in the trunk sclerotomes, suggesting that positive cells in the somites are also sclerotomal cells. This would represent a subdivision within the sclerotome compartment that has not been reported before. In E10.5 embryos (30 somites), expression was still seen in the floor plate of the midbrain although it was weaker in the posterior half (Fig. 9 D). In the gut endoderm derivatives, expression extended into the lung buds. There was no expression in the mesonephric tubules at this stage but weak staining in the adjacent Wolffian tubes (Fig. 9 C). Weak staining in the apical ectodermal ridge of the fore limb was also detected at E10.5.

At stages beyond E10.5, expression became quite widespread in the embryo, because of extensive staining in the gut endoderm and sclerotome derivatives, but was never detected in muscle or heart tissues. Beyond this stage, expression extended to some cells of the surface ectoderm in the skin and to chondrocytes. Expression persisted in the notochord as it became part of the nucleus pulposus of intervertebral discs. Expression was seen in the floor plate until at least E12.5, at which stage expression became more widespread in various parts of the central nervous system, including restricted areas of the brain and neural tube. In newborn and adult mice, expression could be detected in the skin epiderm and in the brain.

Expression of the lacZ reporter gene in homozygous embryos

Since embryos homozygous for the GT insertion are viable (see Table 11), I was also able to analyze the expression pattern with double amount of reporter gene expression. No major qualitative difference could be observed in the expression pattern at all the stages analyzed, but expression was expectedly higher, more easily detectable and homozygous embryos could be distinguished from heterozygous littermates on the basis of the signal intensity alone. A useful application is the possibility to genotype newborn pups by assessment of the

intensity of X-gal staining (after about 5 to 10 minutes of staining, homozygous tissues are distinctively more strongly labeled than the heterozygous ones).

In homozygous embryos, the weaker domains of expression previously observed in some tissues were not clearly detected. For example, the weaker domain of expression observed in the posterior mid-brain (Fig. 9 D) could not be distinguished anymore from adjacent domains. But the anterior limit of expression in the head is more readily seen. Stronger expression also revealed two additional domains of expression that were not identified in the analysis of heterozygous embryos. At the onset of its expression, *cobl* was also expressed in the hindgut pocket at the posterior end of the primitive streak (data not shown) and in extraembryonic tissues.

Extraembryonic expression of cobl

Extraembryonic expression was restricted to a few rows of visceral endoderm cells immediately adjacent to parietal endoderm and to cells located at the tip of the ectoplacental cone. Large numbers of expressing cells were also found in the maternal decidua. To clearly determine the embryonic or maternal origin of positive cells in the ectoplacental cone, I transferred wild-type blastocyst embryos into [GtC101/GtC101] pseudo-pregnant females (Fig. 10 A) and did the reverse experiment with [GtC101/GtC101] embryos (Fig. 10 B). The positive cells of the ectoplacental cone were only detected in [GtC101/GtC101] decidue and are therefore maternally derived. Expression in maternal tissues could be detected in single cells throughout the decidua but was also localized to duct-like structures organized around the embryo and connecting the inner cavity of the decidua with its outermost cells (Fig. 10 A). The only extraembryonic expression derived from the embryo was located in the visceral endoderm cells (Fig. 10 B). Expression in extraembryonic endoderm persisted at later stages in cells immediately adjacent to the placenta. In 12.5 dpc placentae, expression was also present in cells dispersed throughout the labyrinthine layer (data not shown).

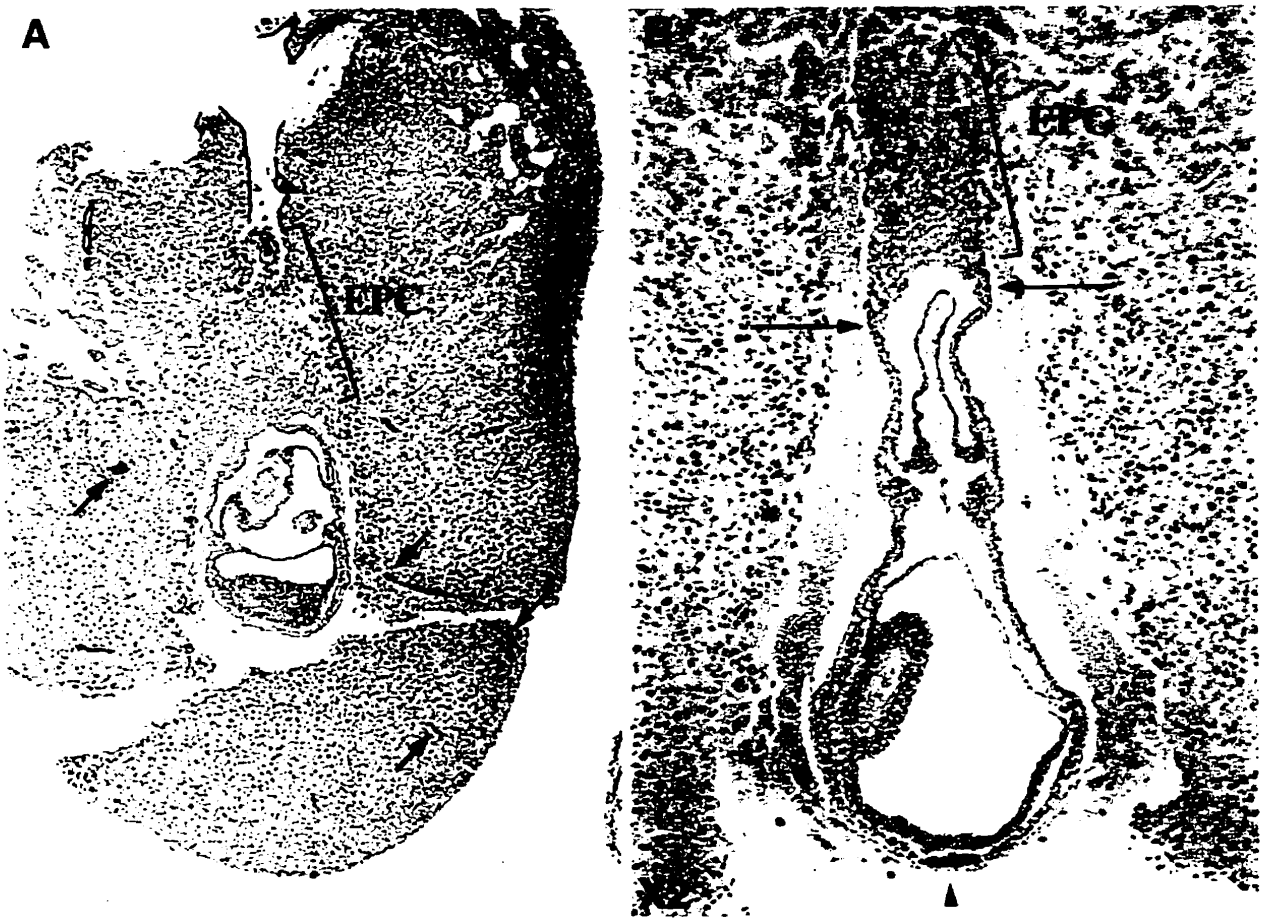
Cloning the endogenous gene

Northern blot analysis with a *lacZ* probe detected a single fusion transcript of 5.4 kb in RNA samples from C101 ES cells or heterozygous embryos (data not shown). *LacZ* containing cDNAs were synthesized and amplified by RACE-PCR from the ho-23 ES cell poly-A+ RNA sample and from E8.5 embryo total

Fig. 10. - Cobl/ß-Gal extraembryonic expression in embryo and maternal decidua at 7.5 dpc.

The respective genotypes of the embryos and decidue are indicated below the pictures. (A) Wild-type embryo implanted into a *GtC101* transgenic decidua. Expression can be seen in various parts of the decidua, especially at the top of the ectoplacental cone (EPC) and in "canals" converging toward the embryo (arrows). (B) *GtC101* embryo implanted into a wild-type decidua. No expression could be detected in the decidua, including the tip of the EPC (top of picture). The enlargement is twice as in (A), focusing on embryonic expression. Expression restricted to the node (arrowhead) and to some extraembryonic endodermal cells (arrows), at the limit between parietal endoderm and yolk sac endoderm.

Abbreviation: EPC: ectoplacental cone. Stain: eosin + nuclear Fast-Red.



Embryo +/+
Decidua GtC101/GtC101

Embryo GtC101/GtC101
Decidua +/+

RNA (see Fig. 11 for the cloning strategy). Two cDNAs (derived from ho-23 cells) were subcloned and sequenced. They contained 250 bp of overlapping endogenous sequence with a single open reading frame (ORF) spliced in frame with the coding portion of the *En-2/lacZ* gene. The longest (367 bp) cDNA was used for further molecular analysis. This sequence (Fig. 12 A) did not show any homology to any previously reported sequence. No recognizable protein domains were observed, although the protein encoded by this partial cDNA is rich in proline (15.7% of the total), serine (15.7%) and lysine (11.4%).

A 278 bp fragment, C101/7, that lacks any *En-2* sequence was subcloned, sequenced and subsequently used as a probe (Fig. 12 A). On Northern blots it hybridized to a 6.5 kb endogenous transcript in both D3 and C101 ES cell RNA samples and to the 5.4 kb fusion transcript in C101 alone (data not shown).

Since the fusion transcript was widely expressed in the brain, I screened an unamplified cDNA library from CD1 adult brain with the probe C101/7. Five different clones were detected, subcloned and partially sequenced. They all overlapped the C101/7 probe and altogether cover a 3.7 kb section of the endogenous transcript, starting 1 kb upstream from the splicing site fused to the reporter sequence and including 2.7 kb of coding sequence located 3' of the insertion site (Fig. 12 B). The 1200 bp b3 cDNA overlaps the C101/7 fragment (Fig. 12 B) and was used as a probe on Northern blots. It detected the same endogenous transcript and fusion transcript as the PCR probe C101/7 (Fig. 12 C). The fusion transcript was also detected with the En-2 probe on the same filter (Fig. 12 D). Levels of endogenous transcript were progressively reduced in heterozygous and homozygous cells, as expected. However low but detectable levels of wild-type transcript were still detected in RNA from the homozygous ES cell line ho-23, suggesting that the expression of wild-type transcript is not completely abolished (Fig. 12 C). This could occur by splicing around the insertion vector (Moens *et al.*, 1992; Skarnes *et al.*, 1992).

Expression of the endogenous gene

C101/7 sense and antisense riboprobes were synthesized and hybridized to sections of 8.5 and 9.5 dpc wild-type CD1 embryos. No signal was detected with the sense probe. The antisense probe detected transcripts in the gut (and presumptive liver at 9.5 dpc), the notochord and the floor plate (data not shown). The pattern of expression at these two stages was identical to the *lacZ* transgene expression. These

Fig. 11. - Diagram of the 5' RACE-PCR cloning strategy.

The 5'RACE-PCR technique was used to isolate endogenous sequence fused to the *lacZ* reporter gene. a) The first strand cDNA was synthesized from the *lacZ* reverse primer #1 (see materials & methods). b) Second strand synthesis and blunt ending was performed. c) EcoRI adapters were ligated to both ends of the cDNAs. d) The cDNAs were amplified by PCR with the EcoRI primers and primer #256. e) PCR products were subcloned and sequenced. f) A 279 bp fragment amplified with the primers #25 and #283 and sub-cloned into the vector pSL301 was used as an endogenous sequence probe.

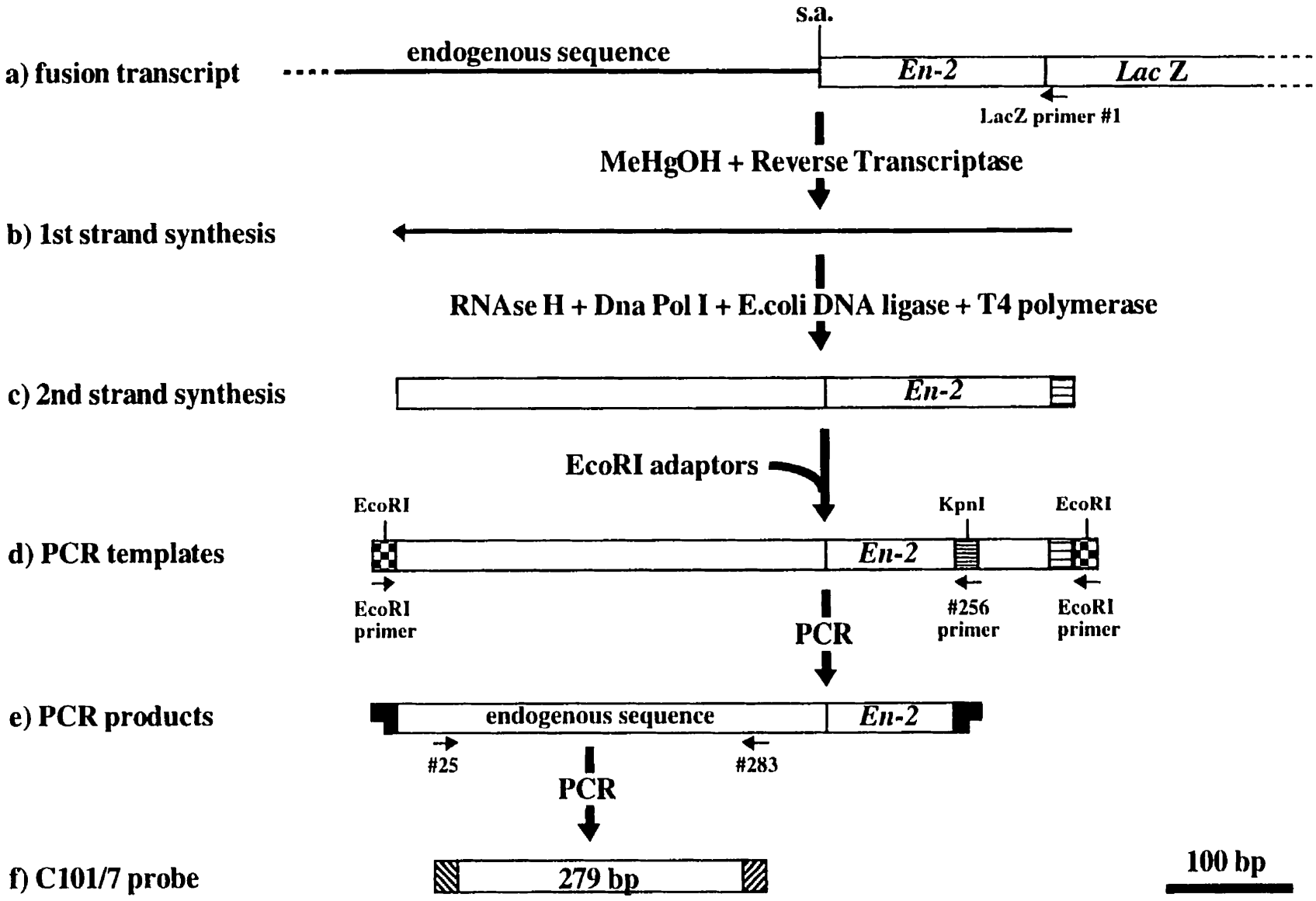
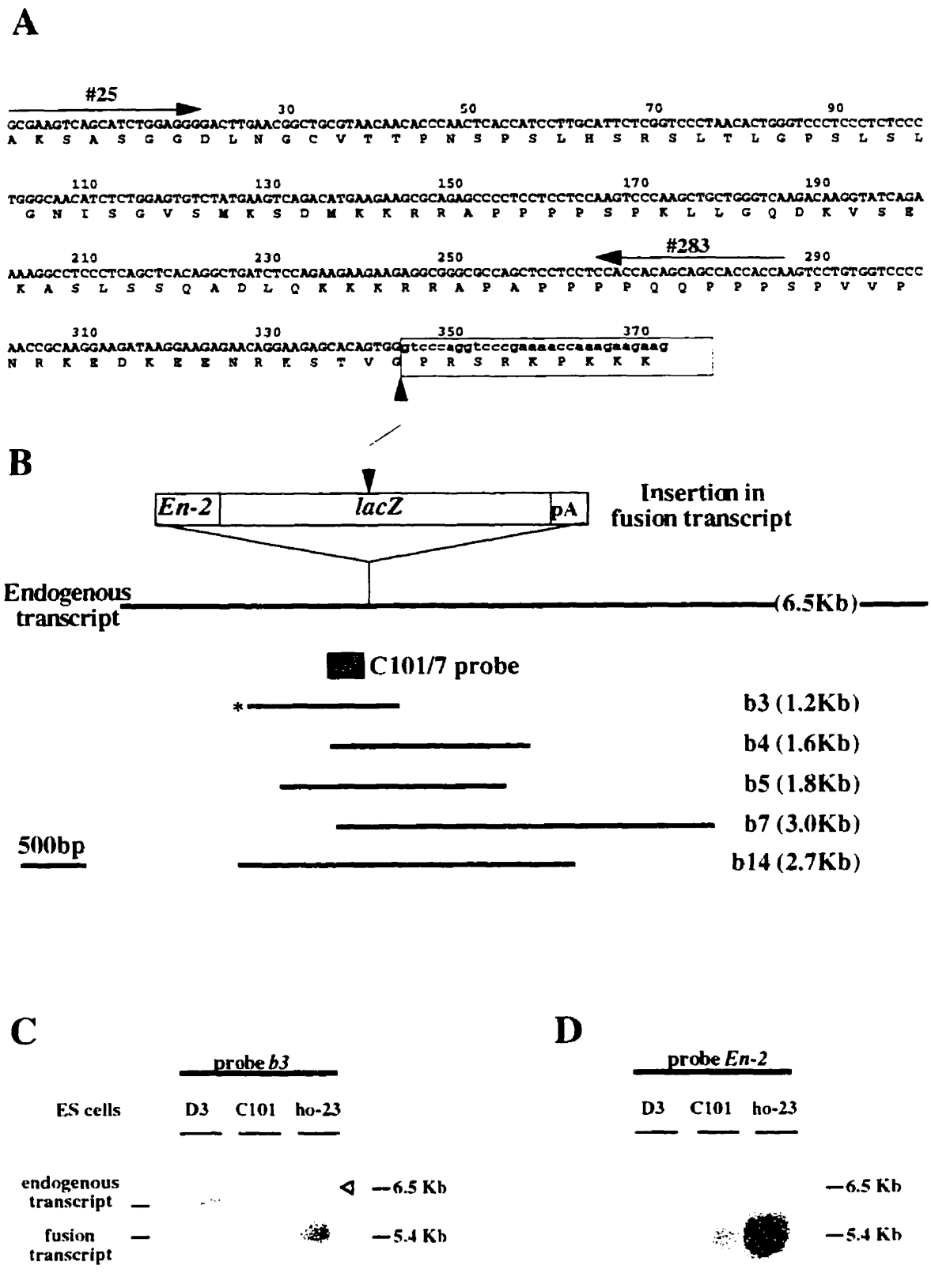

a) fusion transcript
s.a.
endogenous sequence
En-2
Lac Z
LacZ primer #1
MeHgOH + Reverse Transcriptase
b) 1st strand synthesis
RNAse H + Dna Pol I + E.coli DNA ligase + T4 polymerase
c) 2nd strand synthesis
En-2
EcoRI adaptors
d) PCR templates
EcoRI
KpnI
EcoRI
En-2
EcoRI primer
#256 primer
EcoRI primer
PCR
e) PCR products
endogenous sequence
En-2
#25
#283
PCR
f) C101/7 probe
279 bp
100 bp

Fig. 12. - Cloning mouse *Cobl* cDNAs: partial sequence and expression.

(A) 343 bp of endogenous sequence containing an ORF fused in frame (arrow) with En-2 coding sequence in the fusion transcript. 29 bp of the En-2 sequence (lowercase characters) and 9 corresponding amino acids are shown within a box. The sequence of the C101/7 probe is within the two arrows marking the positions of the primers #25 and #283. (B) The C101/7 probe and the five overlapping clones b3, b4, b5, b7 and b14, isolated from a mouse adult brain cDNA library, cover 3.7 kb of endogenous sequence. (C) A probe made from clone b3 (asterisk) was hybridized to a Northern blot carrying 10 µg of total RNA samples from the wild-type D3 cells, from the heterozygous C101 cells and from the homozygous ho-23 cells. It detects both the 5.4 kb fusion transcript and the 6.5 kb endogenous transcript also detected with the probe C101/7. As expected the amount of fusion transcript is increased in the sample from homozygous ho-23 cells but there are also detectable amounts of wild-type transcript (outlined arrowhead). This lane contained more RNA than the wild-type and heterozygous RNA lanes as judged by ethidium bromide staining. (D) The En-2 probe was hybridized to the same blot and only detected the 5.4 kb fusion transcript.


A
#25
#283
B
En-2
lacZ
pA
Insertion in fusion transcript
Endogenous transcript
(6.5Kb)
C101/7 probe
b3 (1.2Kb)
b4 (1.6Kb)
b5 (1.8Kb)
b7 (3.0Kb)
b14 (2.7Kb)
500bp
C
probe b3
ES cells
D3
C101
ho-23
endogenous transcript
fusion transcript
—6.5 Kb
—5.4 Kb
D
probe En-2
D3
C101
ho-23
—6.5 Kb
—5.4 Kb

results confirmed that the C101/7 probe corresponded to sequence from the endogenous gene trapped by the vector.

The longer b3 cDNA (Fig. 12 B) was used as a probe to detect the pattern of expression of the endogenous gene by whole mount in situ hybridization. At E7.5 expression was detected in the node (Fig. 13 A) and at E8.5 the expression was restricted to midline structures of the embryo from the mid brain to the tail bud (Fig. 13 B, C). Upon sectioning this staining was restricted to the roof of the gut, the notochord and the floor plate of the neural tube (Fig. 13 D). Overall the pattern of expression of the endogenous sequence was identical to the pattern revealed by the *lacZ* fusion. In whole mounts, expression in the head process at E7.5 was weak and at E8.5 the expression in the node did not seem as strong and extensive as the *lacZ* marker. These differences could be accounted for by differences in the intrinsic stability of the *lacZ* fusion transcript and the endogenous transcript, by the half-life of the *lacZ* fusion protein itself, or by the detection technique.

However, the expression pattern and the partial sequence, clearly indicated that the C101 insertion had occurred into a novel gene. It was named *cordon-bleu* (*cobl*) because of the specific expression of the lacZ reporter in axial structures of gastrulating embryos.

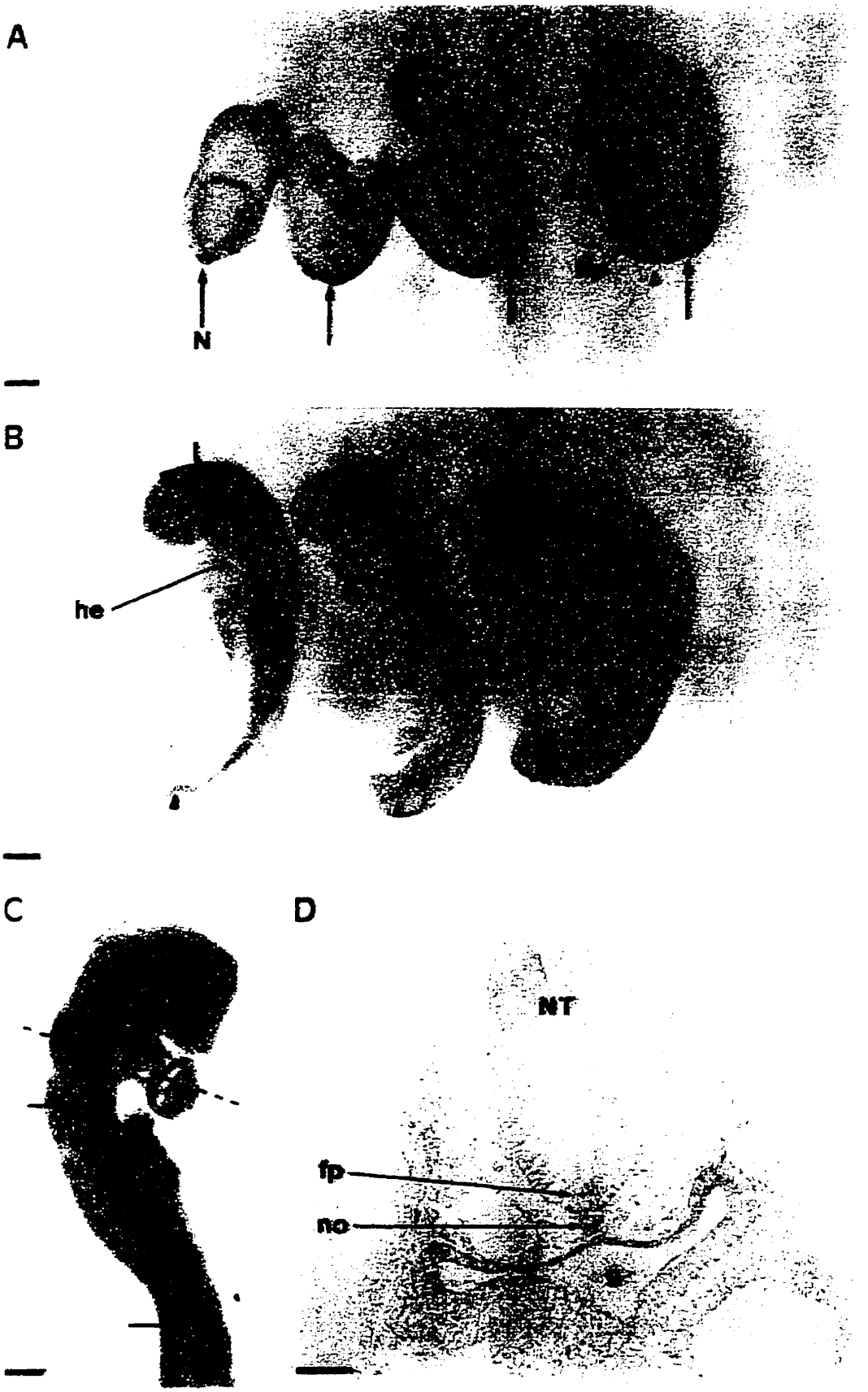
Cloning the complete coding sequence of cobl

I used the b3 cDNA to clone additional cDNAs and I generated probe from the extremities of the cDNAs covering the 5' and 3' ends (b31a and b7). I was thus able to clone cDNAs covering the entire open reading frame and the 3' end of the full length *cobl* cDNA (Fig. 14 B). A 75 bp putative exon coding for 25 a.a. (within brackets in Fig. 15) was only found in 2 cDNAs: the PCR generated C101/7 probe and clone b5. Not all the cDNAs were tested for the presence of this sequence (Fig. 14 B). However, removal of this exon did not alter the reading frame. In addition, some mouse EST sequences homologous to *cobl* (Fig. 14 C) were identified in the expressed sequence tag (EST) public database (Washington University/Merck EST Project). However, these clones appeared to be chimeric cDNAs: the 5' sequence of W20909 is first homologous to *cobl* exons but switches to unknown sequence precisely from the same splicing site fused to the reporter gene in the fusion transcript; the sequence of the clone W81955 goes beyond the poly-A tail of the 3' end of *cobl*. I could not get cDNA clones extending to the expected 5' end of *cobl*. The Northern expression data suggest that the size of the wild-type transcript was 6.5 kb. However, in a previous Northern blot, I had estimated the size of the endogenous

Fig. 13. - RNA *in situ* hybridization on wild-type whole-mount embryos with *b3/cobl* riboprobes.

Distal is at the bottom in (A). Anterior is at the top in (B, C). (A) At E7.5, expression is detected in the node (arrows) and the head-process (between arrowheads). (B, C) At E8.5, expression extends from the anterior midbrain and foregut (arrows) to the tail bud (arrowheads). The dotted line in C indicates the plane of the section shown in (D). Transverse section showing that the axial expression is restricted to the roof of the gut, the notochord and the floor plate of the neural tube.

Abbreviations: floor plate (fp), gut (g), node (N), notochord (no).
Scale bar = 25 µm (D), 100 µm (A,B,C).


A
N
B
he
C
D
NT
fp
no

transcript at 5.5 kb (data not shown). Thus the real length of the full cDNA probably lies between these values. Nevertheless, the complete sequence of the 5.5 kb cDNA revealed the primary structure of the predicted protein.

Sequence of the cobl gene

The cDNA inferred by the contig is 5465 bp long with an additional 16 bases forming the poly-A tail. It contains an ORF coding for a predicted protein of 1340 amino acids (Fig. 15). The leading sequence is 78 bp long and contains a stop codon (nucleotide #19) in frame with the main ORF. The ORF starts with a methionine (nt. #79) which is part of a Kozak consensus sequence:

	-12 -3 +4	
Vertebrate consensus	CCCGCCGCCACCATGG	(Kozak, 1987)
Cobl	CCCGCCGGCTCCATGG	
Human *A-raf-1*	atCtaaGgcTCCATGG	(Beck *et al.*, 1987)

The thymidine at position -3 is only present in 1% of the 699 vertebrate mRNAs analyzed by M. Kozak (Kozak, 1987). It is also the most conserved position of the consensus among higher eukaryotes and is occupied by a purine in 97% cases (generally an adenosine). However, the human *A-Raf-1* oncogene for example, possesses an initiation codon similar to that of *cobl* with a thymidine at the same position (see above).

The predicted 144 kDalton protein is rich in proline (10%) and serine (12%) and is hydrophilic. It contains 12 possible (consensus) potential asparagine (N)-glycosylation sites (Fig. 15), 1 potential amidation site (not shown), 2 potential glycosaminoglycan attachment sites (not shown) and 20 potential N-myristoylation sites (not shown). Because serine and threonine make up 18.3% of the protein there is also a large number of potential phosphorylation sites for cAMP and cGMP-dependent protein kinases (6 sites), casein kinase II (28 sites) and protein kinase C (28 sites).

Two small repeats with 9 out of 11 identical a.a. also contained a potential nuclear localization signal (nls, Fig. 15): the a.a. sequence Lys-(Lys or Arg)-X-(Lys or Arg) has been described as a consensus nuclear localization signal sequence (Chelsky *et al.*, 1989; and reviewed by Garcia-Bustos *et al.*, 1991). Furthermore, these signals are often found in pairs only separated by 10 to 20 a.a. (28 a.a. in Cobl, Fig. 15). This could explain why the Cobl/ß-Gal signal is often localized to nuclei (Fig. 16 A-C). However, Cobl/LacZ is not always localized to the

Fig. 14. - Diagram of the mouse *cordon-bleu* cDNA contig.

(A) Schematic representation of the *cobl* cDNA and C101/7 probe: the ORF is depicted by a white box: the three motif repeats (see Fig. 15 and 17 B and C) are marked by grey boxes and the poly adenylation sites (pA) are indicated. (B) Map of the cloned cDNAs: dotted lines separate different hybridization rounds: cDNAs presented as grey rectangles overlap the entire length of the *cobl* contig: the black box present in b5 and partly in C101/7 is a putative exon missing in cDNAs b3, b14 and b31a. (C) Mouse ESTs identified in public databases (with their GenBank accession numbers).

Abbreviations: bp: base pair: EST: expressed sequence tag: kb: kilobases: ORF: open reading frame: pA: poly-adenylation signal sequence.

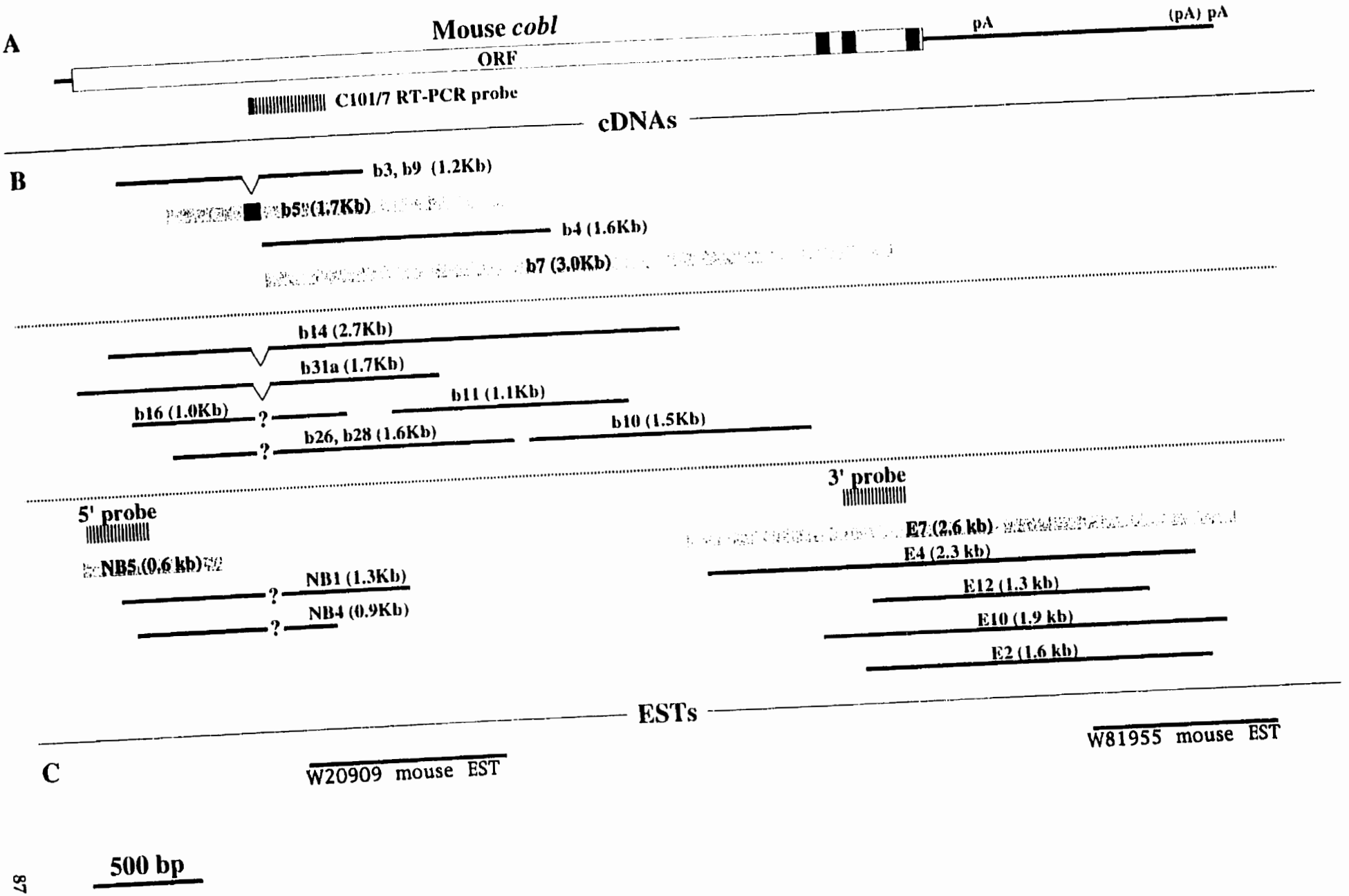

A
Mouse cobl
ORF
pA
(pA) pA
C101/7 RT-PCR probe
cDNAs
B
b3, b9 (1.2Kb)
b5 (1.7Kb)
b4 (1.6Kb)
b7 (3.0Kb)
b14 (2.7Kb)
b31a (1.7Kb)
b16 (1.0Kb)
b11 (1.1Kb)
b26, b28 (1.6Kb)
b10 (1.5Kb)
5' probe
3' probe
NB5 (0.6 kb)
NB1 (1.3Kb)
NB4 (0.9Kb)
E7 (2.6 kb)
E4 (2.3 kb)
E12 (1.3 kb)
E10 (1.9 kb)
E2 (1.6 kb)
ESTs
W20909 mouse EST
W81955 mouse EST
C
500 bp

Fig. 15. - Sequence of the *cordon-bleu* gene.

Numbers indicate positions on the nucleotide sequence. The 1340 amino acids (a.a.) long open reading frame (ORF) extends between two stop codons (each marked by 3 asterisks) at position #19 and #4099. The predicted first codon (methionine codon at position #79) is part of a Kozak consensus sequence (overlined). The asparagine residues of potential N-glycosylation sites are in white over a black circle. The 25 a.a.'s sequence within brackets (starting at position #860) was not encoded by all cDNAs (see Fig. 14). The 2 potential nuclear localization signals are in black boxes (positions #1045 & #1141). The gene trap *lacZ* insertion occurred between the positions #1249/50 (Black triangle). The three 20 a.a.'s motif repeats are in clear rectangles. The 3' untranslated sequence contains 2 poly-adenylation signals (in bold).

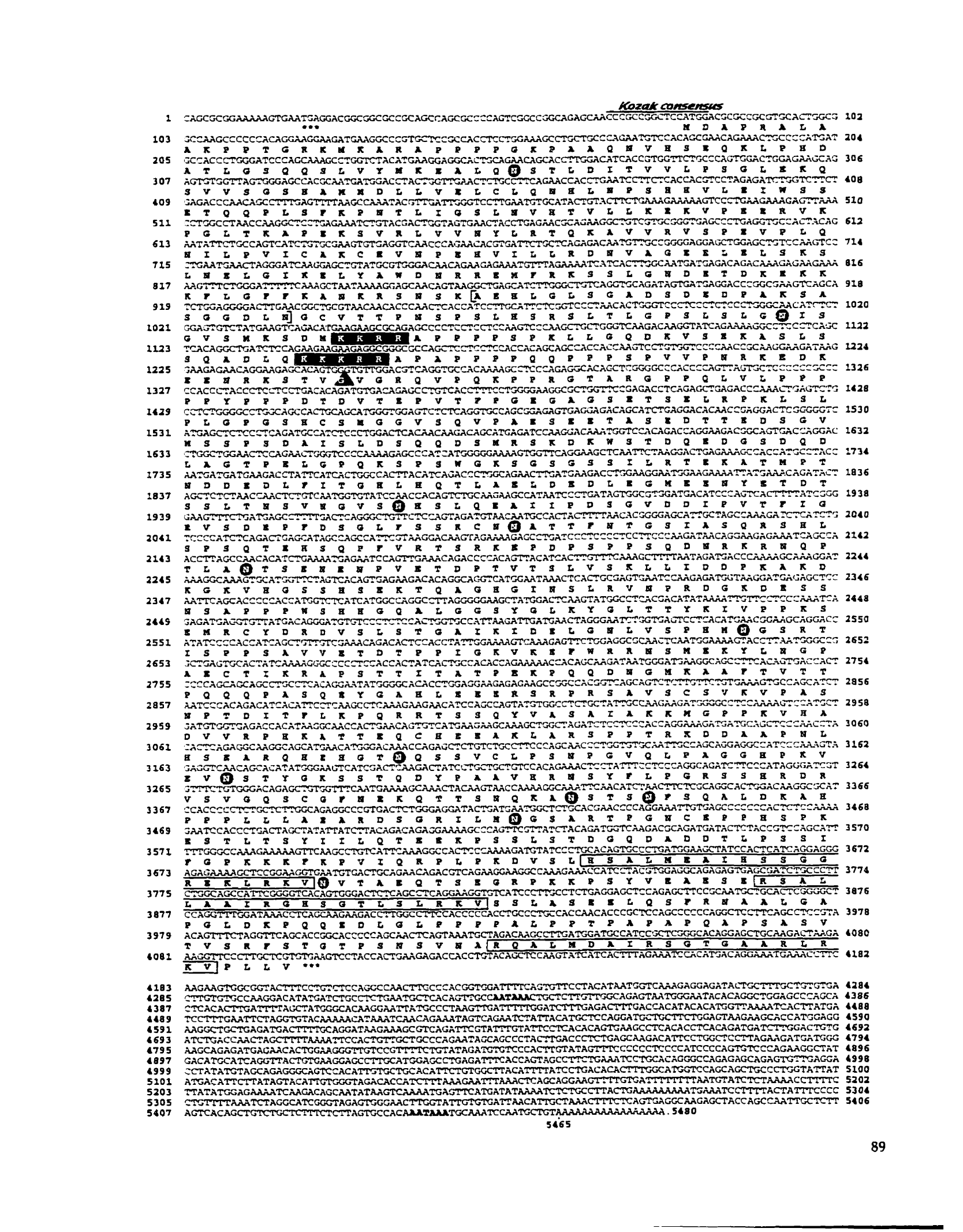

Kozak consensus

Fig. 16. - Nuclear localization of Cobl/ß-Gal.

(A-C) Dorsal skin from E15.5 embryos. (A) Nuclear + cytoplasmic stain. Region showing a mixture of expressing and non-expressing cells: the morphology of the sub-cellular domains expressing ß-Gal suggests that expression is localized to the nucleus. (B) Nuclear stain only. Region where most cells are positive. The stain and the X-gal precipitate clearly co-localized to the nuclei: some cells are negative (small arrow): the staining may be cytoplasmic (large arrow) or delocalized when cells are dividing (arrowheads). (C) Nuclear + cytoplasmic stain. Most cells are positive but negative nuclei are visible (arrows). (D) Protein sequence upstream of the Cobl/En-2/ß-Gal fusion (numbers correspond to the protein sequence). Cobl sequence is in bold type and only En-2 sequence appears as plain text. Two putative nuclear localization signals in Cobl and 1 in En-2 are indicated by subscript typing. (E) Alignment of the nuclear localization signals from the 2 cobl small repeats and from En-2 with a consensus nuclear localization signal.

Abbreviations: En-2: Engrailed-2; lacZ: ß-Galactosidase.
Bare = 50 µM.


A
B
C
D
323
355
391
cordon-bleu
En-2/lacZ junction
E
321DMKKRRAPPPPSPK334
354QKKKRRAPAPPPPQ367
cordon-bleu repeats
EDKRPRTAFTAEQL
En-2 signal
nuclear location signal consensus

nucleus and part of the En-2 sequence present in the fusion protein also contains a potential nls sequence (Fig. 16 D, E). Therefore, the sub-cellular localization observed in transgenic cells could be due to an artifact carried by the hybrid Cobl/ß-Gal protein. Another motif is repeated three times at the C-terminal end of Cobl. I could not detect any other motif such as a signal peptide sequence or any potential transmembrane domain and overall, Cobl does not show any similarity to any other known protein or motif.

The 3' untranslated region is 1.3 kb long and contains 2 poly-A consensus signals (Sheets *et al.*, 1990), with one, located 15 nt. upstream of the poly-A tail, which is probably essential for the poly-adenylation process.

Phenotype and mapping of the insertion

Upon crossing heterozygous mice, we recovered offspring homozygous for the insertion with the expected frequency (Table 11). Homozygosity was confirmed by Southern blot analysis with the *En-2* probe and for the males by breeding with CD1 wild-type females. Homozygous mice were healthy, presented no observable phenotype and were fertile. The homozygous insertion is now maintained on inbred 129/Sv and outbred CD1/129/Sv backgrounds.

The C101/7 cDNA was mapped using the Jackson Laboratories BSS [(C57BL/6JEi X SPRET/Ei) F1 female X SPRET/Ei male] interspecific backcross DNA panel (see Material & Methods and Rowe *et al.*, 1994). It mapped to chromosome 11 between the D11Bir3 (Rowe *et al.*, 1994) and D11Hun1 (McCarthy *et al.*, 1995) loci. It maps near the *Epidermal Growth Factor Receptor* locus (*EGFR* - a.k.a. the *wa-2* mutation locus and the *ErbB* oncogene). The *cobl* locus carrying the GT insertion was named *GtC101* in accordance with mouse genomic nomenclature.

Cloning and mapping the human cobl

I used the full sequence of the *cobl* gene to search for sequence homologies in public databases. The 3' untranslated region and the end of the ORF allowed me to identify human expressed sequence tags (ESTs) presenting a high degree of identity with *cobl*. Sequence data from these ESTs revealed the sequence of the 3' end of the human homologue of *cobl* (Fig. 17 A, B). Some 3' untranslated regions were 80% identical to the mouse gene and the degree of overall identity is 73% in the overlapping C-terminal portion of the ORF (Fig. 17 B). In the human ORF, the 3 repeats are present and show 75%, 95% and 100% identity to the mouse repeats (Fig. 17 B). The higher degree of conservation within the repeats suggests that their

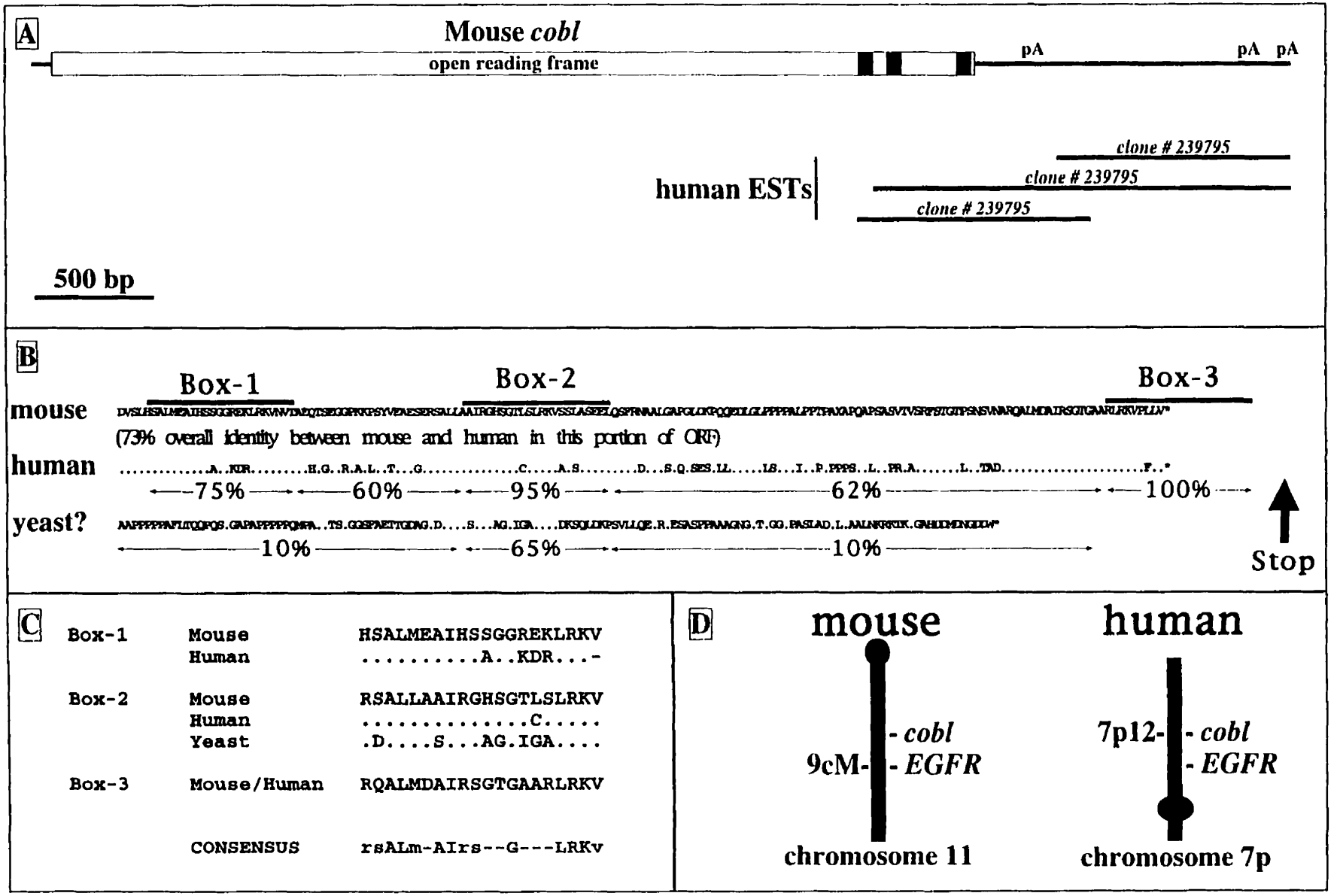
structures may have important roles. Furthermore, a gene from the yeast *S. cerevisiae* (Toh-E. A., 1996. Yeast mutants sensitive to local anesthetics. Unpublished. GenBank #D78487) also contains a motif similar to the second repeat (65% identity with mouse box-2) toward its C-terminal end; like Cobl it is also a proline rich protein. However, the function of this protein is not known and no other similarity was found with the rest of the yeast sequence.

The human ESTs were also used to map the human *cobl* gene. Mouse *cobl* maps to a region of chromosome 11 which is syntenic to a portion of human chromosome 7p where EGFR had already been mapped. Indeed, human *cobl* also mapped to the same syntenic region on chromosome 7p12 (Fig. 17 D). The cobl loci on mouse chromosome 11 or on human chromosome 7p12 do not bear any mutation for which *cobl* could be a potential candidate.

Fig. 17. - Cloning the human homologue of *cobl*.

(A) Schematic representation of the mouse *cobl* cDNA and human ESTs: the ORF is depicted by a white box: the three motif repeats are marked by grey boxes. The numbers over the ESTs refer to their Clone ID number in the EST database. (B) Alignment of the of mouse and human Cobl protein sequences. The three repeated motifs are overlined and were named box-1 to -3: dots indicate conserved a.a. and an asterisk indicates a stop codon. A yeast protein sequence containing a motif with strong homology to box-2 was also aligned. The regions with distinct degrees of homology (indicated in percentage) are separated by opposing arrowheads. (C) Alignment of the 3 conserved motifs with each other and consensus motif (capitals for conserved a.a., lowercase for most frequent a.a. and "-" sign for gap in sequence alignment). (D) Chromosomal mapping of mouse and human *cobl* to syntenic regions.

Abbreviations: a.a.: amino acids; bp: base pair; EST: expressed sequence tag; kb: kilobases; ORF: open reading frame; pA: poly-adenylation signal sequence.


A
Mouse cobl
open reading frame
pA
pA pA
human ESTs
clone # 239795
clone # 239795
clone # 239795
500 bp
B
Box-1
Box-2
Box-3
mouse
(73% overall identity between mouse and human in this portion of ORF)
human
75%
60%
95%
62%
100%
yeast?
10%
65%
10%
Stop
C
Box-1 Mouse HSALMEAIHSSGGREKLRKV
Human .....A..KDR...-
Box-2 Mouse RSALLAAIRGHSGTLSLRKV
Human .....C.....
Yeast .D....S...AG.IGA....
Box-3 Mouse/Human RQALMDAIRSGTGAARLRKV
CONSENSUS rsALm-AIrs--G---LRKv
D
mouse
human
cobl
EGFR
9cM-
7p12-
chromosome 11
chromosome 7p

DISCUSSION

Pattern of expression of cordon-bleu

The GtC101 gene trap insertion identified a novel gene, *cordon-bleu* (*cobl*), expressed in the node, the notochord, the gut and the floor plate. Although the sequence of Cobl gives no clue as to its possible function, it joins a growing number of genes expressed in midline axial structures

Expression begins at 7.5 days in a group of mesodermal cells located at the distal tip of the embryo, in the node region. The node region in the mouse is equivalent to Hensen's node in the chick or the amphibian dorsal lip, which both have the capacity to act as an organizer during embryogenesis by generating and inducing in surrounding tissues the proper formation of the embryonic axis (Spemann and Mangold, 1924; Waddington, 1933). The mouse node was also shown to be able to induce embryonic axis formation in the mouse (Beddington, 1994; and reviewed by Streit *et al.*, 1993). Subsequently at E8.5 expression of *cobl* was restricted to the roof of the gut, the notochord and the floor plate of the neural tube. This pattern of expression is consistent with the close embryological relationships between these tissue types. Fate mapping in chick (Selleck and Stern, 1991) and mouse embryos (Lawson *et al.*, 1986; Lawson *et al.*, 1991; Beddington, 1994; Wilson and Beddington, 1996) showed that cells of the definitive endoderm and the notochord can originate from common precursor cells in the node. Further, it was shown in the chick that, although the floor plate of the neural tube may arise later, it fails to form in the absence of a notochord and that an ectopic graft of an extra notochord can induce the formation of a second floor plate in the adjacent neural tube (van Straaten *et al.*, 1988; Placzek *et al.*, 1990). Mouse mutants lacking a definitive notochord, like T^{wis}/T^{wis} (Herrmann, 1991; Conlon *et al.*, 1995) and *HNF-3β*$^{-/-}$ mice (Ang and Rossant, 1994; Weinstein *et al.*, 1994) do not develop a floor plate. Many inductive interactions have been shown to occur between the notochord, the floor plate (reviewed by Jessell and Dodd, 1992; Ruiz i Altaba and Jessell, 1993) and adjacent mesodermal tissues at different stages of development (Dietrich *et al.*, 1993; Koseki *et al.*, 1993; Pourquié *et al.*, 1993). The expression of *cobl* in the node, the notochord, the gut and the floor plate suggests that it may be involved in their formation or in their inductive interactions.

An important issue is to determine where *cobl* stands in the hierarchy of genes already identified that are expressed in the same structures. Genes displaying expression patterns similar to that of *cobl* but showing earlier onset of

expression, may be part of the genetic pathway regulating its expression. At least three transcription factors, goosecoid, HNF-3ß and Brachyury (T), are expressed before *cobl* in the mouse node. The potential transcription factor goosecoid is expressed in the anterior part of the early primitive streak and the node (Blum *et al.*, 1992) and has been shown to have axis-inducing properties in amphibians by injection into ventralized embryos where it can rescue the formation of dorsal mesoderm and notochord (Cho *et al.*, 1991). However the *goosecoid* gene is not expressed in the notochord or the floor plate at later stages (Gaunt *et al.*, 1993) and is not essential for node formation (Rivera-Perez *et al.*, 1995; Yamada *et al.*, 1995). The winged-helix transcription factor HNF-3ß is also expressed in the anterior primitive streak and the node (Ang *et al.*, 1993; Ruiz i Altaba *et al.*, 1993b; Sasaki and Hogan, 1993). It is essential for notochord and axis formation (Ang and Rossant, 1994; Weinstein *et al.*, 1994) and can affect normal patterning of the neural tube (Sasaki and Hogan, 1994; Ruiz i Altaba *et al.*, 1993a). Because it is expressed before *cobl* in the streak and the node, HNF-3ß could be regulating *cobl* expression. HNF-3ß is a good candidate for regulating *cobl* since its expression pattern is very similar to that of *cobl* in the gut, the notochord and the floor plate at E8.5, in the liver at E9.5 and the gut-derived lung epithelium at E12.5. However, some expression of *cobl* was still detected in HNF-3ß$^{-/-}$ mutant embryos, probably in endoderm cells (Ang and Rossant, 1994) suggesting that, at least in some tissues, its expression is independent of HNF-3ß. The *Brachyury* transcription factor is essential for mouse notochord formation (Herrmann, 1991) and is expressed throughout the primitive streak, the node and the notochord (Wilkinson *et al.*, 1990). It possible role in the regulation of cobl expression is analyzed in the next chapter.

The TGF-ß-related *nodal* (Zhou *et al.*, 1993), expressed early in the primitive streak, could also be a component of the signaling pathway regulating *cobl* expression. However, at the onset of *cobl* expression, it is expressed in a domain surrounding the node, rather than in the node itself where *cobl* is expressed. Furthermore, nodal is also involved in the determination of the left-right axis and expressed asymmetrically in the node (Collignon *et al.*, 1996). A similar asymmetric expression was never observed with cobl. Finally, the phenotype of *nodal* mutants suggests that it is required for primitive streak formation itself rather than axis formation (Conlon *et al.*, 1994).

The pattern of expression of *Shh/vhh*, one of the vertebrate homologues of the *Drosophila hedgehog* gene (Echelard *et al.*, 1993; Roelink *et al.*, 1994; Marti *et al.*,

1995; Iseki *et al.*, 1996), is restricted to the same axial structures expressing *cobl* and *HNF-3β*. The onset of expression of *cobl* also coincides with the first expression of *Shh*. *Shh* is a secreted factor important for the proper patterning of vertebrate embryonic axes in the neural tube (Echelard *et al.*, 1993; Roelink *et al.*, 1994; Ekker *et al.*, 1995; Chiang *et al.*, 1996) in the chick limb (Riddle *et al.*, 1993) and in the somites (Fan and Tessier-Lavigne, 1994; Johnson *et al.*, 1994; Fan *et al.*, 1995). Chick Shh can induce ectopic expression of *HNF-3β* in the neural tube (Echelard *et al.*, 1993) and both genes may be acting through a common genetic pathway where HNF-3ß would be the primary signal for expression of Shh. However, some endodermal expression of Shh occurs in HNF-3ß$^{-/-}$ mutants (Ang and Rossant, 1994). The spatial overlap of expression between these two genes and *cobl* suggests that *cobl* may be part of the same genetic or developmental pathway.

The pattern of expression of *cobl* is also distinct from all other axial genes in some aspects. At its anterior limit of expression, there is no extension of expression from the midline into the rostral and lateral neural tube, as occurs with *Shh* and *HNF-3β* (Echelard *et al.*, 1993; Roelink *et al.*, 1994; Ang *et al.*, 1993; Ruiz i Altaba *et al.*, 1993b; Sasaki and Hogan, 1993). *Cobl* expression was seen in the AER of the limbs but not in the zone of polarizing activity (ZPA) where *Shh* is expressed (Echelard *et al.*, 1993; Roelink *et al.*, 1994). Expression in the notochord persists at stages when *HNF-3β* expression decreases (Ang *et al.*, 1993; Ruiz i Altaba *et al.*, 1993b), and expression in the gut is broader than that of Shh (Echelard *et al.*, 1993; Roelink *et al.*, 1994). Expression in the somites has not been reported before for the genes expressed in axial structures. Furthermore, the domain of expression in the somites marks a distinct compartment that is not revealed by the expression domains of myogenic factors or new bHLH genes such as *paraxis* (Burgess *et al.*, 1995b) and *scleraxis* (Cserjesi *et al.*, 1995). Cobl may therefore play roles outside of axis development and be regulated by different factors than the axial genes. However, the early pattern of expression of *cobl* strongly implicates it in events leading to the formation of the embryonic axis.

Phenotype of the insertion

Although the pattern of expression of *cobl* is very suggestive of an involvement in axis formation, we do not yet have any functional evidence for such a role. Mice homozygous for the GT insertion are viable and fertile. There are different possible explanations that can account for the lack of phenotype following a GT insertion (see general discussion in Chapter V). However, there is a

significant amount of wild-type transcript in ho-23 ES cells homozygous for the insertion. This result suggests that wild-type transcripts may be produced via splicing around the insert, thus providing enough wild-type protein to rescue the normal function. Thus, the lack of phenotype in homozygous mice need not mean lack of essential *in vivo* function for the gene and further functional assays will be needed to assess this result.

Cobl is a novel gene

Further insights into the role of a novel gene can be provided by its coding sequence and its chromosomal localization. The gene identified here was named *cordon-bleu* (*cobl*) based on its pattern of expression. It mapped to the proximal region of chromosome 11, where no plausible related gene or mutation resides. Further, the sequence data available from the 5.5 kb cDNA demonstrates that the GT vector inserted into a novel gene. Apart from the human homologue of *cobl* and the short motif also found in a yeast gene, sequence analysis to date has not revealed any similarity in sequence to any known protein or protein domain, including all the genes associated with the node and the notochord in other vertebrate species. No further insight into the possible function of the gene has yet been provided by this analysis.

Even without knowledge of the function of *cobl*, this GT insertion of the *lacZ* reporter gene into *cobl* demonstrates the usefulness of the gene trap approach in identifying novel genes, that could not be identified by homology with other genes, as well as displaying their pattern of expression during mouse development. It also provides a new, easily detectable marker for phenotypic analyses. In the next chapter, I have used the GtC101 insertion as a developmental marker to analyze the effects of the *Brachyury* and *Danforth's short tail* mutations on early node, notochord and embryonic axis development.

- CHAPTER IV -

CORDON-BLEU EXPRESSION IN *BRACHYURY* AND *DANFORTH'S SHORT-TAIL* MUTANTS.

CHAPTER FOUR : *CORDON-BLEU* EXPRESSION IN *BRACHYURY* AND *DANFORTH'S SHORT-TAIL* MUTANTS.

INTRODUCTION

In the mouse, many classical mutations affect the development of skeletal structures (Lyon and Searle, 1989). Skeletal abnormalities often result from defects occurring during early embryogenesis. Mutations affecting the development of the notochord are particularly interesting because the notochord is not only important for the formation of axial skeleton but also for the patterning of surrounding tissues during early organogenesis. Four semi-dominant mutations, *Brachyury* (*T*), *Danforth's short-tail* (*Sd*), *loop-tail* (*Lp*), *pintail* (*Pt*) and a recessive one, *truncate* (*tc*), affect morphogenesis of the notochord (Lyon and Searle, 1989 and table 12). The *T* and *Sd* mutations have been more extensively studied, partly because of their drastic effects on notochord development. The other mutations present a weaker notochord phenotype. The *Lp* mutation likely affects posterior notochord morphogenesis because of posterior primitive streak defects. The *tc* mutation is recessive and involves random disruption of the notochord along the anterior posterior axis. The *Pt* allele only results in a reduction of the notochord size and length and is only partially lethal. Mutations in the *T* and *Sd* genes have strong effects on the morphogenesis of the notochord in both heterozygous and homozygous embryos. They cause strong but very different phenotypes resulting from the lack of an intact notochord at different embryonic stages and thus provide complementary models for the analysis of axial development and patterning.

The Brachyury gene

The first mutant allele of the *Brachyury* gene, identified by Dobrovolskaïa-Zavadskaïa (1927), is a large deletion encompassing the entire gene. The *T* gene maps to the t region on chromosome 17. This region has been the focus of many genetic studies, and since short-tailed heterozygous *T*/+ mice are easily identifiable, more alleles of the *T* gene have been subsequently isolated. They include other deletions and frame shifts resulting from point mutations or insertion of exogenous DNA. Mice homozygous for mutations in the *T* gene die at mid-gestation from defects in the primitive streak resulting in the lack of posterior axis elongation and allantois formation. The allantois fusion with the

Table 12

Mouse classical mutations affecting notochord development [a]

Stage of notochord development affected	Gene [b] / Mutation	Mutation dominance (Chromosome)	Notochord phenotype	(Time of death) cause of death in homozygous
Notochordal plate formation	***Brachyury (T)*** [c]	Semi-dominant (17)	***T/+***: at E10.5, notochord is interrupted or absent in the tail; mice are born with a short tail or no tail at all. ***T/T***: notochord fails to form; primitive streak does not regress; body truncated posterior to the fore limbs .	**(E10.5)** No chorio-allantoic fusion
Early notochord survival	***Danforth's short tail (Sd)***	Semi-dominant (2)	***Sd/+***: at E9.5, progressive rostro-caudal interruption of the notochord and abnormal formation of the perichordal sheet; reduced antero-posterior axis; short tail. ***Sd/Sd***: increased phenotype with total breakdown of the notochord at E10.5; the tail is absent at birth.	**(birth)** No excretory organs (kidneys and anus); Auto-intoxication. [d]
Posterior elongation	***loop-tail (Lp)***	Semi-dominant (1)	***Lp/+***: delay in neural tube closure causing a looped and crooked tail (not fully penetrant). ***Lp/Lp***: from E9.5, the regression of the primitive streak is delayed causing a shorter neural plate and notochord.	**(birth)** Open neural tube craniorachischisis
Notochord growth	***pintail (Pt)*** [c]	Semi-dominant (4)	***Pt/+***: at E10.5, the rate of cell division in the notochord is reduced leading to a progressive rostro-caudal reduction of nucleus pulposus (notochord) and intervertebral disc size; kinked tail with a threadlike tip. ***Pt/Pt***: represent a more severe phenotype than Pt/+.	**High mortality (pre-weaning)** Survivors are healthy but smaller.
Posterior elongation	***truncate (Tc)***	Recessive (6)	***Tc/Tc***: between E9.5 and E10.0, the notochord is disrupted and sclerotome cells degenerate; the posterior spinal column is interrupted; the tail is short or absent. Penetrance is incomplete.	**(viable)** Posterior truncation with paralysis of the hindlegs

([a]) From Lyon and Searle (1989).
([b]) The *T* gene encodes a potential transcription factor; The other loci have not been cloned.
([c]) The *T* locus comprises many alleles (T^c, T^h, T^{hp}, T^{or}, T^{orl}, T^{w18}) presenting more or less severe phenotypes. Most of them were induced mutations.
([d]) Some heterozygotes also die within the first three weeks.
([e]) Induced by Methylcholanthrene.

chorion is essential for the formation of the placenta and therefore survival of the embryo.

The earliest defects in *Brachyury* mutants appear in axial mesoderm. In homozygous embryos, only a few notochord precursors are present at day 8.5 and the notochord never forms. Furthermore, somites do not form or are very abnormal and structures posterior to the future fore limb position do not develop normally. The notochord of E10.5 heterozygous embryos is interrupted or altered in the sacral and tail region thus blocking growth of the tail bud. The *T* gene was cloned (Herrmann *et al.*, 1990) and encodes a putative transcription factor (Kispert and Herrmann, 1993) with cell autonomous function (Rashbass *et al.*, 1991). The *T* gene is conserved among vertebrates (Schulte-Merker *et al.*, 1994). Its expression in the nascent primitive streak and notochord is also conserved across species consistent with a role for the *T* product in the development of these tissues. Mutant alleles such as T^{ct} or T^{Wis} (Shedlovsky *et al.*, 1988), encoding a truncated *T* protein (Herrmann, 1991), have a stronger phenotype than *T* null alleles, suggesting that *T* probably interacts with other proteins. Heterozygous mice carrying these alleles are tail-less. Analysis of the *T* gene expression in T^{Wis} mutants suggested that the *T* protein is only required for its own expression and for mesoderm formation after day 8.0 (Herrmann, 1991; reviewed by Herrmann and Kispert, 1994; and especially by Beddington *et al.*, 1992). At this stage, electron microscopy analysis of early mutant embryos revealed defects in the structure of the node from which notochord precursor cells arise (Fujimoto and Yanagisawa, 1983). This observation suggested that the *T* product could play a more direct role in the fate of notochord precursors, independent of the phenotype observed in the primitive streak and allantois.

The Danforth's short tail mutation

The *Danforth's short-tail* (*Sd*) mutation is the only known mutant allele for *Sd* and it was isolated by C. H. Danforth (hence *Sd* for short-Danforth) from his stock of "posterior duplication" mutant mice (Danforth, 1930) and later described by Dunn *et al.* (1940). The *Sd* mutation is semi-dominant lethal but could become almost dominant lethal in the Bagg albino genetic background (Dunn *et al.*, 1940). Variations in the severity of the phenotype in heterozygous individuals have also been observed between the NMR1 and C57BL/6 strains (Dietrich *et al.*, 1993). The phenotype is mostly represented by defects in axial skeleton and urogenital and digestive systems in heterozygous and homozygous mice (Dunn *et al.*, 1940; Theiler,

1988; Gluecksohn-Schoenheimer, 1943). Some lethality occurs among heterozygous mice presenting excessive defects, whereas the phenotype in homozygous embryos is stronger and always lethal. In heterozygous mice, there can be no tail at all, a tail that may never exceed half the normal length or any intermediate length. The sacral region is shortened and occasionally the spine is crooked (scoliosis or lordosis), one or both kidneys may be missing or be smaller. Homozygous mice have no tail. Their spinal column is shorter and there are no vertebrae posterior to the second lumbar vertebra. Furthermore, they completely lack kidneys and have an imperforate anus, no rectum but have a genuine cloaca. The bladder, urethra and genital papilla may be reduced or missing. Their survival up to birth is not compromised by this array of defects and they can compete for suckling with their littermates. Nevertheless, they die from auto-intoxication within 24 hours of birth because of their incomplete excretory system. The *Sd* gene and the molecular alteration underlying the mutation are still unknown.

The earliest defect visible in both heterozygous and homozygous embryos is the breakdown of the notochord starting at day 9 of development. In heterozygous embryos, disruptions in the notochord first appear in the cervical and trunk region and then extend along the entire length of the notochord. By the stage E11.5, the notochord has almost completely disappeared. In homozygous embryos, the same disruptions occur but the notochord is also abruptly interrupted in the lower thoracic region and only forms discontinuous fragments in more posterior positions.

Expression analysis in T and Sd mutants

The *Brachyury* and *Danforth's short-tail* genes have not been assigned to any genetic pathway and no target gene for the *T* transcription factor has been identified yet. Therefore, the molecular defects at the basis of the *Brachyury* or *Sd* phenotypes are still unknown. There are not enough mutations or genes cloned to use a direct approach to unravel genetic pathways or target genes. However, cloned genes have been used as novel molecular and genetic markers to analyze the phenotype of the *T* (Herrmann, 1991; Rashbass *et al.*, 1994; Dietrich *et al.*, 1993; Conlon *et al.*, 1995) and *Sd* mutations (Koseki *et al.*, 1993; Dietrich *et al.*, 1993; Phelps and Dressler, 1993). The detection of abnormal expression pattern in specific tissues may help to understand the extent of alterations generated by these mutations. More importantly, it may help to identify early effects on gene expression and tissue patterning in mutant embryos. GT insertions now also

provide a growing collection of easily detectable genetic markers that can be used to analyze embryonic development.

I used the *GtC101* GT insertion of the *lacZ* reporter gene into *cordon-bleu* (*cobl*) as a genetic marker to analyze the early phenotypes resulting from mutations in the *Brachyury* and *Danforth's short-tail* genes. The *GtC101* insertion was previously used to assist in the analysis of the phenotype of targeted mutations in *HNF-3β* (Ang and Rossant, 1994), GTPase-activating protein (Henkemeyer *et al.*, 1995), SHP-2 (Saxton *et al.*, 1997) and Gli-2 (Ding *et al.*, *submitted*). *Cordon-bleu* is expressed in axial structures and in a variety of tissues whose formation or patterning are affected by the *Brachyury* and *Sd* mutations. Furthermore, the onset of *cobl* expression in the node precedes the appearance of morphological abnormalities for both mutations. The possible function of *cobl* is not known but its expression pattern in mice carrying the *GtC101* GT insertion provides a versatile and easily detectable marker to analyze the phenotype of mutations altering axis development and embryo patterning. The aims of this study are to document further the description of the *T* and *Sd* phenotypes and to analyze regulation of *cobl* expression in mutant embryos lacking a notochord.

MATERIALS AND METHODS

Mouse strains

Heterozygous [C57BL/10ScSn-*T*/+] mice (breeding pair with *T*/+ male: Jackson Laboratories) carrying the *Brachyury T* mutation (large deletion) were back-crossed into [129/Sv/CD1 *GtC101*/*GtC101*] outbred mice. *T*/+ short-tailed offspring were genotyped for *GtC101* by ß-Gal staining. Double heterozygous female mice were back crossed with [129/Sv/CD1 *GtC101*/*GtC101*] males to obtain [*T*/+: *GtC101*/*GtC101*] outbred mice. These mice were then maintained by back-crossing into the [129/Sv/CD1 *GtC101*/*GtC101*] line and by inter-crosses between [*T*/+: *GtC101*/*GtC101*] littermates.

[T^{Wis}/+/+; *GtC101*/*GtC101*] mice were derived by back crossing [129/Sv/cp:T^{Wis}/+] mice (kindly provided by A. Shedlovsky and W. Dove) into the [129/Sv/CD1 *GtC101*/*GtC101*] outbred mice.

[*Sd*/+; *GtC101*/*GtC101*] mice, carrying the *Danforth's short-tail* mutation, were derived from backcrosses between [C57BL/6By-*Re Sd* Va^{J} [NLR] N3] (breeding pair with *Sd*/+ female: Jackson Laboratories) and [129/Sv/CD1 *GtC101*/*GtC101*] mice.

The genotyping of the mice at the *cobl* locus was performed as described previously (Chapter III).

Brachyury T^{Wis} mutant genotyping

Before the late head-fold stage of development, embryos homozygous for the *Brachyury* mutations are not distinguishable by morphology alone from their littermates. Furthermore heterozygous embryos are not morphologically different from their wild-type littermates until E12.5 when tail growth is altered. At later stages, the embryos' morphology provided a good criterion to determine their genotype.

The sequence for the primers BGH-037 (5'ACGTTGCGAGCTGCTGCGGC3'), BGH-039 (5'ACCCATGTCAAACCCATCAG3') and BGH-052 (5'CCTATGCGGACAATTCATCTG3') was obtained from Dr. B. G.. Herrmann. The BGH-052/BGH-037 primer pair is specific to the T^{Wis} allele and amplifies a 150 bp DNA sequence: The BGH-052/BGH-039 primer pair amplifies a wild-type 200 bp DNA fragment absent in the T^{Wis} allele.

Extraembryonic membranes from each embryo were lysed o/n at 55°C in 100 µl of non-ionic lysis buffer: 50 mM KCl, 10 mM Tris.HCl (pH 8.3), 2 mM $MgCl_2$, 0.1 mg/ml gelatin (Sigma), 0.45% Nonidet P-40, 0.45% Tween-20, supplemented with Proteinase K at 100 µg/ml and stored at -20°C. One to 5 µl of each sample was

denatured at 94°C for 10 minutes and used for a PCR assay. Template denaturation was done at 94°C for 1 minute, annealing at 60°C for 30 seconds and elongation at 72°C for 1 second in 35 cycles. Samples were analyzed after electrophoresis through a 2.5% (1.5% agarose/1% Nusieve® agarose) gel.

Genotyping Danforth's short-tail

The molecular nature of the *Sd* gene and mutation are not known. Therefore, no molecular assay is available to genotype embryos at the *Sd* locus. Hence, genotyping relied solely on embryo morphology and on the *GtC101* pattern of expression. From day E8.5, I found anomalies in axial expression that seemed to represent the homozygous class, but no further discrimination was possible at this stage. At E9.0 however, expression in notochord presents large gaps in *Sd/Sd* embryos and is fragmented and branching with the neural tube or the gut in *Sd/+* embryos. At later stages, differences in expression are more conspicuous and morphological defects also become apparent. Overall, *GtC101* expression domains provided good criteria to differentiate heterozygous, homozygous and wild-type littermates at stages when morphological defects are not yet visible.

Beta-galactosidase detection in whole mount embryos

The protocol followed was slightly different from the procedure used in Chapter II or III. Embryos were dissected from the uterus in cold PBS (supplemented with Calcium, $CaCl_2.2H_2O$ at 0.133g/liter, and Magnesium, $MgCl_2.6H_2O$ at 0.1g/liter) and separated from their extraembryonic membranes. When it was necessary to use the extraembryonic tissues for genotyping, embryos were dissected in individual drops and kept separated throughout the entire staining procedure.

Embryos were fixed in PBS containing 0.2% gluteraldehyde (Fisher) or 3.7% formaldehyde for approximately 5 to 30 minutes according to the stage of development: E7.5 (5 min.) E8.5 (10 min.), E9.5 (15 min.), E10.5 & E11.5 (30 min.), E12.5 or older (2 hours). Embryos were then washed 3 times in Wash buffer (0.1M sodium phosphate, 2 mM $MgCl_2$, 0.2% NP-40) and stained o/n in the dark at 37°C in X-gal buffer (Wash buffer containing 0.5 mg/ml X-gal (Vector Biosystems), 5 mM $K_3Fe(CN)_6$ and 5 mM $K_4Fe(CN)_6$). Embryos were post-fixed in Wash buffer containing 3.7% formaldehyde and preserved in the same buffer at 4°C.

Whole-mount alizarin red/alcian blue bone and cartilage staining

Embryos were dissected out on embryonic day 17. After removal of extraembryonic tissues and viscera, embryos were fixed in 95% Ethanol o/n in scintillation vials. The skin was removed before incubating in alcian blue (Sigma)/acetic acid/ethanol (150 mg/L in 1 vol./4 vol. of glacial acetic acid/95% ethanol) for 48 hours. Embryos were then rinse in 95% ethanol for a few hours and incubated in 2% (w/v H_2O) potassium hydroxide (KOH) for 24 hours. The solution was replaced by 1% KOH containing 75 mg/L alizarin red-S (Sigma) for 24 hours. Embryos were cleared in 20% glycerol (v/v) 1% KOH for a week, changing the solution daily and finally transferred to 50% glycerol/50% ethanol for photography and storage.

Clarification of the embryos

Embryos were dehydrated through serial washes in Methanol at 25%, 50%, 75%, 80%, 95% twice and 100% methanol and then cleared in benzyl alcohol : benzyl benzoate (1:2) just prior to photography.

RESULTS

The GT insertion into the cobl gene provides a novel genetic marker to study axis development and embryo patterning

No phenotype has been observed to date in embryos and mice heterozygous or homozygous for the *GtC101* GT insertion (see Chapter 3). It is therefore a viable marker that should not interfere with embryonic development in control or mutant embryos. The *E. coli lacZ* reporter gene has been shown to be innocuous during embryonic development and the enzymatic activity of its product is easily detectable *in vivo*. Expression of *cobl* during early embryogenesis (described in Chapter 3) makes it a suitable marker to analyze the phenotype of notochord developmental mutants.

When introduced into the *T* or *Sd* mutant strains, *GtC101* caused no variation from the expected ratio of embryos from [*T*/+ x *T*/+] or [*Sd*/+ x *Sd*/+] intercrosses (see Table 13 A & B, 13 A and 14), regardless of whether the mice were heterozygous or homozygous for *GtC101*. Expression of *cobl* persisted in mutant embryos thus permitting its use as a marker. Unless specified otherwise, intercrosses between mice homozygous for the *GtC101* allele and heterozygous for one of the mutant allele were used to produce control and mutant embryos. Therefore, the embryos described here are always homozygous for *GtC101* and the terms wild-type, heterozygous or homozygous usually refer to the genotype at the *T* or *Sd* loci.

Expression of cobl in the axis of mice carrying the T and T^{Wis} mutant alleles of the Brachyury gene

The expression of *cobl* was analyzed between E7.5 and E11.5 (see Table 13 A,B) in embryos carrying the original *T* deletion (Dobrovolskaïa-Zavadskaïa, 1927) or the T^{Wis} allele (Shedlovsky *et al.*, 1988). *T* is a complete deletion of the *Brachyury* gene and T^{Wis} allele encodes a truncated protein. Because the phenotype of T^{Wis} heterozygous and homozygous mutants is more severe than that of *T* mutants, it has been suggested that the T^{Wis} truncated protein may interfere with the wild-type protein in heterozygous and with other peptides normally interacting with *T* in homozygous embryos (Herrmann, 1991). I found the same phenotypic differences observed previously between the 2 alleles, i.e. lesser development of the somites

and tail bud. Expression in extraembryonic endoderm was not altered by these mutations.

Table 13.
No. of control and *Brachyury* mutant embryos analyzed for *cobl* expression

A: Embryos collected from [*T*/+] intercrosses[1]

Stage	# embryos (# litters)	No. embryos : % of total +/+	*T*/+	*T*/*T*	n.d.
E7.5	35 (6)	28	79%	6 18%	1 3%
E8.5	43 (5)	13 30%	18 42%	12 28%	
E10.5	18 (2)	3 17%	9 50%	6 33%	
Total [2]	96	16 26%	27 44%	24 25%	1 1%

B: Embryos collected from [T^{Wis}/+] intercrosses[1]

Stage	# embryos (# litters)	No. embryos : % of total +/+	T^{Wis}/+	T^{Wis}/T^{Wis}	n.d.
E7.5	48 (6)	8 17%	23 48%	15 31%	2 4%
E7.5 *GtC101*/+	17 (2)	6 35%	9 53%	2 12%	
E8.5	20 (2)	5 25%	10 50%	3 15%	
E9.5	14 (2)	3 21%	9 64%	2 14%	
E10.5	27 (3)	7 26%	12 44%	7 26%	1 1%
E11.5	8 (1)	2 25%	5 63%	1 13%	
Total [2]	134	31 23%	68 51%	30 22%	3 2%

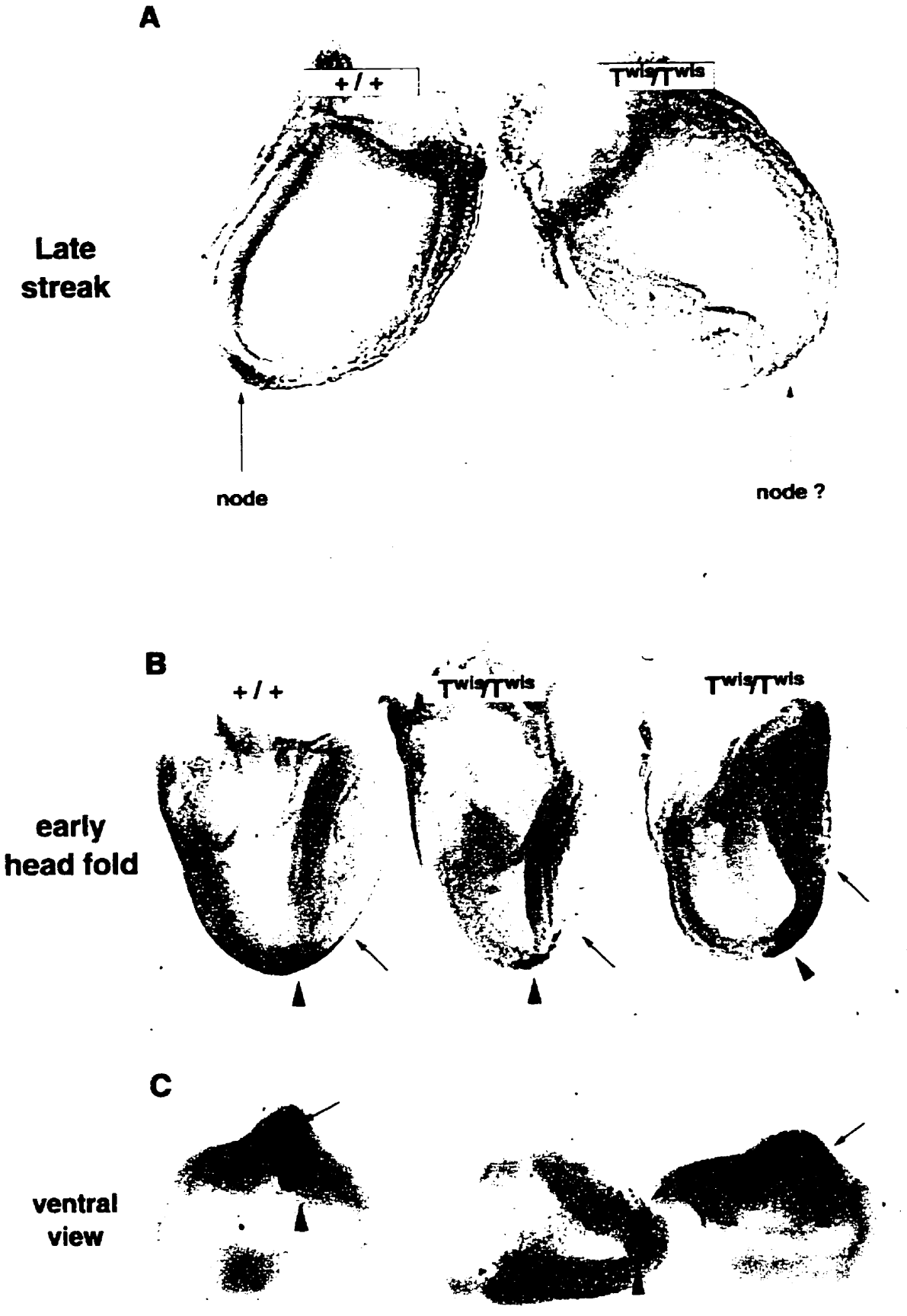
[1] All embryos were *GtC101*/*GtC101* carrier, unless specified otherwise.
[2] Another 17 [*T*/+; *GtC101*/+] and 30 [T^{Wis}/+; *GtC101*/+] embryos between E10.5 & E14.5 were generated when testing [*T*(or T^{Wis})/+; *GtC101*/*GtC101*] males by breeding with wild-type CD1.

Cobl onset of expression

It has been previously reported that the node of *T*/*T* embryos has an abnormal morphology at the early head fold stage (Fujimoto and Yanagisawa, 1983). Since *cobl's* expression appears before the first morphological defects reported for *T* or

Fig. 18. - Expression of *cobl* in the node of *Brachyury* mutant embryos.

Expression of *cordon-bleu* (*cobl*) as revealed by detection of the *COBL*/ß-Gal fusion protein activity in the node of control and mutant T^{Wis}/T^{Wis} embryos at the late streak (A) and early head fold stage (B, C). The genotypes of the embryos are indicated in the figure. (A) Lateral view with anterior to the left. Expression in the node of the wild-type embryo (left arrow) is already strong in the node whereas it is much reduced in a slightly older mutant embryo (right arrow). (B) Lateral view with anterior to the right. The node (arrowheads) of the wild-type embryo is a clear morphological structure at the tip of the embryo. In the 2 mutant embryos, this structure appears missing and the expression area of *cobl* is consequently reduced. (C) Ventral view of the embryos shown in (B) with anterior to the top. Fewer cells expressing *cobl* are present in the region of the node (arrowheads) in mutant embryos, and they do not outline the node as a distinct tissue as in wild-type embryos. Expression in the anterior midline is present in all embryos (arrows in B and C). However, in the two T^{Wis}/T^{Wis} embryos, it did not condense in the axis proper and probably represents mostly endoderm expression.


A
Late streak
+ / +
Twis/Twis
node
node ?
B
early head fold
+ / +
Twis/Twis
Twis/Twis
C
ventral view

T^{Wis} homozygous mutants. I analyzed ß-Gal activity in the node of E7.5 control and mutant embryos from intercrosses between [*T*/+; *GtC101*/*GtC101*] mice or [T^{Wis}/+; *GtC101*/*GtC101*]. A phenotype could be detected in all homozygous E7.5 embryos (n=15). At the late streak stage, no difference could be detected between wild-type and heterozygous embryos at the onset of *cobl* expression. In homozygous embryos however, *cobl* expression was delayed or reduced in the node region (Fig. 18A). Nevertheless, weak expression was still detectable (Fig. 18A), suggesting that the *T* transcription factor is not strictly required for *cobl* expression. I observed the same early defect of *cobl* expression in the node of T^{Wis}/T^{Wis} embryos (n=2) from 2 litters heterozygous for the *GtC101* insertion (data not shown). Therefore, the alteration of *cobl* expression is likely to be due solely to the effects of the T^{Wis} mutation. At E7.5, 20% (n=7) of the embryos from [*T*/+; *GtC101*/*GtC101*] parents showed altered expression of *cobl* in the node (data not shown). However, I did not genotype embryos carrying the *T* allele because genotyping by PCR is not as conclusive as for the T^{Wis} allele.

At the head fold stage, *cobl* expression anterior to the node appeared in all embryos but axial condensation was less prominent in homozygous embryos (Fig. 18 B & C). Most of the ß-gal activity appeared to come from the definitive endoderm, consistent with later expression in the gut and the failure of notochord precursors to populate the anterior midline. Fewer *cobl* positive cells were detected in the node region of homozygous embryos, and they appear scattered (Fig. 18C). The node size was also considerably reduced at this stage when compared with wild-type embryos (Fig. 18 B). It corresponds to the stage at which morphological defects were previously identified in the node by electron microscopy (Fujimoto and Yanagisawa, 1983). It is not clear whether the node is completely missing or if a sub-population of node cells is absent or incorrectly patterned. This deficiency in *cobl* expressing cells represents the earliest defect detected in *Brachyury* mutant embryos.

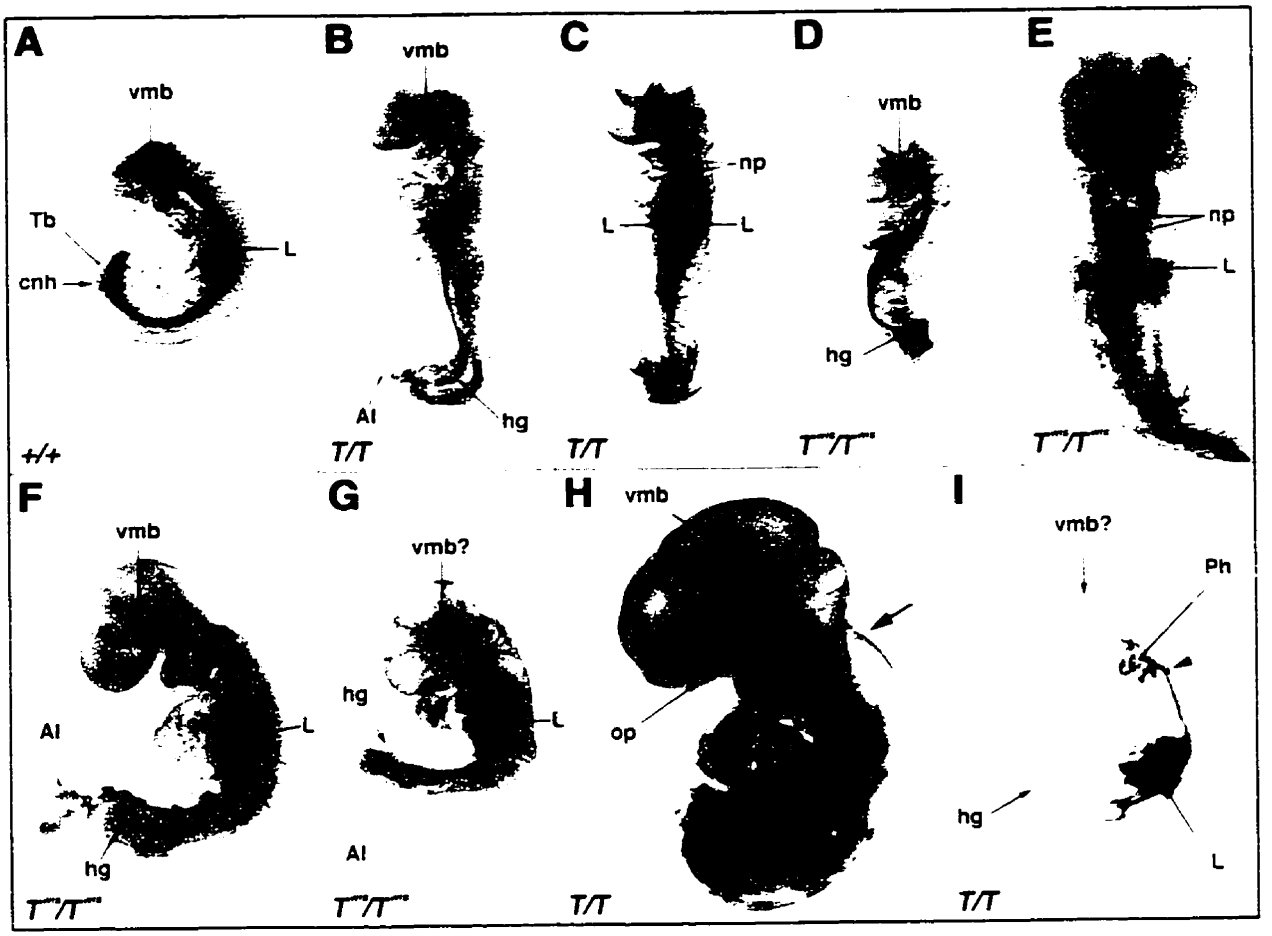
Cobl expression in the axis

At E8.5 (n=3 T^{Wis}/T^{Wis}; n=12 *T*/*T*) *cobl* expression was restricted to anterior structures and dorsal gut endoderm (Fig. 19 A-E). In some embryos, groups of cells resembling notochord segments were also observed in the anterior midline of both *T*/*T* and T^{Wis}/T^{Wis} embryos (Fig. 19 C and E). The node itself, as detected by *cobl* expression in wild-type littermates (Fig. 18), was either totally absent or smaller and of abnormal shape. It was also found at a more anterior position when in

Fig. 19. - Expression of *cobl* in the axis of *Brachyury* mutant embryos.

Expression of *cordon-bleu* (*cobl*) as revealed by detection of the COBL/ß-Gal fusion protein activity. The genotypes are indicated in the lower left corner. (A-E) 10 somite stage 8.5 dpc embryos. (A) Wild-type embryo presenting the full *cobl* expression pattern. The induction of the liver (L) and the chordoneural hinge (cnh) are distinct. (B, C) Lateral and ventral views of a T/T embryo. The development of the posterior trunk and allantois are much reduced. Most of the detectable expression is in the gut, liver (L), ventral midbrain (vmb). Condensations of cells (possibly notochord precursors) are also visible in the anterior axis (np in C). (D, E) Lateral and ventral view of a T^{Wis}/T^{Wis} embryo (E - higher magnification after removal of the heart). Although posterior development of this embryo is more severely affected, expression remains similar. (F, G) 9.5 dpc embryos. Expression is maintained in the gut endoderm and developing liver. Expression in the ventral midbrain was sometimes missing (G). (H) 11.5 dpc embryos at the time of embryonic death. Posterior structures are completely missing but expression of *cobl* remains in anterior structure. Expression has also appeared at an appropriate time on surface ectoderm dorsal to the hindbrain (big arrow) and in the anterior olfactory placode. The forelimb is present but small and has an abnormal morphology (white asterisk). (I) Cleared 10.5 dpc embryo. Expression was only seen in gut endoderm and liver. Some cells dorsal to the gut could represent remaining notochord precursors (arrowhead).

Abbreviations: Al: allantois; cnh: chordoneural hinge; hg: hindgut; L: liver; np: notochord precursors; op: olfactory placode; Ph: pharyngeal pouches; Tb: tailbud; vmb: ventral midbrain.


A
vmb
Tb
L
cnh
+/+
B
vmb
Al
hg
T/T
C
np
L
L
T/T
D
vmb
hg
T⁻/T⁻
E
np
L
T⁻/T⁻
F
vmb
Al
L
hg
T⁻/T⁻
G
vmb?
hg
L
Al
T⁻/T⁻
H
vmb
op
T/T
I
vmb?
Ph
hg
L
T/T

control embryos it is moving posteriorly with the regressing streak. Expression of *cobl* was not detected in the ventral midline of the neuroepithelium, consistent with the inability of mutant notochord to induce floor plate. However, expression of *cobl* could still be detected in midline ventral midbrain cells (Fig. 19 B & D). Induction of expression in the forming liver was timely and comparable in level with control embryos (Fig. 19 A-E).

At E9.5 (n=2 T^{Wis}/T^{Wis}), expression of *cobl* remains mostly restricted to the gut endoderm, the liver and ventral midbrain (Fig. 19 F-I). Homozygous embryos could be collected up to the stage of embryonic death, between E10.5 and E11.5 (n=8). Anterior structures appear normal but development of the forelimb and posterior tissues was always altered in homozygous embryos. Expression of *cobl* in the anterior dorsal neural tube also appeared normally at E11.5 (Fig. 19 H). Expression in the gut remained normal but the hindgut structure was greatly altered by the lack of posterior development. Expression in the ventral midbrain could be normal (Fig. 19 H), whereas in some embryos only weak (Fig. 19 F) or no expression (Fig. 19 G & I) could be detected. Overall, expression in all tissues was weaker in T^{Wis}/T^{Wis} embryos. The lack of somites and the posterior truncation did not permit further analysis of *cobl* expression in homozygous embryos. The structure of the trunk notochord was also altered in 2 $T^{Wis}/+$ embryos. One embryo showed interrupted axial expression in the trunk (Fig. 23 J) and another, branching of the notochord with the gut (Fig. 21 R). Abnormal expression of *cobl* in the somites of other heterozygous embryos further suggests the occurrence of trunk notochord defects (see below and Fig. 23).

Expression of cobl in the axis of mice carrying the Danforth's short-tail mutation

The GtC101 insertion does not alter the Sd phenotype

In order to confirm that the *GtC101* insertion did not interfere with the *Sd* phenotype, I analyzed the penetrance of some skeletal defects in mutant embryos with or without the *GtC101* insertion. I dissected 6 litters from [*Sd*/+; *GtC101*/*GtC101*] parents and 7 litters from [*Sd*/+; +/+] parents at E17.5 (Table 14 A). At this stage, morphological defects alone (e.g. tail and body length) indicate the embryos' genotype. I processed all the homozygous and subset of heterozygous and wild-type embryos for bone and cartilage staining. I scored embryos for their number of ribs, vertebrae -with and without an ossified centrum, and neural arches.

Because heterozygous mice often develop only one kidney. I scored the number of kidneys in sacrificed *Sd/+* adult females as a criterion for the penetrance of the phenotype. All mutant embryos had no visible nucleus pulposus (notochord-derived) in intervertebral discs and no dens on the axis vertebra (C2). All heterozygous embryos had 13 ribs and an ossified centrum in the atlas (C1).

I detected no significant differences in the expressivity of the phenotype between the two groups (with or without the *GtC101* insertion) (see Table 14 B). Furthermore, the numbers are comparable to previous observations (Koseki *et al.*, 1993; Dietrich *et al.*, 1993).

Table 14.

Penetrance of the *Sd* phenotype in *GtC101/GtC101* mice

A: No. of embryos analyzed for skeletal defects[1]

			No. embryos : (% of total)		
Stage	*GtC101 locus*	# embryos (# litters)	+/+	*Sd/+*	*Sd/Sd*
E17.5	+/+ [2]	49 (7)	13 27%	22 45%	14 29%
"	*GtC101/GtC101*	38 (6)	7 18%	21 55%	10 26%
Total		87	20 23%	43 49%	24 28%

B: Comparison between skeletal and urogenital defects[3]

Sd locus	*Sd/Sd*		*Sd/+*		+/+
cobl locus	*GtC101/GtC101* (n=10)	+/+ (n=13)	*GtC101/GtC101* (n=17)	+/+ (n=22)	
Ribs	11.2 ±1.4	11.9 ±2.0	13 ±0	13 ±0	13 ±0
Vertebrae	18.9 ±3.2	20.0 ±1.8	30 ±2.5	33 ±2.4	61 ±2.0
Centrum	n.d.	n.d.	21 ±3.0	22.5 ±4.4	25 ±3.0
Neural arches	22.0 ±4.0	24.6 ±1.0	30 ±1 9	32.5 ±1.7	n.d.
Missing kidney[4]	N.A.	N.A.	7 (22%) (n=32)	2 (22%) (n=9)	n.d.

1 Alcian Blue (cartilage) and Alizarin Red (bone) staining.

2 The [*Sd/+; +/+*] parents were derived by backcrossing [*Sd/+; GtC101/GtC101*] mice into the 129 Sv/cp background.

3 Mean value and standard deviation are indicated.

4 Analyzed in adult *Sd/+* females.

Abbreviations: n= sample size; N.A. not applicable; n.d. not determined.

Cordon-bleu expression in the axis

Expression of *cobl* was analyzed between E7.5 and E12.5 in control and mutant embryos of [*Sd*/+; GtC101/GtC101] parents (see Table 15). The embryos' genotype at early stages (up to day E10.5) could only be inferreed from the pattern of *cobl* expression or at later stages from morphological traits. However, the ratios observed and the correlation between *cobl* expression and morphology proved that GtC101 is a reliable marker to classify embryos into genotypic groups.

At E7.5, in 45 embryos from 6 different litters, no change in *cobl* expression could be detected. At E8.5, 70 embryos from 10 different litters were examined. Before the 10 somite stage and turning of the embryo, no distinction could be made between control and mutant littermates (38 embryos). However, in some embryos the notochordal plate *cobl* expression appeared reduced in width or discontinuous and did not show the homogenous structure observed in wild-type litters (data not shown).

Table 15

No. of control and *Sd* embryos analyzed for *cobl* expression[1]

Stage	# embryos (# litters)	No. embryos : % of total							
		+/+		*Sd*/+		*Sd*/*Sd*		n.d.	
E7.5	40 (6)	n.d.[2]							
E8.5 < 10 somites	38 (5)	n.d.[2]							
E8.5 > 10 somites	32 (5)	**26**	81%[3]			**6**	19%	**1**	3%
E9.5	52 (6)	**10**	19%	**32**	62%	**10**	19%		
E10.5	40 (5)	**6**	15%	**24**	60%	**9**	23%	**1**	3%
" *GtC101*/+	31 (4)	**6**	19%	**19**	61%	**5**	16%	**1**	3%
E11.5	12 (3)	**3**	25%	**7**	58%	**2**	17%		
E12.5	9 (1)	**3**	33%	**3**	33%	**3**	33%		
Total	254	**28**	19%	**85**	59%	**35**	20%	**3**	%

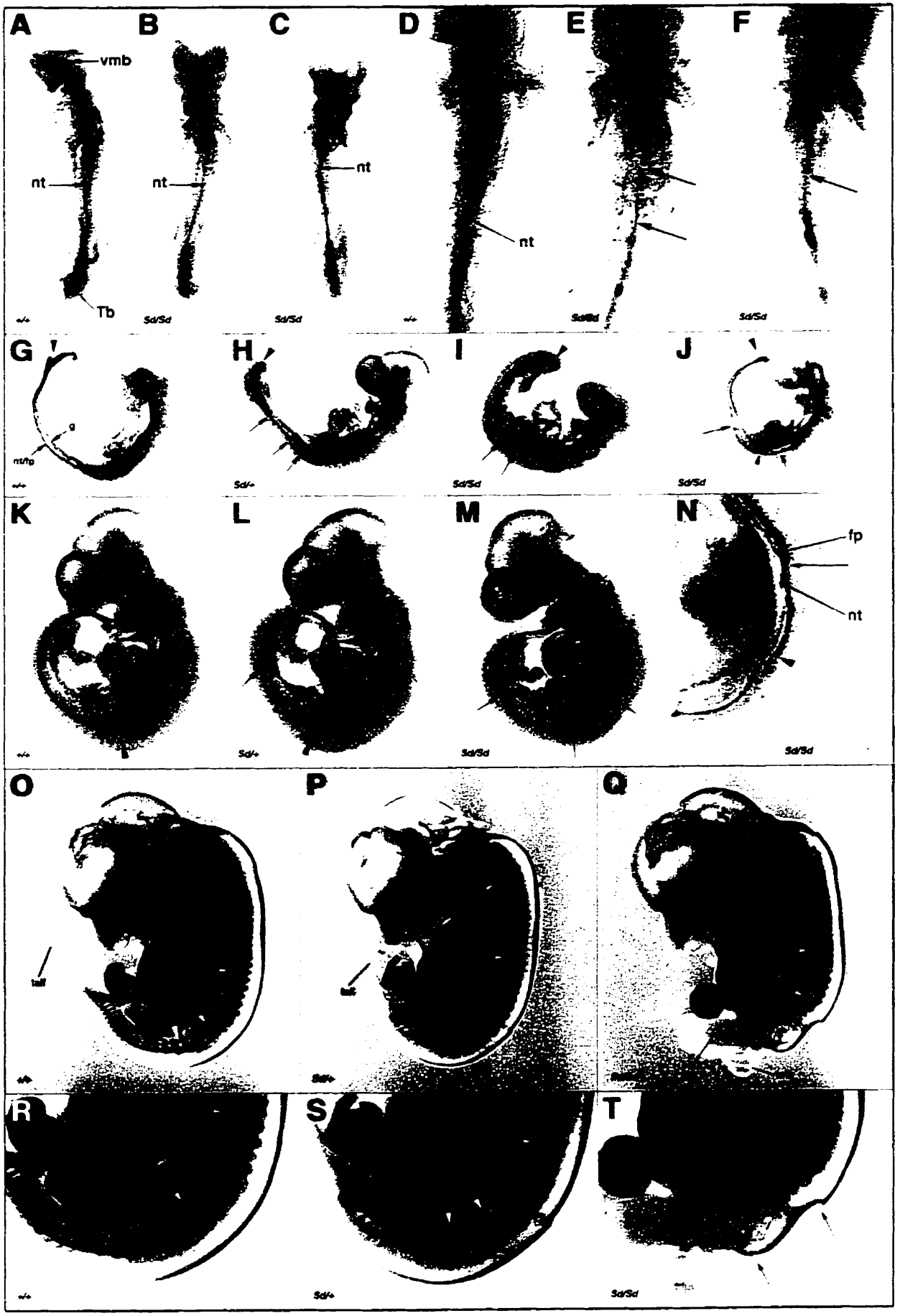
[1] All embryos were *GtC101*/*GtC101* carrier, unless specified otherwise.
[2] All embryos displayed similar staining patterns.
[3] No difference was found between wild-type and heterozygous embryos.

In 32 more advanced E8.5 embryos (from 10 somites stage), and especially after embryos had started turning, the structure of the notochord was clearly altered (n=6, Fig. 20 A-F). *Cobl* expression along the axis of these embryos was

Fig. 20. - Expression of *cobl* in the axis of *Danforth's short-tail* mutant embryos.

Expression of *cordon-bleu* (*cobl*) as revealed by detection of the COBL-ß-Gal fusion protein activity. The genotypes are indicated in the lower left corner. (A-F) 10 somite stage 8.5 dpc embryos. The notochord (nt) is poorly condensed and does not form a clear structure in candidate *Sd/Sd* embryos (B, C). (D-F) Higher magnification of (A-C): note the gaps in axial staining (arrows in E and F). (G-J & N) 15 somite stage 9.0 dpc embryos. (G) wild-type embryo with a well developed chordoneural hinge (cnh) and a clear separation between the gut (g) and the notochord/floor plate (nt/fp) (arrows). (H) *Sd/+* embryo with incomplete expression in the cnh (arrowhead) and an abnormal notochord, branching with the gut (arrows). (I) *Sd/Sd* embryo presenting the same defects as *Sd/+* embryos; in addition, expression in the notochord and floor plate are interrupted (arrows). (J, N) Cleared *Sd/Sd* embryo. (J) The notochord (arrows) and the floor plate (arrowhead) are interrupted. (N) Higher magnification of embryo in (J) showing diminution (arrow) and interruption (arrowhead) of floor plate induction consistent with more severe defects in the notochord. (K-M) 9.5 dpc embryos from the same litter. (L) Lower levels of floor plate induction in the lumbar region where the notochord branches with the gut (arrow). Expression in the somites first appears (arrowheads in K, L). (M) The notochord/floor plate staining is interrupted (arrows); no somite expression is detectable. (O-T) 12.5 dpc embryos. (O,R) Strong *cobl* expression in the posterior notochord (white arrowhead in R) and in intervertebral discs (iv discs, arrows in O and R). (P, S) The notochord is highly disorganized (arrowheads in S) and seems ectopic when compared to iv discs (arrows); the iv discs are missing in the cervical/thoracic region (between arrows in P) and smaller in the thoracic/lumbar region (white arrow in S). (Q,T) No notochord was detected; the body axis is truncated in the thoracic lumbar region and the neural tube became kinked and folded ventrally in the lumbo-sacral region (arrows in T); none or little iv disc expression is detectable in the cervico-thoracic region (between arrows in Q).

Abbreviations: cnh: chordo-neural hinge; fp: floor plate; g: gut; HL: hindlimb; nt: notochord; Tb: tailbud; vmb: ventral midbrain.


A
B
C
D
E
F
vmb
nt
nt
nt
nt
Tb
+/+
Sd/Sd
Sd/Sd
+/+
Sd/Sd
Sd/Sd
G
H
I
J
g
nt/fp
+/+
Sd/+
Sd/Sd
Sd/Sd
K
L
M
N
fp
nt
+/+
Sd/+
Sd/Sd
Sd/Sd
O
P
Q
tail
tail
+/+
Sd/+
R
S
T
+/+
Sd/+
Sd/Sd

weaker and appeared disorganized. A clear reduction in the density of axial staining in the middle trunk region prefigured the later notochord defects (Fig. 20 E & F). Cross sectioning of these embryos revealed abnormal *cobl* expression in the three germ layers. Expression in the roof of the gut was not continuous along the A-P axis. The notochord was interrupted or some segments did not express *cobl*. Induction of *cobl* expression in the floor plate was very reduced including levels adjacent to notochordal cells (data not shown).

Between the stage E9.0 and E9.5, wild-type, heterozygous and homozygous mice could be easily identified after analysis of *cobl* expression (Fig. 20 G-N). In heterozygous embryos, the notochord extends along the entire axis and expresses *cobl* from head to tail bud, but its structure is very abnormal. It is interrupted by constrictions and often branching ventrally to contact the gut. The constrictions are localized to anterior notochord and posterior to the thoracic region, whereas branching occurred more frequently and at any position along the notochord. The notochord posterior end, in the tail bud's chordoneural hinge, is reduced. The induction of the floor plate, although uneven and weaker than in wild-type littermates, has occurred along most of the axis. Differences in the penetrance of the phenotype could be observed between individual embryos. Homozygous embryos were identifiable from stage E9.0 (Fig. 20 I) by loss of expression in the notochord and the adjacent floor plate. Wherever present, the notochord was irregular and branching with the gut or the neural tube as in heterozygous embryos (Fig. 20 M & N). It was also abruptly interrupted at thoracic levels and, in more posterior regions, presented discontinuous fragments which nevertheless reached the tail bud. The notochord bud of the tail bud was generally reduced or absent in heterozygous embryos and absent in homozygous embryos (see below). The induction of the floor plate also ended at the level of anterior notochord truncation (Fig. 20 J, M & N). Posterior expression of *cobl* in ventral neural tube was only distinct in some areas overlying notochord (Fig. 20 M).

At E10.5, morphological defects of the tail became a clear indicator for homozygous embryos (n=9). Their tail is shorter, thinner and often kinked; embryos are also shorter and often had back lordosis (Fig. 23 C-E). In homozygous embryos, almost no notochord expression was detectable except in the hindlimb and tail region. Anterior notochord was absent or had become incorporated into the forming vertebrae and expression in the floor plate barely extended beyond the liver anlage. Some embryos had a less severe phenotype and showed weak axial expression in trunk and lumbar regions (Fig. 23 C). Defects in the somitic

patterning of homozygous embryos became evident at this stage (see below). The notochord defects observed at E9.5 persisted in heterozygous embryos.

Between stages E11.5 and E12.5, the tail of heterozygous embryos became thinner than in wild-type littermates and the genotype of all embryos could be determined without doubt. The tail in homozygous embryos (n=5) had become very thin and did not express *cobl*. At these stages, axial expression in wild-type embryos has become restricted to the forming intervertebral discs and the notochord disappears within forming vertebrae in an anterior-posterior gradient (Fig. 20 O.R). Expression was absent from anterior and diminished in posterior intervertebral discs of heterozygous embryos (Fig. 20 P). In the lumbar region, only small fragments of notochord could be identified but they were not located between vertebral segments (Fig. 20 S). In homozygous embryos, only a few and weak thoracic intervertebral discs were visible and the notochord is totally absent (Fig. 20 Q.*T*). The numbers of intervertebral discs detected (determined by their most hindmost level of expression) correlates with the numbers of vertebrae counted in E17.5 in *Sd/+* and *Sd/Sd* embryos skeletal preparations (see Table 15 B). Dorsal expression in neural tube of all embryos was comparable (Fig. 20 O.P.Q).

Expression of cobl in the tail bud of Brachyury and Sd mutant embryos

The nature of the tissues composing the vertebrate tail bud has long been a controversial subject among developmental biologists. Holmdahl (1925; 1939) proposed that the tail bud was made of an undifferentiated blastomere but Pasteels (1939; 1943) argued that it was highly organized and that the different germ layers were organized around the chordoneural hinge (CNH) structure (Pasteels, 1943). Holmdahl's model suggested that the three germ layers of the tail arose *de novo* by a process named "secondary body development" and Pasteels proposed that the morphogenesis of the tail was in direct continuity with gastrulation ("primary body development") and that the CNH was the perpetuation of the primitive streak. The former theory still receives support (Griffith *et al.*, 1992), but more recent data now favours the latter model. The existence of the chordoneural hinge structure is supported by the domains of *XBrachyury* and *Xnotl* expression in the frog tail bud (Gont *et al.*, 1993) and by cell lineage analyses in avian and mouse tail buds (Catala *et al.*, 1995; Wilson and Beddington, 1996). The vertebrate CNH probably represents the continuation of the node/primitive streak into the tail as it has also been shown to have axis inducing activities (Gont *et al.*, 1993; Catala *et al.*, 1995). *Cobl* is

expressed continuously in axial structures of the trunk and tail as the latter grows posterior to the allantois from E8.5. Interestingly, *cobl* expression also marks Pasteels's chordoneural hinge structure in the tail bud, and thus supports arguments depicting the organization of the tail bud as a continuation of trunk axial structures.

The morphogenesis of the tail bud is affected in heterozygous *Brachyury* and *Sd* embryos, and in homozygous *Sd* embryos. As a result, mutant mice are born with a shortened tail or no tail at all. I analyzed *cobl* expression in control and mutant tails to investigate the effects of the *Brachyury* and *Danforth's short-tail* mutation on CNH and tail morphogenesis.

Expression of cobl in the tail bud of Brachyury heterozygous embryos

On the basis of their morphology, *T*/+ or T^{Wis}/+ embryos are indistinguishable from their wild-type littermates before E11.0. However, from the formation of the tail bud at stage E8.5, I detected abnormal *cobl* expression in the CNH of T^{Wis}/+ and *T*/+ embryos (Fig. 21 A-G). The notochord bud (its posterior end) is smaller in mutant embryos (Fig. 21 A-E). Moreover, the lateral width of the CNH is reduced (Fig. 21 F, G), and axial expression of *cobl* did not extend to the tip of the tail bud as in wild-type embryos (Fig. 21 A-E). At this stage, it is not clear whether mis-expression posterior to the CNH is a primitive streak or a hind gut defect. The persistence of these defects at E9.5, suggested that it was not a mere developmental delay in heterozygous embryos, but a reduction of the notochord bud and tail gut structures illustrating incomplete CNH formation (Fig. 21 H-L). In the most severe cases, expression in the tail is limited to weak axial staining no longer outlining any CNH structure (Fig. 21L). The lack of expression in the tail tip appeared more clearly as a gut defect at this stage (Fig. 21 I & L).

From day 10.5, differences became apparent between the tails of *T* + and T^{Wis} + embryos. *T*/+ embryos grow a tail to different length, varying from no tail at all to a shorter tail (up to 3/4 of normal length); T^{Wis}/+ embryos however, always develop into tailless mice. In *T*/+ embryos, the notochord stopped at any position from the lumbar region to the tip of the tail (Fig. 21 N-P). When the notochord extended to the tail bud however, *cobl* expression was never found in the notochord bud of the CNH unlike in wild-type littermates (Fig. 21 M). The tail gut appeared normal and always extended to the CNH even in the absence of a notochord (positive cells at the tip of the tail in Fig. 21 O,P). In all T^{Wis}/+ embryos, the notochord never extended

Fig. 21. - Expression of *cobl* in the tail bud of *Brachyury* heterozygous mutant embryos.

Expression of *cordon-bleu* (*cobl*) as revealed by detection of COBL/ß-Gal fusion protein activity. Genotypes are indicated in the lower left corner. All figures compare expression in the notochord posterior tip (or notochord bud, nb, indicated by an arrow) and in the hindgut/tailgut (indicated by an arrowhead) between wild-type and heterozygous mutant tail buds. (A-C) 10 somite stage 8.5 dpc embryos (anterior at the top and dorsal to the right): in mutant embryos, the nb is reduced in the chordoneural hinge (cnh) and the tailgut does not reach the tip of the tail bud (B, C, E). (D-G) Enlarged lateral (D, E) and dorsal (F, G) views of the tail buds from embryos in (Fig. 19 A) and (C): the dorsal view reveals the narrower cnh of mutant embryos (G). (H, I) 15 somite stage E9.5 dpc embryos: the defects observed at 8.5 dpc remained in the tail bud. (J-L) Higher magnification of the tail bud of 25 somite stage 9.5 dpc embryos just before closure of the posterior neuropore: only the nb is reduced in (K) and the tail gut appears missing in (L). (M-R) Tail buds from 10.5 dpc embryos. (M) In wild-type embryos, the notochord/floor plate and the tail gut are still distinct structures in the tail's cnh. (N-P) *T*/+ embryos' tail buds showing increasing degrees in the expressivity of the phenotype: the notochord/floor plate (arrows) may extend to various length in the tail bud. The tail gut appears to always extend to the tip of the tail bud (arrowheads). (Q and R) T^{Wis}/+ embryos' tail buds: the notochord always stops at the level of the rump and no more expression is detectable in the tail bud. The tail gut (arrowheads) is usually not detectable in the tail (Q) but rarely extended normally as in (R): the notochord sometimes branches with gut endoderm (asterisks in R). (S-U) Tail buds from 11.5 dpc embryos. (S) In the wild-type tail, the notochord and tail gut are still distinguishable in the cnh: expression in the tail somites has also been induced. (T,U) The tail and somites are resorbing posterior to the notochord tip (arrow) and where notochord is missing (small arrows).

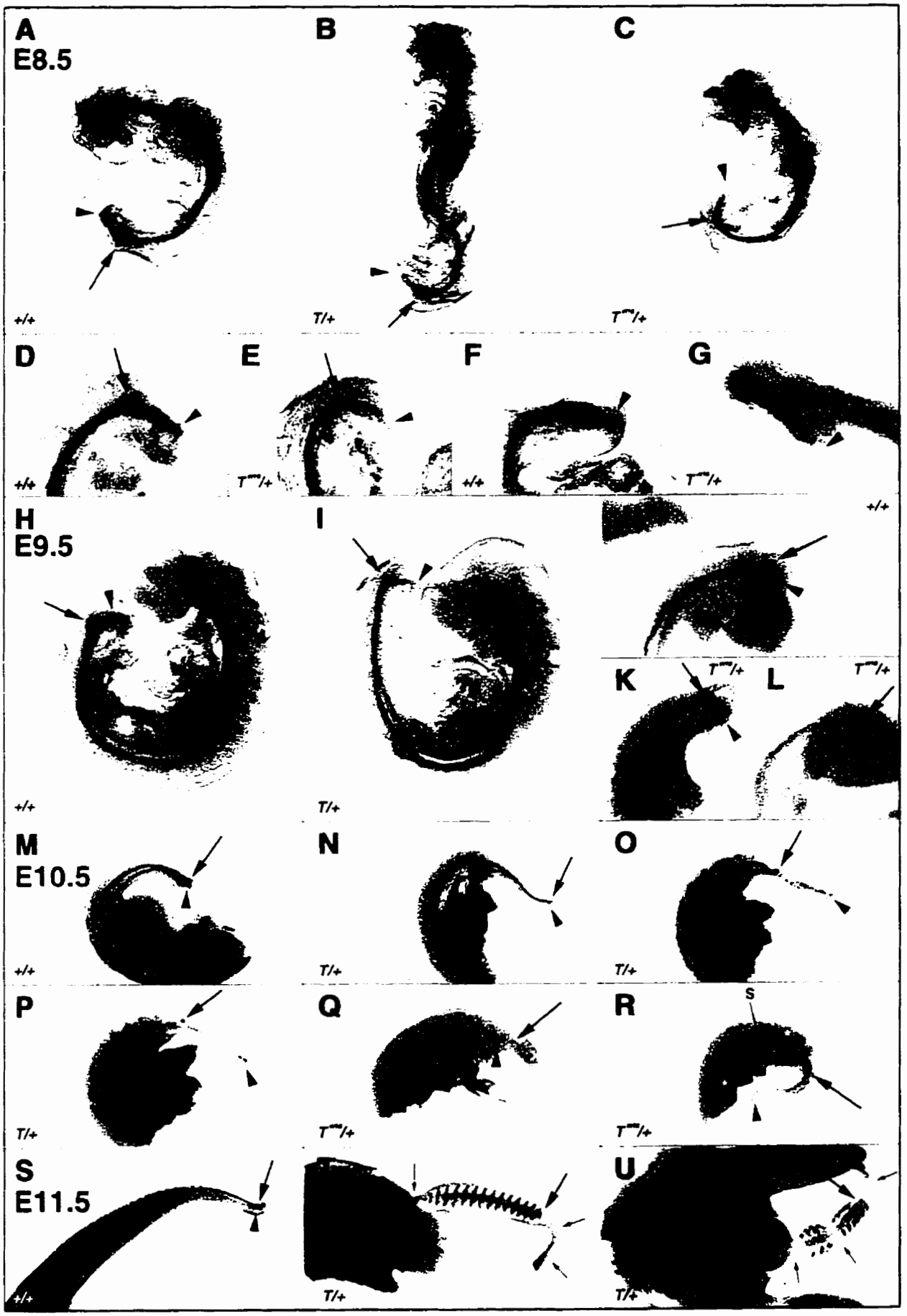
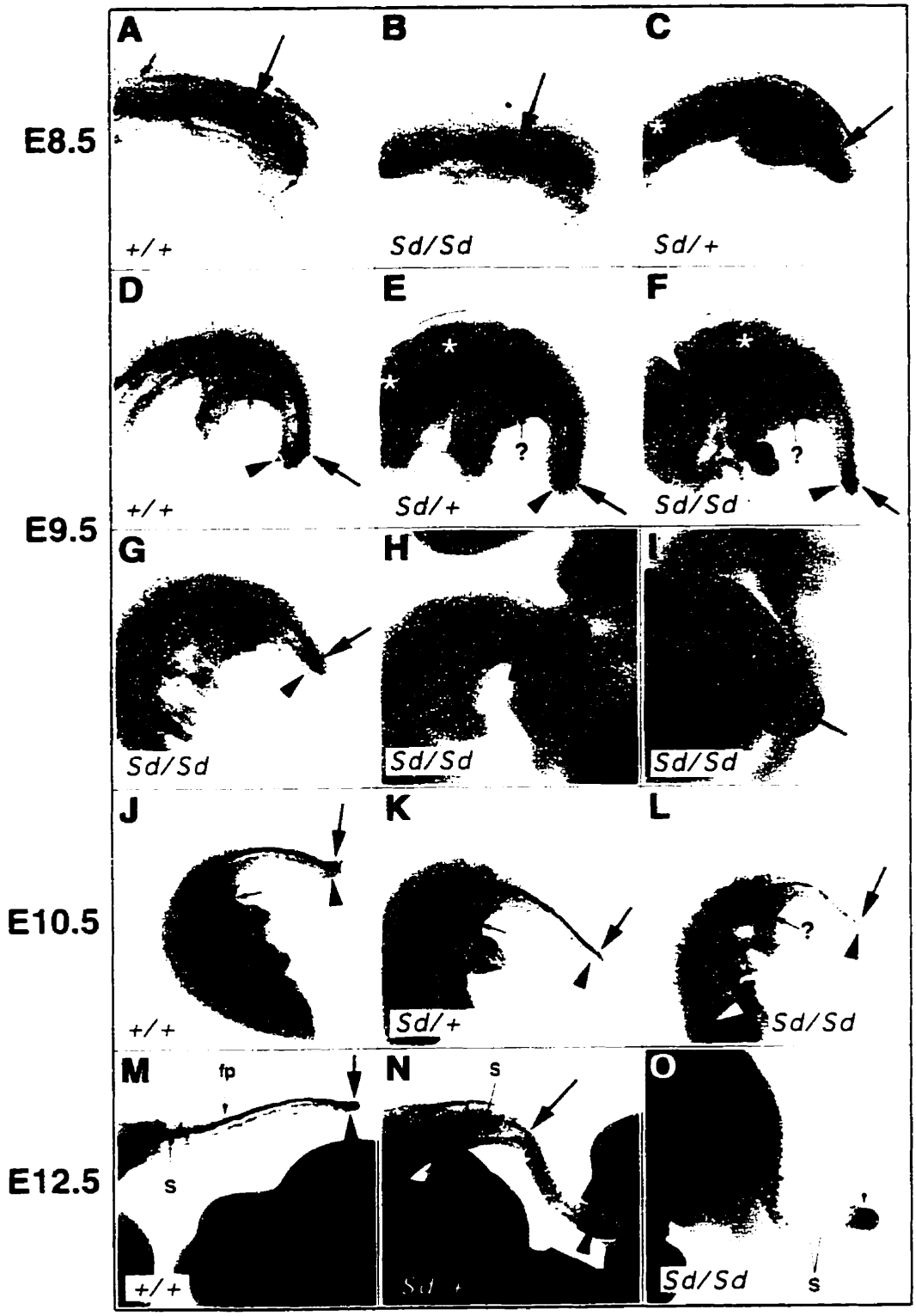

A
E8.5
B
C
+/+
T/+
T⁻⁻/+
D
E
F
G
+/+
T⁻⁻/+
+/+
T⁻⁻/+
H
E9.5
I
+/+
T/+
K
T⁻⁻/+
L
T⁻⁻/+
M
E10.5
N
O
+/+
T/+
T/+
P
Q
R
S
T/+
T⁻⁻/+
T⁻⁻/+
S
E11.5
U
+/+
T/+
T/+

Fig. 22. - Expression of *cobl* in the tail bud of *Danforth's short-tail* mutant embryos.

Orientation of the tails: All figures are lateral views with dorsal at the top or right side and posterior at the right or bottom. Genotypes are indicated in the lower left corner and developmental stages to the left. All figures compare expression in the notochord bud (nb. indicated by an arrow) and in the hindgut/tailgut (indicated by an arrowhead) between wild-type and mutant tail buds. Trunk notochord defects are also indicated by asterisks. (A, B) E8.5 embryos shown in Fig. 20 A & B; at this stage, no strong difference was found between control and mutant embryos; however, the nb was often smaller in mutant tailbuds (B). (C) From day 9.0, the lack of nb was visible in both *Sd/+* (branching notochord, asterisks) and *Sd/Sd* embryos (not shown). (D) The nb and the tail gut are well developed; expression in the future cloaca was also visible (small arrow). (E) The tail gut appeared normal but expression in the nb and cloaca (small arrow) was much reduced. (G-I) In *Sd/Sd* embryos, the chordoneural hinge (cnh) was abnormal because of both defects in the notochord and the tail gut; the nb was missing in all cases; the tail gut was always reduced and had failed to form a ventral pocket or was even misshapen (G & H); expression in the cloaca was not detectable (small arrow in F) or clearly ectopic (left arrowhead in G). (J-L) One day after, *Sd/Sd* tails started to appear shorter than *Sd/+* tails. (K) The notochord extended to the tip of the tail but the cnh was missing in *Sd/+* embryos. (L) The decay of the notochord was very advanced in homozygous embryos and the tail gut was not visible; note also the very weak expression in the endoderm of the cloaca and allantois (small arrows) which contrasts with the strong expression in wild-type (J) and *Sd/+* embryos (K). (M-O) In E12.5 wild-type tails (M), tail gut and notochord extended to the tail tip and the cnh structure was still manifest; expression was also induced in the tail somites (S) and floor plate (fp). (N) The notochord has stopped prematurely but the tail gut extended to the tip of the tail; somites also formed beyond the notochord in the tail; however, induction of *cobl* expression in the somites (S) and the survival of the latter was only observed in positions adjacent to notochord tissue. (O) No *cobl* expression was detectable; the tail contained some somites (S) but was short and necrotic with hemorrhages (small arrow).

Abbreviations: fp: floor plate; nb: notochord bud; S: somite.


E8.5
E9.5
E10.5
E12.5
A
B
C
+/+
Sd/Sd
Sd/+
D
E
F
+/+
Sd/+
Sd/Sd
G
H
I
Sd/Sd
Sd/Sd
Sd/Sd
J
K
L
+/+
Sd/+
Sd/Sd
M
fp
s
N
s
O
+/+
Sd/Sd
s

beyond the rump. It could either stop abruptly or end as one or several short individual fragments (Fig. 21 Q). The tail gut expression was not visible in the tail of $T^{Wis}/+$ embryos (Fig. 21 Q) and no expressing cell could be found at the tip of the tail (Fig. 21 Q.R). In one embryo the notochord was branching with the gut and gut expression was visible down to the tip of the tail (Fig. 21 R). In all mutant tails, somites kept forming in the tail well beyond the end of the notochord.

From stage E11.5, expression is induced in tail somites and the notochord and tail gut extend to the CNH at the tip of the tail (Fig. 21 S). In mutant embryos, the survival of the tail appeared to be dependent on the presence of notochordal tissue. Induction of expression in the somites and conservation of the normal tail diameter strictly correlates with the presence of underlying axial expression in notochord (Fig. 21 *T*.U). From that stage, the breakdown of tail sections lacking a notochord has started (see necrosis in Fig. 21 *T*.U).

Expression of cobl in the tail bud of Danforth's short-tail embryos

Sd/+ and *Sd*/*Sd* mice having a short tail or no tail respectively also provide a good experimental system to study abnormal development of the tail bud. Although 40% of E8.5 embryos had reduced expression in the CNH (Fig. 22 B), a clear difference between wild-type and mutant embryos did not appear before E9.5 before posterior neural tube closure (Fig. 22 C). At this stage the notochord bud was absent or reduced in the CNH of all mutant embryos. As the tail elongates, defects in the tail of *Sd*/+ embryos remained restricted to the notochord bud of the CNH, suggesting that the notochord was reduced rather than missing since it was still being formed during elongation (Fig. 22 E). Alterations in the CNH of *Sd Sd* embryos however, became more severe so that both notochord and gut expression were very reduced and the CNH, if present at all, had an abnormal structure: the notochord bud was reduced or missing and the tail gut either stopped prematurely or failed to expend ventrally as in wild-type or heterozygous embryos (Fig. 22 F-I). The ventral expansion of the gut and its contact with the ventral ectodermal ridge (VER, Grüneberg, 1956) are important for the formation of the cloaca and anus. The latter never forms in *Sd*/*Sd* embryos, and the bladder, urethra and genital papilla, also partly derived from the cloaca are often missing.

From E10.5, the CNH of *Sd*/+ embryos started to disappear. The tail gut is still present and the cloaca appeared normal (Fig. 22 K). At later stages however, the notochord of the tail breaks down and only patchy expression remained in the proximal tail while the tail gut extended to the tail bud. Somites also formed down to

the tail bud but *cobl* expression was only induced in somites at levels adjacent to the notochord (Fig. 22 N). In *Sd/Sd* embryos the tail was already shorter at E10.5 and the notochord expression was reduced or missing in many segments of the tail. The cloaca or tail gut were forming distinct structures and expression was also missing in the allantois endoderm (Fig. 22 L). By E12.5, the tail was necrotic with sites of hemorrhages and no *cobl* expression was detectable, but somitic segments were still visible (Fig. 22 O). At birth however, the tail of *Sd/Sd* mice consisted in the best cases of a skin thread, suggesting the degeneration of all mesodermal and endodermal tissues.

Cordon-bleu expression in the somites of Brachyury and Sd mutant embryos

The notochord has been shown to play an important role in the dorso-ventral patterning of the neural tube (reviewed by Placzeck, 1995) and of the sclerotome (Pourquié *et al.*, 1993; Fan and Tessier-Lavigne, 1994; Johnson *et al.*, 1994) and myotome of the somites (Rong *et al.*, 1992; Stern and Hauschka, 1995; Pownall *et al.*, 1996).

Between E9.5 and E10.5, expression of *cobl* appears first in anterior somites and extends to the tail somites in a rostro-caudal order. At this stage, the morphology of trunk somites in *T/+*, *Sd/+* and *Sd/Sd* mutant embryos is not distinguishable from that of wild-type littermates. However, definite somites do not form in homozygous *Brachyury* mutants and expression of somite specific markers has been shown to be altered in *Sd* mutant embryos (see Table 17 & Koseki *et al.*, 1993; Dietrich *et al.*, 1993). Therefore, I analyzed *cobl* expression in the somites of *Brachyury* and *Sd* mutant embryos. No *T/T* embryo expressed *cobl* in somite regions as predicted. However, T/+ or Sd/+ and Sd/Sd did show expression but in all cases, expression was reduced or lost in regions of absent notochord.

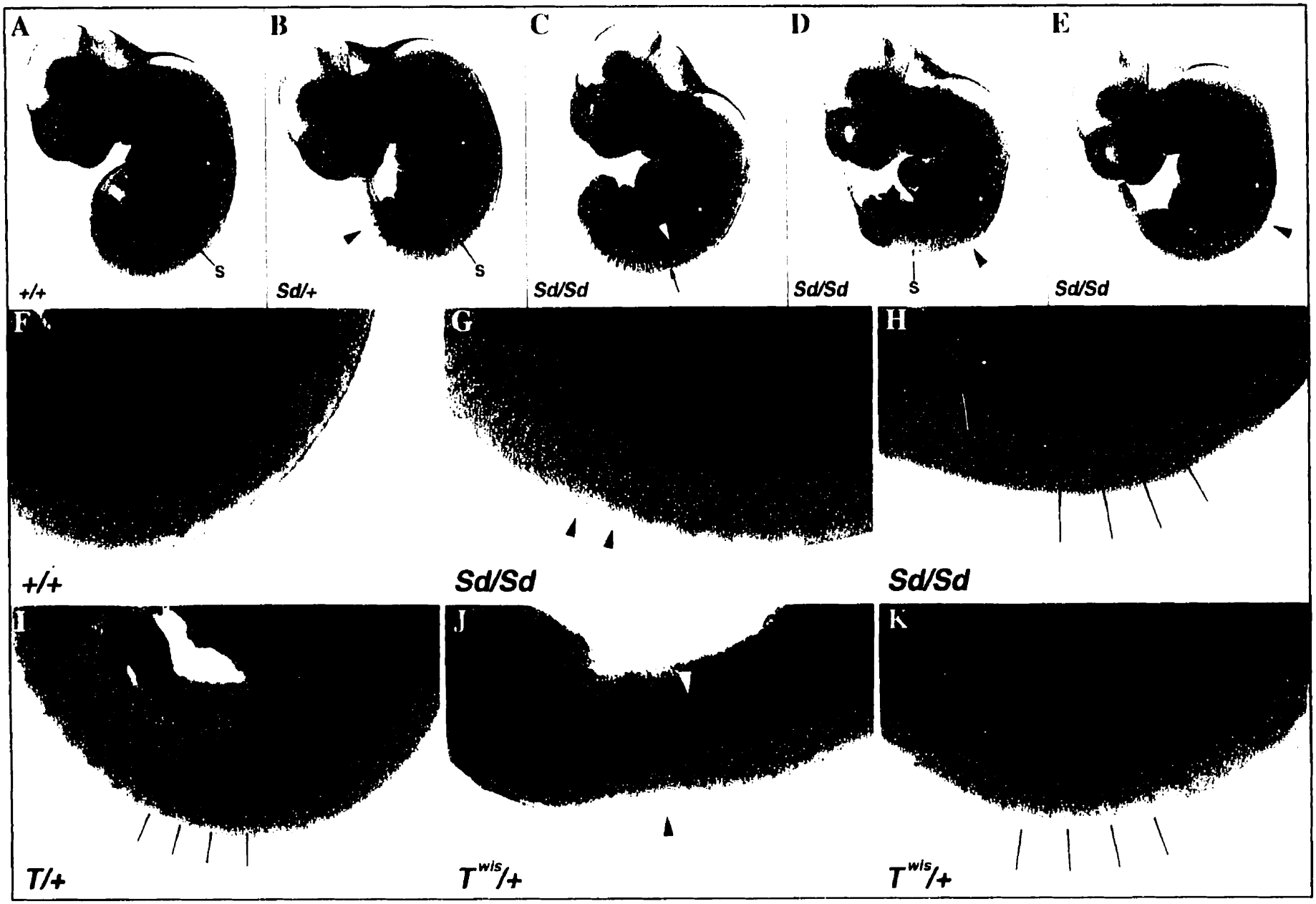
For example, many T^{Wis}/+ embryos had reduced somitic expression of *cobl* in posterior regions (typically in the rump and tail regions) where the notochord became interrupted, reduced or missing at axial positions (e.g. Fig. 21 R). One T^{Wis}/+ embryo had reduced expression in two trunk somites (contralateral) at a level where axial expression (notochord and floor plate) was missing (Fig. 23 J). This phenotype was the most severe in *Sd/Sd* embryos which lack notochord almost entirely from day E10.5.

Seven embryos (20%) showing the severe tail defects typical of *Sd/Sd* embryos and reduced axial *cobl* expression anterior to the hindlimb, had clear defects in somite expression. Defects ranged from weaker expression in individual somites (Fig. 23 C, D, G) to almost complete loss of expression (Fig. 23 E, H). However, a few cells expressing *cobl* at low levels could always be identified in some somites of the most severely affected homozygous embryos (Fig. 23 H). The same altered somite expression was observed in [*Sd/Sd; GtC101/+*] embryos (n=5). In *Sd/+* embryos, expression was absent only in tail segments missing axial expression of *cobl*. Thus, a direct correlation between notochord expression of *cobl* and somite expression was observed, supporting the notion that intact notochord is required for correct patterning of the somites.

In addition, 1 *T/+* and 1 T^{Wis1}/+ embryos (out of 9 *T/+* and 41 T^{Wis1}/+ E10.5 or E11.5 heterozygous embryos) displayed abnormal *cobl* expression in all somites. These embryos had ectopic expression of *cobl* in the anterior and posterior compartments of somites despite apparently normal expression along the axis (Fig. 23 I & K).

Fig. 23. - Somite patterning defects in 10.5 dpc *Danforth's short-tail* and *Brachyury* mutant embryos.

Cobl/LacZ expression in the somites. Dorsal is at the right (A-E) or bottom (F-K) and posterior is at the bottom or left respectively. The genotypes are indicated in the lower left corner. (A) Wild-type embryo showing normal tail development and expression of *cobl* in the somites (S) from the fore limb (white asterisk) to the rump. (B) *Sd/+* embryos, identifiable by altered *cobl* expression in the notochord (starting at the arrowhead level) but showing normal expression in the somites. (C-E) The tails were kinked and *cobl* expression in the axis was interrupted from the trunk down (arrowheads): expression in the somites was reduced (C) or almost completely missing (D, E). (F-H) Details of embryos shown in (A, C & E). Lower levels of somite expression (arrowheads) and the division between individual somites (bars) are indicated. (I) Ectopic expression in the somites of a *T/+* embryo. (J) Reduced expression in the axis (white arrowhead) and contralateral somites (only 1 visible, black arrowhead) of a $T^{Wis}/+$ embryo. (K) Ectopic expression in the somites of a $T^{Wis}/+$ embryo.


A
B
C
D
E
S
S
S
+/+
Sd/+
Sd/Sd
Sd/Sd
Sd/Sd
F
G
H
+/+
Sd/Sd
Sd/Sd
I
J
K
T/+
T^wis/+
T^wis/+

DISCUSSION

This study demonstrates the efficacy of using gene trap insertions as developmental markers. Since the GtC101 insertion into cordon-bleu has no phenotype on its own, I could use the *lacZ* reporter gene as a neutral marker to analyze development in embryos carrying mutations at the *Brachyury* or *Danforth's short-tail* loci. These analyses provided new information about the time of onset of anomalies caused by these mutations in both heterozygous and homozygous embryos. Analysis of *cobl* expression allowed the detection of the earliest defects reported so far in Brachyury homozygous and heterozygous embryos: the node of E7.5 homozygotes only showed weak expression of *cobl*, suggesting a patterning defect or the deficiency of a specific population of node cells. Similarly, incomplete expression of *cobl* marked the tail bud of E8.5 Brachyury heterozygotes embryos. Despite the lack of a molecular assay to determine the genotype at the Sd locus, the COBL/ß-Gal marker proved to be a good indicator to distinguish wild-type, heterozygous and homozygous embryos. Most Sd homozygotes could be identified at E8.5 by abnormal expression of *cobl* in trunk notochord precursors and all embryos could be classified into genotypic classes from day 9.5. Analysis of *cobl* expression in the tail bud of T and Sd heterozygotes and Sd homozygotes revealed early alterations of the chordoneural hinge that are consistent with the lack of tail growth in mutant embryos and the urogenital phenotype in Sd homozygous mice. Furthermore, these analyses provided novel information on the patterning of the somites by notochord during embryonic development.

Node formation in Brachyury mutants

The best unitary explanation of the *Brachyury* phenotype has come from the analysis of chimeric embryos comprising wild-type and mutant cells. Cell movements are impaired in *T/T* cells, and cells tend to accumulate in the posterior streak thus blocking posterior development (Wilson *et al.*, 1995). Cell movements are also involved in the migration of anterior notochord precursor cells and in the formation of the allantois posterior to the streak. This study does not explain however, the dominant phenotype found in the tail of heterozygous mice. Furthermore, the defects observed in the notochord, the primitive streak and the allantois could be independent. Previous analyses of *Brachyury* mutant embryos

using genetic markers did not reveal defects in the primitive streak before the formation of the first somites at E8.0 (see Table 16). In this study, I show that expression of *cobl* is clearly altered in the node of T^{Wis}/T^{Wis} embryos at E7.5, before the decrease of *T* expression observed in T^{Wis} mutants at E8.0 (Herrmann, 1991). Expression of *cobl* in anterior midline appears normal in T^{Wis}/T^{Wis} embryos at the early head fold stage but the lack of expression in the node region persists. Expression of *Sonic hedgehog* (*Shh*), another node-notochord marker has been reported to be normal in the node and head process of T^{Wis} homozygous embryos at E7.5 (Conlon *et al.*, 1995). Expression of *Shh* has been interpreted as suggesting that the node forms in homozygous embryos, while the lack of notochord has been usually described as a defect in the maintenance of notochord precursor cells or in their differentiation into notochord. The results presented here suggest that the node is already altered at an early stage, missing cell sub-populations or not being properly patterned. This node defect could precede the morphological defects observed by electron microscopy (Fujimoto and Yanagisawa, 1983). The fate of notochord precursor cells may be already compromised within the node. Dorsal gut endoderm and notochord are derived from cells of the ventral node (Lawson *et al.*, 1986; Beddington, 1994). The missing or weaker *cobl* expression in the gut endoderm of some T^{Wis}/T^{Wis} embryos (data not shown) could also be linked to a common node defect.

The role of the T transcription factor in the formation or patterning of the mouse node is demonstrated here. Furthermore, the early defect of *cobl* expression prior to any morphological defect suggests that *cobl* expression could be partly regulated by the T protein in the node. The recent identification of a node and notochord expression specific element in the promoter regulatory region of the *T* gene (Clements *et al.*, 1996) also suggests that T could play distinct roles in the node/notochord and in the primitive streak.

Notochord formation in Danforth's short-tail mutants

The analysis of the *Sd* mutation was impaired by the lack of genetic markers to confirm the genotype of embryos from early stages presenting no abnormal morphological traits. However, expression of *cobl* turned out to be a very good indicator of the possible genotype of embryos. In 1958, Grüneberg made an extensive histological analysis of the *Sd* phenotype and proposed that the alteration of the notochord was the primary defect responsible for the whole skeletal and urogenital phenotype (Grüneberg, 1958). He also postulated that the

Table 16

Genetic marker analyses in *Brachyury* mutants.

Marker	Normal expression [a]	Defect onset	Expression in homozygous *Brachyury* mutants	*Allele*	Ref. [b]
Node and notochord formation					
goosecoid	Anterior streak in presumptive node region.		Normal at E6.5/7.0.	l^{Wis}	(4)
cobl	Node, notochord, floor plate, gut endoderm and liver.	E7.5 E8.5	Abnormal onset in the node. Abnormal in chordoneural hinge of l + and l^{Wis} + embryos.	l, l^{Wis}	
T	Primitive streak Node and notochordal plate Notochord.	E8.0 (1 som.)	Decrease in notochord precursors, node and primitive streak. No mRNA detectable after E8.5.	l^{Wis} l^{c}	(1) (5)
T protein	*idem*	E8.0 (1 som.)	Up to E8.5, residual T protein in notochord precursors and primitive streak. After E9.5, no more T protein is detectable.	l^{Wis}	(2)
Shh	Node, notochord, floor plate, gut endoderm and liver.	E8.5 (5 som.)	Normal onset in the node & head process Later, only in ventral forebrain, liver, hindgut. Not in notochord precursors containing T^{Wis} protein.	l^{Wis}	(4)
Primitive streak patterning					
BMP-4	Posterior streak and ventral mesoderm.		Normal at E7.5 and E8.5.	l	(3)
Evx-1	Primitive streak ectoderm & mesoderm.	E8.5 (7 som.)	Decrease to complete extinction by E9.5.	l	(3)
Wnt-3a	Primitive streak ectoderm & mesoderm.	E8.5 (5 som.)	Expression is normal at E7.5 Decrease from the 5 som. stage.	l	(3)
Wnt-5a	Primitive streak, lateral mesoderm and all posterior tissues from last som. to allantois.	E8.5 (2 som.)	Normal at E7.5 Decrease and complete loss by the 6 som. stage.	l	(3)
cdx-4	Expression in tail bud and base of allantoic bud (overlap with *T*).	E8.5	Ectopic in anterior embryo and posterior extraembryonic mesoderm (also observed to a lesser extent in l^{Wis} + embryos).	l^{Wis}	(4)

(*continued*)

Table 16 (*continued*)

Genetic marker analyses in *Brachyury* mutants.

Marker	Normal expression ([a])	Defect onset	Expression in homozygous *Brachyury* mutants	*Allele*	Ref. ([b])
Neural tube patterning					
Hox-A.7 *Tgm6lacZ1*([c])	Anterior limit of expression in NT marks the 8th somite's level.		Normal anterior posterior boundaries.	*T*	(3)
Msx-1	Dorsal NT and caudal ectoderm.	E8.5 (5 som.)	Dorso-ventral unchanged but posterior extension in caudal ectoderm.	*T*	(3)
Pax-3	Dorsal NT and dermomyotome.	E9.5	Ectopic expression in sclerotome and caudal extension in somites.	*T* T^c	(3) (5)
Pax-6	Dorsal and ventral NT excluding the floor plate.	E9.5	Dorsal and ventral shift including the floor plate.	*T*	(3)
Somite patterning					
Mox-1	Paraxial mesoderm (somites).	E8.5	Expression in somites but no more than 8 paired pre-somitic blocks are formed.	T^{Wis}	(4)
Pax-1	sclerotome (ventromedial som.) and posterior pre-somitic mesoderm.	E9.5	Normal onset but reduced and limited to anterior somites.	*T* T^c	(3) (5)
twist	Cephalic mesoderm, sclerotomes and base of the allantois (no overlap with *T*).	E8.5	Reduced in cephalic mesoderm and absent in paraxial mesoderm (presence of *Mox-1* suggests pattern defect or lack of sclerotome cells).	T^{Wis}	(4)

([a]) Only a relevant subset of expressing tissues is indicated here.
([b]) (1) Herrmann, 1991; (2) Kispert and Herrmann, 1994; (3) Rashbass *et al.*, 1994; (4) Conlon *et al.*, 1995; (5) Dietrich *et al.*, 1993.
([c]) Transgenic *lacZ* reporter gene reflecting *Hox-A7* expreston.
Abbreviations: E8.5= embryonic day 8.5; NT= neural tube; som.= somites.

initial defect may be within the notochord precursors in the node/primitive streak since the notochord never seemed to form properly as it emerged from the streak (Grüneberg, 1958). Later histochemical analyses also suggested that the newly formed notochord is abnormal in *Sd* mutant embryos (Paavola *et al.*, 1980). However, earlier studies analyzed formation of posterior notochord at stages when the genotype is revealed by phenotypic traits, or they focused on NT (Bovolenta and Dodd, 1991; Phelps and Dressler, 1993) and on paraxial (Dietrich *et al.*, 1993; Koseki *et al.*, 1993) or lateral mesoderm (Phelps and Dressler, 1993) patterning (see Table 17). They could not easily analyze the earlier formation of anterior notochord. I could not detect abnormal *cobl* expression in the node and head process of any E7.5 embryos derived from [*Sd*/+; *GtC101*/*GtC101*] parents, either in the node or posterior streak before late E8.5 stage. This suggests that the appearance of anterior notochord precursors occurs normally in mutant embryos. This is consistent with the normal induction by axial mesoderm of anterior structures such as the floor plate (analyzed at E13.0 by Bovolenta and Dodd, 1991). From E8.5 (10 somites) however, *cobl* provided a good marker to follow notochord morphogenesis in mutant embryos. In a substantial proportion of late E8.5 embryos (19%), the structure of the notochord emerging from the node was abnormal. This result supports the idea that trunk notochord precursors are abnormal and never condense into a normal notochord in *Sd* mutant embryos. At later stages, the notochord was also found to have an anomalous structure along its entire length. It partly supports Grüneberg's theory (1958) and contrasts with a role for the survival of differentiated notochord initially attributed to *Sd*.

Posterior axis and tail bud development

In the tail bud, the domain of *cobl* expression corresponds to the mouse chordoneural hinge (CNH). The CNH structure was discovered by Pasteels (1943) and later inferred in the frog by the complementary expression domains of *XNotl* and *XBra* (Gont *et al.*, 1993) and in the mouse and chick by cell lineage analyses (Wilson and Beddington, 1996; Catala *et al.*, 1995). In *Xenopus*, it has been shown to retain the organizer activity of the dorsal blastoporal lip (Gont *et al.*, 1993) and to be necessary for complete axial patterning and growth of the tail bud (Tucker and Slack, 1995). The continuous expression of *cobl* in the node, posterior streak and chordoneural hinge of the tail bud and the defects observed in mutant embryos support these findings. Furthermore, the lack of tail growth in mutant embryos can be traced back to early alterations in the dorsal component of the

Table 17

Genetic marker analyses in *Danforth's short-tail* mutants.

Marker	Normal expression [a]	Defect onset	Expression in Sd mutants	Ref. [b]
cobl	Node, notochord floor plate and gut endoderm. Subset of cells between anterior and posterior somite compartments.	E8.5 E9.5 E10.0	Weak and discontinuous trunk notochord in *Sd/Sd*. Incomplete notochord in *Sd/Sd* & *Sd/+* and abnormal hindgut in *Sd/Sd*. Incomplete onset in all somites.	*herein*
T	Primitive streak, node, notochord and tail bud.	E9.5	Reduced and diffused in *Sd/Sd* & *Sd/+*. Interrupted where notochord breaks down.	(1,2)
Pax-1	Sclerotomes, in all somites and body segments (future vertebrae) from E12.5.	E9.5 (<25 som.) E10.5 (<40 som.)	Reduced in hindmost 5 somites of *Sd/Sd* embryos. The 17 anterior somites are positive in *Sd/Sd* embryos, versus 35-40 in wild-type embryos. Reduced in hindmost somites of *Sd/+* embryos. At E13, only 18 vertebral bodies will form in *Sd/Sd* embryos, compared to more than 40 in +/+ embryos.	(1,2)
twist	Sclerotomes	E10.5	Reaches the 20th somite sclerotome, beyond the last somites expressing *Pax 1* (18th) in adjacent sections.	(1)
Pax-3	Dorsal and ventral NT excluding the floor plate. Dermomyotomes and tail bud.	E9.0	Ventral expansion in NT posterior to 16th somite. Absent from the tail bud in *Sd/Sd* or *Sd/+*.	(1,2)
Pax-2	Intermediate zone of NT (from E12);	E12.0	Expands to ventral NT (basal and floor plate). Ectopic expression in dorsal alar plate.	(3)
	Metanephros and ureteric bud of forming kidneys (at E12).	E12.0	No expression in metanephros (E12). Reduced expression in ureteric bud (E13).	

([a]) Only a relevant subset of expressing tissues is indicated here.
([b]) (1) Koseki *et al.*, 1993; (2) Dietrich *et al.*, 1993; (3) Phelps & Dressler, 1993.
Abbreviations: E8.5= embryonic day 8.5; NT= neural tube; som.= somites.

chordoneural hinge, the notochord bud - representing the hindmost notochordal cells. In all embryos fated to become short-tailed or tailless animals, expression in the notochord bud was reduced or missing (Fig. 21 & 22). For each genotype, I always found a strong correlation between corruption or premature arrest of the notochord at E10.5 and predicted tail length: in *T/+* embryos, the notochord and the tail gut appeared normal but the former stopped prematurely at different levels in each embryo even though tail gut always reached the tail bud (Fig. 21 N-P); in $T^{Wis}/+$ embryos, the notochord always stopped just posterior to the rump and tail gut was not detected in the tail (Fig. 21 Q); in *Sd/+* embryos, the notochord and tail gut initially extended to the tail tip but both failed to reach further at later stages (Fig. 22 K); finally in *Sd/Sd* embryos, no tail gut was visible and the notochord was decaying at all levels of the tail (Fig. 22 L). The tail bud always elongated and formed somites beyond the hindmost level of expression in the notochord. However these tail segments eventually decayed suggesting that the tail gut (present in *T/+* embryos) is not sufficient to support tail growth or survival. Axial tissues posterior to the CNH of the tail bud have been shown to form posterior and somitic mesoderm and could be responsible for this elongation (Wilson and Beddington, 1996). Grafting experiments in Xenopus have also shown that tail segments anterior to the CNH are also responsible for tail elongation (Tucker and Slack, 1995). Therefore, apparent tail growth could alternately be the result of proximal extension accompanied by distal tail resorption.

Urogenital defects of Sd/Sd embryos

Grüneberg (1958) postulated that the urogenital defects in Sd mutant embryos were a direct consequence of the notochord alterations. In *Sd/Sd* embryos, alterations in the gut structure were found at the level of the presumptive cloaca at E9.5 (Fig. 22 F-I). At E10.5, the cloaca is not visible and expression in the allantois endoderm is also reduced (Fig. 22 L). Abnormal cloaca development was also observed in some heterozygous embryos, a proportion of which suffers urogenital defects. This abnormal development of the cloaca may represent the primary cause for the whole urogenital phenotype (lack of kidneys, bladder, urethra and genital papilla). During kidney development, the ureter originates from the mesonephric duct but becomes part of the cloaca as it grows into the metanephric mesenchyme to induce kidney formation. It is possible that this process is prevented in *Sd/Sd* embryos by an abnormal cloaca thus preventing further development of the kidneys. However, analysis of *Pax-2* expression in E12.5 embryos clearly

demonstrated abnormal patterning of the metanephric mesenchyme at day E12.5 whereas the epithelium of the ureter expressed *Pax-2* (Phelps and Dressler, 1993). Therefore causes other than cloaca and ureter defects are still conceivable. The influence of the hindgut defect on the development of the bladder and urethra can be inferred by their mode of formation which is more closely associated with differentiation of anterior and posterior cloaca into urogenital tissues and rectum respectively. These results however, do not support or suggest any causal effect between the notochord and the hindgut defects as suggested by Grüneberg (1958).

Expression of cordon-bleu and embryo patterning

The maintenance of its expression indicates that *cobl* does not strictly require the presence of wild-type *Brachyury* protein or wild-type *Danforth's short-tail* gene. Expression in gut, notochord and floor plate occurred independently at different positions along the anterior-posterior axis. It suggests that expression of *cobl* is not dependent on *cobl* expression in adjacent tissues but rather on the differentiation stage of each tissue. For example, the floor plate expression was seen in the absence of adjacent notochord. This is consistent with earlier observations in the mouse and zebrafish that floor plate could be induced in embryos lacking a definite notochord (Rashbass *et al.*, 1994; Halpern *et al.*, 1993).

An important alteration in the regulation of *cobl* expression happened in the somites of mutant embryos. Expression of *Pax-1* (sclerotome marker) and *m-Twist* (somite marker) have already been showed to be reduced in *Sd* mutant embryos (Dietrich *et al.*, 1993; Koseki *et al.*, 1993 and Table 17). Somitic cells expressing *cobl* are localized in the medio-ventral myotome compartment but later seemed to migrate toward the vertebral column where expression of *cobl* was principally seen in the chondrocytes of forming cartilage. This domain of expression was severely reduced or totally missing in *Sd/Sd* embryos (Fig. 23). Subsequently, *Sd/Sd* embryos have reduced vertebrae and completely lack intervertebral discs. Their position at the limit between sclerotome and myotome compartments together with expression of *cobl* suggest that they could represent a specific groups of sclerotome cells giving rise to part of the vertebrae and/or to intervertebral discs. Furthermore, the localization of these cells between the anterior and posterior compartment of each somite correlates with the position of intervertebral discs along the anterior-posterior axis. The lack of expression in the somites closely

correlates with the absence of notochord. Together with the reduced expression in one $T^{Wis}/+$ embryo also lacking axial staining, it suggests that the missing notochord rather than the mutation is responsible for this phenotype. The notochord has been shown to be involved in the dorso-ventral patterning of the sclerotome (Pourquié *et al.*, 1993; Fan and Tessier-Lavigne, 1994) and myotome (Rong *et al.*, 1992; Stern and Hauschka, 1995; Pownall *et al.*, 1996) compartments in the somites. Interestingly, expression of *Pax-1* was reduced in posterior somites only (Koseki *et al.*, 1993) whereas I found alterations of *cobl* up to the forelimb.

The ectopic expression of *cobl* in the somites of $T/+$ and $T^{Wis}/+$ embryos could be due to an abnormal underlying notochord or to a role for the T protein in somitogenesis. The latter is already suggested by incomplete somitogenesis in *Brachyury* homozygous embryos and by the lack of T/T cell contribution to paraxial mesoderm in <T/T cells + wild-type cells> chimeric embryos (Rashbass *et al.*, 1991; Wilson *et al.*, 1995). Therefore in these embryos, a threshold of T activity required for somitogenesis may not have been reached.

These experiments illustrate how novel gene trap markers can be used to investigate specific cell lineages development or detect novel genetic interactions. Using the *GtC101* insertion into *cordon-bleu* to analyze the *Brachyury* and *Danforth's short-tail* phenotypes provided novel data about node/notochord formation, tail morphogenesis and somite patterning. The results presented here also support previous reports suggesting that the node was altered in Brachyury homozygotes, that the tail bud chordoneural hinge is essential for tail growth and that patterning of the somites depends in part on the presence of an intact notochord.

- CHAPTER V -

DISCUSSION

CHAPTER FIVE : DISCUSSION

During embryogenesis of a metazoan, important developmental genes are expressed in specific regions or tissues where they act as regulators of embryo patterning and organogenesis. I have used gene trap vectors and mouse ES cells to identify and characterize a novel gene showing a restricted pattern of expression during early embryogenesis. Various aspects and potentials of the gene trap approach have been investigated. In a large scale screen, 303 independent insertion events, activating the gene trap vector *in vitro*, were assayed for expression of the *lacZ* reporter gene *in vivo* in 8.5 dpc chimeric embryos. Thirty-nine candidate developmental genes were identified together with a large number of ubiquitously expressed genes. I characterized further one insertion and cloned a novel gene, *cordon-bleu*, expressed in axial structures during gastrulation. This gene trap insertion was also used as a marker to analyze the phenotypes of mutations in the *Brachyury* and *Danforth's short-tail* genes.

I will now discuss these results, the potential utilization of these biological tools, and finally the future experiments necessary to address some of the questions that arose from the work presented here.

THE GENE TRAP SCREEN

The pilot screen using spontaneous differentiation of ES cells *in vitro* did not provide sufficient criteria to pre-select for insertions into gene with restricted pattern of expression during early embryogenesis. However, our pre-screen was rather crude and one should not reject the idea of pre-screening ES cells for the regulation of reporter gene expression following *in vitro* differentiation. Other pre-screens using more specific *in vitro* differentiation protocols successfully enriched the pools of lines tested further for insertions into genes which are developmentally regulated *in vivo* (in particular, Forrester *et al.*, 1996).

In our screen, the chimeras derived from 167 (55%) ES cell clones did not express *lacZ* at E8.5 whereas 97 (32%) showed ubiquitous expression and 39 (13%) displayed a restricted pattern of expression. However, the screen was certainly not fully exploited. These 39 insertions allowed us to identify candidate developmental genes possibly involved in embryo patterning. Determining further the role of these genes however, would have required their cloning and molecular characterization as well as the genetic analysis of the potential phenotypes created

by the insertions. The identification of mutant embryos is still the most convincing evidence for the importance of a gene product during embryogenesis. Genes that are ubiquitously expressed should not be overlooked either. In the screen performed by Friedrich and Soriano (1991), the majority of recessive embryonic lethal mutations corresponded to insertion into genes showing widespread expression. Furthermore, the phenotype of genes that are ubiquitously expressed can affect very specific tissues (e.g. BTF3 in Deng and Behringer, 1995). In our large screen, the classification of embryos into the "ubiquitous" group was often decided after an overnight staining. For practical reasons, little attention was given to the relative levels of expression between different tissues and expression was not analyzed in sectioned embryos. The insertion H201 (pilot screen) provides a good example. Expression appeared ubiquitous after an overnight staining; however, after a short staining, expression in the head CNS was clearly stronger and sections showed that H201 was not expressed in endoderm. Therefore, some insertions into genes with subtle restriction of expression may have been wrongly placed into the "ubiquitous" class. Still, this screen and others gave insights into the potential of gene trapping approaches and suggested possibilities for improving the screening strategies or for alternative uses of trapping vectors.

Potential of trapping approaches

The potential to identify, clone and mutate genes by the gene trap approach has been demonstrated by different screens (see Introduction). However, these different screening strategies also revealed some of the advantages and drawbacks of the trapping approaches.

Advantages of the trapping approaches

There are two main advantages to the use of trapping vectors in the mouse: i) insertion events can be selected and screened *in vitro* and ii) the screening strategy may rely on three main criteria: the analysis of gene expression, the cloning of trapped genes and the detection of mutant phenotypes.

The possibility to pre-screen insertions *in vitro* via the analysis of expression or the direct cloning of trapped genes presents important advantages over mutation screens when one wants to perform large scale experiments on mammalian cells. First, mutation screens require testing the effects of an insertion *in vivo* when the other two approaches can be performed *in vitro*. Furthermore,

the implementation of better strategies controlling *in vitro* differentiation of multipotent cells will expand the scope of developmental processes which can be targeted by future screens. Next, if one wants to isolate any gene, mutation screens also bear some limitations. In the mouse, some null mutations into genes expected to play an important developmental role at specific stages of development present no phenotype at all or phenotypes appearing in unexpected times and tissues. In the yeast *S. cerevisiae*, 60% of gene disruptions caused no detectable phenotype affecting cell growth and division (Goebl and Petes, 1986; Burns *et al.*, 1994). Genes are likely to escape a mutagenic screen because their function is redundant or the phenotype is too subtle to be seen during the screen. Isolating genes on the basis of their expression pattern or by direct cloning circumvents these limitations. However, it is easy to argue that what determines the real importance of a gene's product is the phenotype created by its disruption. When this trapping approach was selected, the average number of gene trap insertion resulting in a phenotype is comparable to that obtained with random gene disruption in yeast.

Drawbacks of the trapping approaches

Specificity of insertion

If one wants to saturate the genome with insertions, the efficiency and randomness of such events are important factors. Recombined retroviruses are very efficient vectors but may not integrate totally randomly in the genome. Retrovirus do not integrate randomly in the genome of birds (Shih *et al.*, 1988) and probably not in the mouse (Rohdewohld *et al.*, 1987; Scherdin *et al.*, 1990; Bonnerot *et al.*, 1992). In *D. melanogaster*, the *P* element transposon has some preferential sites for insertion (Tsubota *et al.*, 1985; Kelley *et al.*, 1987) and the specificity of insertion of the *Hobo* transposon is different from that of the *P* element (Smith *et al.*, 1993a). In the mouse, plasmid DNA may not integrate completely randomly either (Sutherland *et al.*, 1993; Macleod *et al.*, 1991). The two independent gene trap insertions that occurred into the *LAR* gene in a relatively small screen could represent such a limitation (Skarnes *et al.*, 1995). It is probably difficult to design a vector that would integrate perfectly randomly into the mouse genome. However, if different approaches are used, they should compensate each other for some of the bias that they contain.

Mutagenesis efficiency

The efficiency of mutagenesis is another important issue. Although the number of insertions that created a detectable phenotype affecting embryogenesis

(46%) is comparable with results from other random mutagenesis experiments. trapping vectors may not always be reliable mutagenic agents. Because GT vectors insert into introns. another problem is the splicing around the insertion site (Skarnes *et al.*. 1992; Gasca *et al.*. 1995) also observed with knock-out vectors (Moens *et al.*. 1992). During maturation of mRNAs. an endonuclease first cuts transcripts at the polyA signal sites before the addition of a poly-A tail by a terminal transferase. However. the poly-A signal sequence is not a strong signal in itself (Day. 1992). thus increasing the probability of splicing around the insert prior to the cutting of the RNA. Poly-AT vectors may prove to be even poorer mutagenic agents because they can be activated after insertion into 3' untranslated regions.

Reporter gene expression

The expression pattern of the reporter gene does not always reflect the expression of the disrupted gene. The high stability of the ß-Gal reporter protein may give false information about the intensity and duration of gene expression. In a more misleading case. a GT *lacZ* reporter was only expressed in the brain heart and testis after insertion into the BTF3 gene which was known to be expressed ubiquitously (Deng and Behringer. 1995).

Possible vector modification

A number of vector modifications are also possible that would improve gene detection and induction of mutations.

Gene expression and detection

Improved efficiency of isolating ES clones with gene-trap integrations in active genes has been achieved by development of the *ßgeo* gene-trap vectors. in which 95% of $G418^R$ colonies show ß-Gal activity (Friedrich and Soriano. 1991). The addition of a translation initiation sequence to the *lacZ* gene makes it independent of the endogenous protein coding sequence and increases the frequency of expressing clones by at least three-fold (PT-3 in the pilot screen). Replacement of the *ß-galactosidase* gene by a gene that can be assayed *in vivo* would eliminate the replica plating step of the screen. One candidate reporter gene is the green fluorescent protein (GFP) used as a reporter gene in *C. elegans* (Chalfie *et al.*. 1994). in *D. melanogaster* (Yeh *et al.*. 1995) and. in a modified form. in the mouse (Yang *et al.*. 1996; Zhang *et al.*. 1996; Zernicka-Goetz *et al.*. 1997).

Mutagenesis

Retroviral vectors generally insert at the 5' end of genes and their insertions are less likely to result in fusion proteins with residual activity. Introducing in the trapping vector sequences that promote release of the RNA polymerase downstream of the reporter gene could prevent transcription of sequences located in 3' and thus would decrease the risk of splicing around the insertion. However, such sequence motifs have not been clearly identified. Another useful addition to the gene-trap vector would be the insertion of sequences that respond to site-specific recombinases. Addition of these sequences would allow manipulation of the locus where insertion occurred to create genetic mosaics, revertants, or place other genes under the control of the trapped endogenous promoter.

What is the best trapping strategy?

The results of gene trapping depend both on the approach chosen and on the type of vector used. The approaches vary mostly with the type of target cells and with the screening criteria defining potentially interesting insertion events.

Even with improved trapping vectors, the rate-limiting step of the screen remains the production of transgenic mice or the expression screen in chimeric embryos. Better ES cell lines such as R1 ES cells and improved techniques for generating chimeric embryos, such as aggregation between ES cells and single embryos, can reduce the number of embryos and the effort needed to generate chimeras (Nagy *et al.*, 1993). A reduction in the number of insertions screened *in vivo* could also be achieved by prescreening ES clones for genes that have a higher possibility of being involved in developmental processes. For example subcellular localization of the ß-galactosidase activity could be used as a criterion. Another potential prescreen would take advantage of the ability of ES cells to differentiate spontaneously (Doetschman *et al.*, 1985), or in response to growth and differentiation factors (Forrester *et al.*, 1996; Baker *et al.*, 1997). In addition, systematic molecular characterization of the tagged genes by RACE-PCR could identify novel genes and candidate developmental genes (von Melchner *et al.*, 1992). These types of approaches, coupled with the induction of specific ES cell differentiation pathways *in vitro*, should make large scale mutagenic screens using trapping vectors feasible in mouse ES cells.

Alternate uses for trapping vectors

In *D. melanogaster* enhancer trap/*P*-element insertions have been widely

used to disrupt and isolate novel genes, and to saturate the genome with genetic markers (reviewed by Spradling *et al.*, 1995). These insertions also provide valuable markers to detect subtle pattern alterations in mutation screens, for cell tagging in mosaic experiments and they can reveal novel cell types within apparently homogenous tissues (e.g. in the D. melanogaster olfactory lobes: Riesgo-Escovar *et al.*, 1992; Yang *et al.*, 1995). The use of trapping insertion as genetic markers is also applicable to vertebrates.

Insertions that do not affect the viability of the mouse can be very useful for genetic studies. The ability to analyze cell lineage in chimeric embryos is essential to the understanding of cell fate and potential. The insertion ROSA-ßgeo-26 ubiquitously expresses the *ßgeo* reporter gene throughout development and adult life thus providing a marker equivalent to the nucleoli in the avian chick/quail chimeras system (Le Douarin, 1969; Le Douarin, 1973), since Rosa-26 carrying ES cells can be used to produce chimeric embryos. Other insertions, marking specific tissues, could also be used to analyze the effects of different mutations on specific cell types (as for GtC101 in Chapter IV).

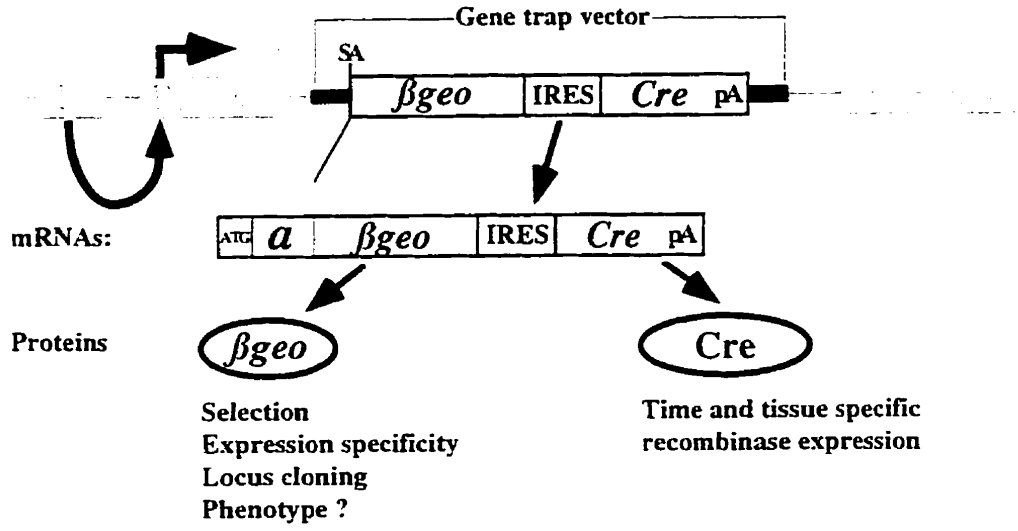
In addition to the reporter gene, GT insertions could put other genes under the control of endogenous promoters. Thus, proteins such as the Cre (Sauer and Henderson, 1989) or FLP (Dymecki, 1996a) recombinases could be expressed as "editor" proteins in specific tissues and at specific stages of development. These proteins can promote recombination between specific target sequences located in cis or in trans. Banks of cell lines or mouse strains with specific patterns of ß-Gal-Cre expression could be generated using trapping vectors (Fig. 24 A). The most efficient vector to generate these insertions would certainly be an enhancer trap, since ETs give the highest frequency of active insertion (see Introduction). However, if a screen for novel trapping insertions is undertaken, the use of a gene trap vector would concomitantly provide better tags for cloning the trapped genes and would generate mutations more efficiently in these genes (Fig. 24 A). Tissue specific Cre recombinase expression would allow the recombination (excision or inversion) of sequences flanked by the Cre specific loxP sites ("floxed" sequences). This approach has already been used to generate a tissue-specific mutation in the mouse (Gu *et al.*, 1994). It could also be used to activate oncogenes or toxins in specific tissues (see Fig. 24 B).

Fig. 24. - Diagram of possible experiments using trapping vector tools to generate tissue specific expression of Cre recombinase.

(A) Gene trap construct activated following insertion into gene *A*. Expression of the *βgeo* selection reporter gene and the *Cre* editor gene is regulated by the endogenous promoter. When insertion occurred into a developmentally regulated gene, the recombinase activity should be restricted to specific tissues and time points. (B) Possible target sequences for the Cre recombinase. Recombination will be targeted to sequences flanked by the LoxP sites ("floxed"); if the loxP sequences are in the same orientation, the floxed intervening sequence will be deleted by the action of the recombinase. (1) In a conditional knock-out of gene *B*, exon B2 will only be deleted in cells expressing Cre. (2) Deletion of the floxed stop codons will enable translation of the gene *C*. This approach may be used to induce tumor formation (a), to selectively kill cell lineages (b) and for partial phenotype rescue (c).

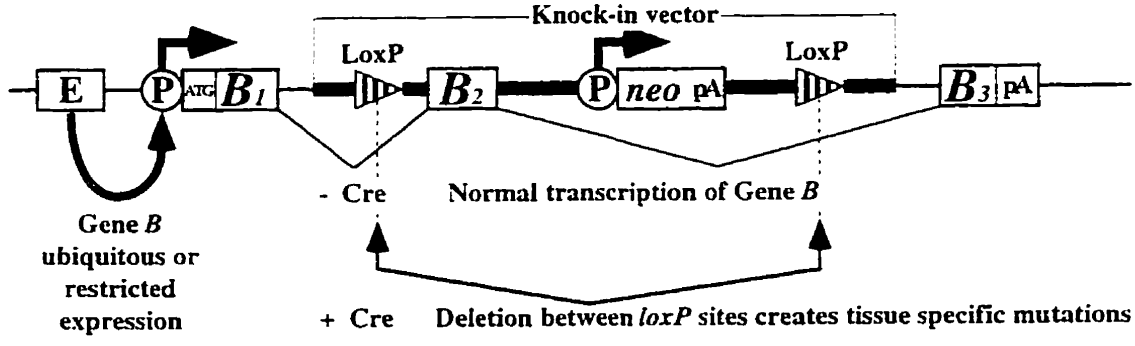
Abbreviations: ßgeo: ß-galactosidase/neomycin fusion gene; E: enhancer; IRES: internal ribosome entry site; P: promoter; pA: poly-adenylation signal; SA: splice acceptor site; SD: splice donor site.

A GENERATING TISSUE SPECIFIC Cre EXPRESSING MOUSE STRAINS

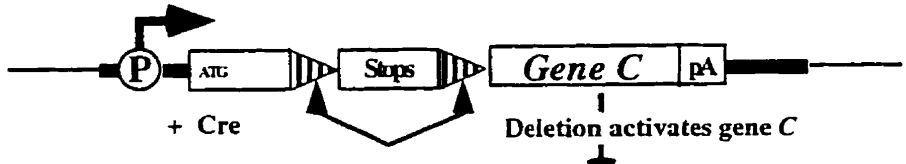


B TARGET SEQUENCES

1. Conditional knock-out



2. Tissue specific gene activation



a. Tissue specific transformation — *Oncogene* — Deletion activates transforming gene

b. Specific in vivo Cre expressing cells ablation — *Toxin* — Deletion activates toxin

c. Tissue specific phenotype rescue — *cDNA* — Deletion activates wild type allele

Future studies

In the gene trap screen presented in Chapter 2, we identified genes by their pattern of expression in 8.5 dpc chimeric embryos. Further analyses could follow 2 approaches: the genetic study of the possible phenotype resulting from the insertion and the cloning and molecular characterization of the trapped genes. The genetic analysis, even limited to the 36 developmentally regulated genes, would pose a number of feasibility problems. One would have to generate germline chimeras for each of the selected ES cell lines, which would require considerable breeding space. It would require that the ES cell clones, some of them generated in 1990, have conserved their totipotence for the colonization of the germline. Finally the average number of gene trap insertions into spatially regulated genes that result in a recessive embryonic lethal mutation is too low to justify this approach. All these potential drawbacks do not make this approach very attractive at this point of the screen. The cloning of the trapped genes would certainly represent the safest approach to characterize further these insertions. For each clone, one would only need that the ES cells still express the reporter gene, whether or not they conserved their totipotence. The cloning could use 5'-RACE-PCR, which is the most current strategy to clone trapped genes. However, other approaches are also possible: RNA samples from all the clones under study could be pooled to make a cDNA library primed with a *lacZ* specific oligonucleotide that should only allow the cloning of fusion transcript sequences. cDNAs corresponding to fusion transcripts could be identified because they would all contain the En-2 exon sequence upstream of *lacZ*. Each unique clone could be analyzed by Northern blot or by whole-mount in situ hybridization at E8.5 to match its expression pattern with one trapping insertions.

USING GENE TRAP INSERTIONS AS DEVELOPMENTAL MARKERS

Gene trap markers

Once the locus carrying the GT insertion of interest has been backcrossed into a mutant strain, detection of the marker expression is technically very easy (most GT vectors carry the *lacZ* reporter gene). Detection of ß-Gal activity is fast and very sensitive in whole-mount embryos up to the stage E12.5. *LacZ* is a very good marker for genetic analyses where few mutant embryos can be generated. Furthermore, the ß-Gal enzymatic activity can amplify the signal given by genes expressed at low levels. In this study and in previous studies where the *GtC101* insertion was

used as a marker to analyze mutations in the *HNF-3ß*, *GAP*, *SHP-2* and *Gli2* genes (Ang and Rossant, 1994; Henkemeyer *et al.*, 1995; Saxton *et al.*, 1997; Ding *et al.*, *submitted*), the insertion was innocuous to the development of control and mutant embryos. In another study of the Brachyury phenotype by Wilson and co-workers (1995), 4 gene trap insertions that expressed the *lacZ* reporter gene ubiquitously were generated *de novo* in BTBR *T/T* and BTBR *T/+* ES cells. They did not alter the developmental potential of the cells nor did they interfere with the cell fate resulting from the *T* mutation.

T and Sd analysis with GtC101

In the study of the *Brachyury* and *Danforth's short tail* mutants, I also compared the phenotypes observed in the presence or absence of the GtC101 insertion. In all cases no difference could be observed according to the genotype at the *cobl* locus except for the intensity of ß-gal signal. The GtC101 GT insertion therefore acted as a neutral genetic marker for the phenotypic analysis of these mutations.

This marker was clearly useful in the identification of early defects in heterozygous and homozygous *Brachyury* and *Sd* mutants. It provides more evidence suggesting that node development is affected in Brachyury mutants. This defect is unlikely to involve the general cell movement defect proposed to explain the Brachyury phenotypes (Wilson *et al.*, 1995) but could indicate patterning defects in the primitive streak. This alternate, or additional, explanation is supported by the chordoneural hinge patterning defects found in the tail bud of heterozygous embryos. In Sd mutants, the formation of the notochord rather than solely its maintenance is altered thus supporting the unification model proposed by Grüneberg (1958).

The maintenance of *cobl* expression in the Brachyury T/T homozygous mice clearly indicated that it did not require the presence of the T transcription factor. Early expression of *cobl* occurred independently in the 3 germ layers, thus showing that *cobl* does not induce its own expression from one tissue to another, but is probably induced by patterning genes. The expression pattern of *cobl* was clearly altered by the effects of the mutations on the development of the notochord: in tissues such as the floor plate and the somites the patterning action of an intact notochord is required for induction of *cobl* expression.

Future studies

Further analyses could use [T^{Wis} +], [T^{Wis} T^{Wis}], [*Sd* +] and [*Sd Sd*] GtC101 GtC101 ES cells to test their developmental potential when competing with wild-type cells in chimeric embryos. These studies have already been performed for the *T* mutation (Rashbass *et al.*, 1991; Wilson *et al.*, 1993; Wilson *et al.*, 1995). These analyses first require the establishment of ES cell lines carrying the proper genotype. In the case of the *Sd* mutation, the genotype of such cell lines could only be inferred by the phenotype of chimeric embryos carrying a large proportion of them. These ES cell lines could help us to investigate further the phenotype of the *Sd* mutation and answer the following questions. Is there any rescue of mutant cells by wild-type cells or is the mutation cell-autonomous? What is the relationship between the notochord and urogenital defects in Sd mutant embryos? What is the potential of *Sd Sd* ES cells to colonize the ureter or the kidneys?

THE CORDON-BLEU GENE

The analysis of the GtC101 gene trap insertion led to the identification of a novel gene, named *cordon-bleu* because of its expression pattern in the embryonic axis. The expression pattern of *cobl* is strikingly similar to that of *HNF3-β* and *Shh* which have been shown to play essential roles in the formation of the node and dorso-ventral patterning of the embryo respectively (Ang and Rossant, 1994; Weinstein *et al.*, 1994; Chiang *et al.*, 1996). However, the GtC101 insertion does not cause any phenotype. The production of the fusion transcript should result in the production of a non-functional truncated Cobl protein fused to LacZ. As the GT vector inserted into an intron, the wild-type protein can still be produced when transcription continues throughout the vector-containing intron and splicing removes the intron containing the GT insert. In this case and if enough wild-type protein is made, the potential phenotype of the mutation may be rescued. In the case of the GtC101 insertion, this situation is suggested by the detection of wild-type *cobl* RNA in homozygous ES cells. However, other explanations are possible. The truncated Cobl peptide may still fulfill the normal function of the entire protein. The function of *cobl* could be redundant with that of another gene. The knock-outs of the *MyoD* and *myf-5* genes provided good evidence for the existence of functional redundancy in the mouse (Rudnicki *et al.*, 1993). Thus the function as well as the importance of *cobl* in the node and axis formation processes remain a conundrum. However, the degree of conservation between the mouse and human

cobl sequences supports the idea that Cobl's function is important. Other node specific genes such as *noggin* (Smith *et al.*, 1993b), *Eck* (Chen *et al.*, 1996) and *chordin* (Sasai *et al.*, 1994) have been characterized in vertebrates. However, apart from the essential role of HNF-3ß in the node, little is known about the hierarchy of the genetic components controlling node development.

Cobl's expression pattern indicates two possible directions for speculations on the plausible functions for *cobl*. First it is almost exclusively expressed in epithelial cells where it could be important for the architecture of the cells: this could involve a role in the elasticity of cells (it is expressed in the notochord, chondrocytes and the skin) or a more general feature of epithelial tissues such as cell adhesion properties. However, the important domains of expression found in the CNS after mid gestation are not completely compatible with these hypotheses. Another possibility is that *cobl* is involved in cell secretion mechanisms in the CNS. Cobl is expressed in signaling tissues (the notochord, the floor plate, chondrocytes) and in epithelial structures (the gut, the tongue, the olfactory epithelium and the skin) that are the centre of important secretory processes.

The sub-cellular localization of cobl in the cytoplasm and/or the nucleus also leaves multiple options regarding the levels at which cobl may be acting (i.e. does cobl function as a transcription regulator in the nucleus, or as structural or enzymatic element in the nucleus or cytoplasm?). Only direct functional studies will be able to resolve these questions.

Future studies on the cordon-bleu gene

Functional analysis

The next priority should be to determine the function of the *cobl* gene. Functional analyses could be performed in the mouse by targeted knock-out of the *cobl* gene or in the frog by injection of the *cobl* cDNA into *Xenopus laevis* eggs.

Only the making of a real null allele of the *cobl* gene will determine whether its function is needed during mouse development. Targeted mutagenesis by homologous recombination is now a standard genetic technique in the mouse. To knock-out the *cobl* gene however, it will be necessary to first clone its genomic sequence in order to build a targeting vector.

Some genes are able to induce the formation of secondary axis or to rescue axis formation in *X. laevis* embryos following their injection into normal or ventralized eggs, respectively. This type of assay has been widely used to determine

if novel genes are involved in the control of axis formation in vertebrates. The full length cDNA is already available for injection into frog eggs. Mouse node proteins have been shown to be able to induce secondary axes in *X. laevis* (Blum *et al.*, 1992); therefore, the frog *cobl* gene may not have to be cloned for this experiment if mouse Cobl is able to function in amphibian embryos.

Structural analysis

The cloning of human *cobl* sequences already allowed the identification of protein domains that appear to be more conserved between the two species. The conservation of such domains may mark structural or functional constraints and presumably indicates important features in a protein. The complete sequencing of the human *cobl* gene and cloning of *cobl* homologues from other species should strengthen this type of analysis. Furthermore, other species may possess proteins containing specific domains of *cobl* or carry mutations at loci corresponding to *cobl* homologues, thus helping the study of *cobl* function.

Another important consequence of a protein's structure is its sub-cellular localization. I showed previously that the Cobl ß-Gal protein could be localized to the nucleus; however, there is a possibility that it is an artifact due to the structure of the chimeric fusion protein. The Cobl protein encoded by the full-length cDNA can be used to raise antiserum. This antiserum should be useful to determine whether native Cobl is also found in the nucleus and to analyze more completely the expression pattern of *cobl*.

Expression study

A gene's expression pattern is another source of information for the unraveling of its function. First, the immunohistochemical analysis would indicate if wild-type Cobl detection matches the domains revealed by the Cobl ß-Gal fusion protein. In another gene trap insertion into the *BTF3* gene, the expression pattern of the *lacZ* reporter gene is restricted to the brain, heart and testis and does not reflect the expression of the disrupted gene which is expressed ubiquitously (Deng and Behringer, 1995).

Expression of *cobl* beyond E12.5 becomes much more complex and was not extensively analyzed. For example, expression in the CNS appears quite dynamic from E12.5 and after birth, and may reflect important aspects of *cobl* function in brain and spinal cord development.

Mutagenic screen around the GtC101 insertion locus

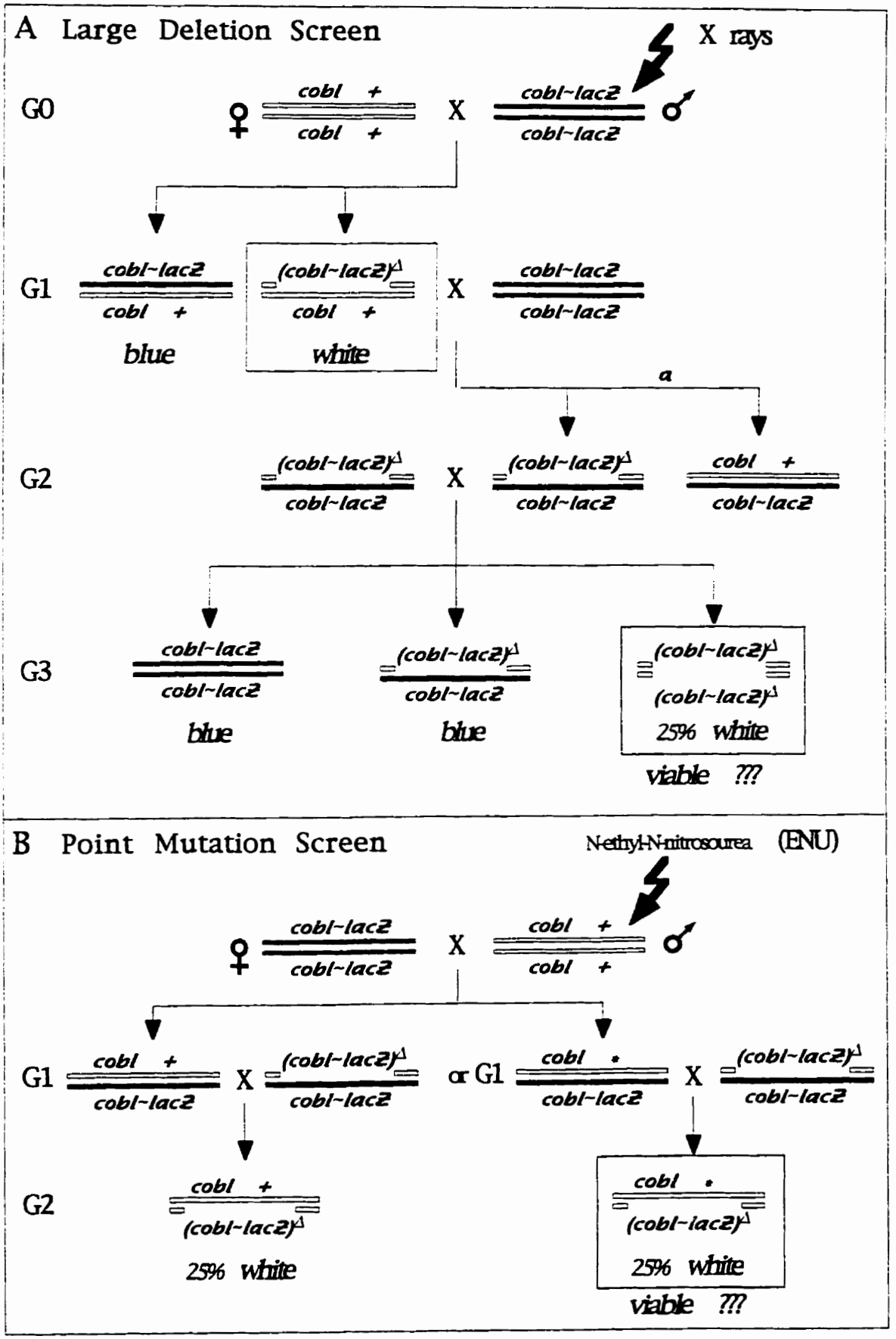
A genetic approach taking advantage of the gene trap insertion could be used as an alternative strategy to characterize the function of *cobl* and other linked genes on chromosome 11. Because expression of Cobl/ß-Gal is observable in skin biopsies from birth into adulthood, it is almost as conspicuous as an external trait. Furthermore, its expression is distinguishable between heterozygous and homozygous tissues. In specific chromosomal regions carrying easily detected markers, mutagenesis screens were designed that allowed the identification of novel mutant alleles (Shedlovsky *et al.*, 1988; Rinchik *et al.*, 1990; Rinchik *et al.*, 1994). In the screen presented in Figure 25, I propose a strategy inspired by these authors, that would allow for a rapid selection of mutations affecting genes linked to the *cobl* locus on chromosome 11. This screen would require a minimum breeding time and space. Mice carrying radiation induced deletions overlapping the *cobl* locus, could be identified in the first generation (G1) by the lack of ß-Gal activity (Fig. 25 A). Only selected animal strains would need to be bred to the third generation to analyze the phenotypes associated with the deletions (Fig. 25 A). A second screen, using chemical mutagens such as N-ethyl-N-nitrosourea (ENU), and targeted to the same region of chromosome 11 could isolate point mutations or small deletions affecting one locus at a time (Russell *et al.*, 1979). The largest deletion generated in the first screen would then be used to identify lethal mutations in the second generation (G2, see Fig. 25 B). .

This type of screen would be theoretically possible for any other gene trap insertion where a biopsy for *lacZ* expression is possible (e.g. gene trap insertion expressing *lacZ* in blood cells).

Fig. 25. - Forward genetic screen using the GtC101 insertion as an anchor locus to detect mutations in *cobl* and in adjacent loci on chromosome 11.

(A) Screen for radiation-induced large deletion. Candidate mice possibly carrying a deletion encompassing the GtC101 locus can be identified in the first generation (G1) by the lack of ß-Gal expression in tail biopsies. Only the potential mutation carrier strains need to be bred to G3 to identify recessive embryonic lethal deletion. (B) Secondary screen for point mutations (using N-ethyl-N-nitrosourea, ENU, as a mutagen) that are not complemented by deletion alleles at the *cobl* locus. Potential carrier, heterozygous for recessive embryonic lethal mutations, will be identified by their lack of G2 *lacZ* negative progeny.

Abbreviations: *cobl*: *cordon-bleu*; *cobl_lacZ*: *GtC101* allele driving ß-Gal expression in the skin; ENU: N-ethyl-N-nitrosourea; Δ indicates a deletion carrying allele; * indicates a point mutation carrying allele; chromosomes are represented by a white bar or by a black bar when carrying an active *lacZ* allele.


A Large Deletion Screen
X rays
G0
cobl +
cobl +
X
cobl~lacZ
cobl~lacZ
G1
cobl~lacZ
cobl +
blue
(cobl~lacZ)Δ
cobl +
white
X
cobl~lacZ
cobl~lacZ
a
G2
(cobl~lacZ)Δ
cobl~lacZ
X
(cobl~lacZ)Δ
cobl~lacZ
cobl +
cobl~lacZ
G3
cobl~lacZ
cobl~lacZ
blue
(cobl~lacZ)Δ
cobl~lacZ
blue
(cobl~lacZ)Δ
(cobl~lacZ)Δ
25% white
viable ???
B Point Mutation Screen
N-ethyl-N-nitrosourea (ENU)
cobl~lacZ
cobl~lacZ
X
cobl +
cobl +
G1
cobl +
cobl~lacZ
X
(cobl~lacZ)Δ
cobl~lacZ
or G1
cobl *
cobl~lacZ
X
(cobl~lacZ)Δ
cobl~lacZ
G2
cobl +
(cobl~lacZ)Δ
25% white
cobl *
(cobl~lacZ)Δ
25% white
viable ???

REFERENCES

Allen, N.D., Cran, D.G., Barton, S.C., Hettle, S., Reik, W. and Surani, M.A. (1988). Transgenes as probes for active chromosomal domains in mouse development. *Nature*. **333**. 852-855.

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

Ang, S.L. and Rossant, J. (1994). HNF-3 beta is essential for node and notochord formation in mouse development. *Cell*. **78**. 561-574.

Ang, S.L., Wierda, A., Wong, D., Stevens, K.A., Cascio, S., Rossant, J. and Zaret, K.S. (1993). The formation and maintenance of the definitive endoderm lineage in the mouse: involvement of HNF3/forkhead proteins. *Development*. **119**. 1301-1315.

Ausubel, F.M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A. and Struhl, K. (1989). *Current Protocols in Molecular Biology*. John Wiley & Sons, New York.

Avner, P., Arnaud, D., Amar, L., Cambrou, J., Winking, H. and Russell, L.B. (1987). Characterization of a panel of somatic cell hybrids for regional mapping of the mouse X chromosome. *Proc Natl Acad Sci USA*. **84**. 5330-5334.

Avraham, K.B., Hasson, T., Steel, K.P., Kingsley, D.M., Russell, L.B., Mooseker, M.S., Copeland, N.G. and Jenkins, N.A. (1995). The mouse Snell's waltzer deafness gene encodes an unconventional myosin required for structural integrity of inner ear hair cells. *Nat Genet*. **11**. 369-375.

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

Baker, R.K. and Lyons, G.E. (1996). Embryonic stem cells and *in vitro* muscle development. *Curr Top Dev Biol*. **33**. 263-279.

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

Beck, T.W., Huleihel, M., Gunnell, M., Bonner, T.I. and Rapp, U.R. (1987). The complete coding sequence of the human A-raf-1 oncogene and transforming activity of a human A-raf carrying retrovirus. *Nucleic Acids Res*. **15**. 595-609.

Beddington, R.S., Rashbass, P. and Wilson, V. (1992). *Brachyury* - a gene affecting mouse gastrulation and early organogenesis. *Development (Suppl)*. 157-165.

Beddington, R.S.P. (1994). Induction of a second neural axis by the mouse node. *Development*. **120**. 613-620.

Beddington, R.S.P. and Smith, J.C. (1993). Control of vertebrate gastrulation: inducing signals and responding genes. *Curr Opin Genet Dev*. **3**. 655-661.

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

Bhattacharyya, M.K., Smith, A.M., Ellis, T.H., Hedley, C. and Martin, C. (1990). The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. *Cell*. **60**. 115-122.

Bier, E., Vaessin, H., Shepherd, S., Lee, K., McCall, K., Barbel, S., Ackerman, L., Carretto, R., Uemura, T., Grell, E. *et al.* (1989). Searching for pattern and mutation in the Drosophila genome with a P- lacZ vector. *Genes Dev*. **3**. 1273-1287.

Blum, M., Gaunt, S.J., Cho, K.W.Y., Steinbeisser, H., Blumberg, B., Bittner, D. and DeRobertis, E.M. (1992). Gastrulation in the mouse: The role of the homeobox gene *goosecoid*. *Cell*. **69**. 1097-1106.

Boer, P.H., Potten, H., Adra, C.N., Jardine, K., Mullhofer, G. and McBurney, M.W. (1990). Polymorphisms in the coding and noncoding regions of murine Pgk-1 alleles. *Biochem Genet*. **28**. 299-308.

Boncinelli, E. and Mallamaci, A. (1995). Homeobox genes in vertebrate gastrulation. *Curr Opin Genet Dev*. **5**. 619-627.

Bonhomme, F., Benmehdi, F., Britton-Davidian, J. and Martin, S. (1979). [Genetic analysis of interspecific crosses Mus musculus L. x Mus spretus Lataste: linkage of Adh-1 with Amy-1 on chromosome 3 and Es-14 with Mod-1 on chromosome 9]. *C R Seances Acad Sci D*. **289**. 545-548.

Bonnerot, C., Legouy, E., Choulika, A. and Nicolas, J.F. (1992). Capture of a cellular transcriptional unit by a retrovirus: mode of provirus activation in embryonal carcinoma cells. *J Virol*. **66**. 4982-4991.

Bonnerot, C., Rocancourt, D., Briand, P., Grimber, G. and Nicolas, J.F. (1987). A beta-galactosidase hybrid protein targeted to nuclei as a marker for developmental studies. *Proc Natl Acad Sci USA*. **84**. 6795-6799.

Bovolenta, P. and Dodd, J. (1991). Perturbation of neuronal differentiation and axon guidance in the spinal cord of mouse embryos lacking a floor plate: analysis of Danforth's short tail mutation. *Development*. **113**. 625-639.

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

Breitman, M.L., Clapoff, S., Rossant, J., Tsui, L.C., Glode, L.M., Maxwell, I.H. and Bernstein, A. (1987). Genetic ablation: targeted expression of a toxin gene causes microphthalmia in transgenic mice. *Science*. **238**. 1563-1565.

Brenner, D.G., Lin-Chao, S. and Cohen, S.N. (1989). Analysis of mammalian cell genetic regulation in situ by using retrovirus-derived "portable exons" carrying the Escherichia coli lacZ gene. *Proc Natl Acad Sci USA*. **86**. 5517-5521.

Brown, A., Copeland, N.G., Gilbert, D.J., Jenkins, N.A., Rossant, J. and Kothary, R. (1994). The genomic structure of an insertional mutation in the Dystonia musculorum locus. *Genomics*. **20**. 371-376.

Bultman, S.J., Michaud, E.J. and Woychik, R.P. (1992). Molecular characterization of the mouse agouti locus. *Cell*. **71**. 1195-1204.

Burgess, D.L., Kohrman, D.C., Galt, J., Plummer, N.W., Jones, J.M., Spear, B. and Meisler, M.H. (1995a). Mutation of a new sodium channel gene, Scn8a, in the mouse mutant 'motor endplate disease'. *Nat Genet*. **10**. 461-465.

Burgess, R., Cserjesi, P., Ligon, K.L. and Olson, E.N. (1995b). Paraxis: a basic helix-loop-helix protein expressed in paraxial mesoderm and developing somites. *Dev Biol*. **168**. 296-306.

Burns, N., Grimwade, B., Ross-Macdonald, P.B., Choi, E.Y., Finberg, K., Roeder, G.S. and Snyder, M. (1994). Large-scale analysis of gene expression, protein localization, and gene disruption in Saccharomyces cerevisiae. *Genes Dev*. **8**. 1087-1105.

Cachon-Gonzalez, M.B., Fenner, S., Coffin, J.M., Moran, C., Best, S. and Stoye, J.P. (1994). Structure and expression of the hairless gene of mice. *Proc Natl Acad Sci USA*. **91**. 7717-7721.

Callahan, C.A. and Thomas, J.B. (1994). Tau-beta-galactosidase, an axon-targeted fusion protein. *Proc Natl Acad Sci USA*. **91**. 5972-5976.

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

Carmeliet, P., Ferreira, V., Breier, G., Pollefeyt, S., Kieckens, L., Gertsenstein, M., Fahrig, M., Vandenhoeck, A., Harpal, K., Eberhardt, C., Declercq, C., Pawling, J., Moons, L., Collen, D., Risau, W. and Nagy, A. (1996). Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele. *Nature*. **380**. 435-439.

Casadaban, M.J. and Cohen, S.N. (1979). Lactose genes fused to exogenous promoters in one step using a Mu-lac bacteriophage: in vivo probe for transcriptional control sequences. *Proc Natl Acad Sci USA*. **76**. 4530-4533.

Castle, W.E. and Allen, G.M. (1903). The heredity of albinism. *Proc Amer Acad Arts Sci*. **38**. 603-622.

Catala, M., Teillet, M.A. and Le Douarin, N.M. (1995). Organization and development of the tail bud analyzed with the quail- chick chimaera system. *Mech Dev.* **51**. 51-65.

Chalfie, M., Tu, Y., Euskirchen, G., Ward, W.W. and Prasher, D.C. (1994). Green fluorescent protein as a marker for gene expression. *Science.* **263**. 802-805.

Chelsky, D., Ralph, R. and Jonak, G. (1989). Sequence requirements for synthetic peptide-mediated translocation to the nucleus. *Mol Cell Biol.* **9**. 2487-2492.

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

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

Chiang, C., Litingtung, Y., Lee, E., Young, K.E., Corden, J.L., Westphal, H. and Beachy, P.A. (1996). Cyclopia and defective axial patterning in mice lacking *Sonic hedgehog* gene function. *Nature.* **383**. 407-413.

Cho, K.W.Y., Blumberg, B., Steinbeisser, H. and De Robertis, E.M. (1991). Molecular nature of the Spemann's Organizer: the role of the Xenopus homoeobox gene *goosecoid*. *Cell.* **67**. 1111-1120.

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

Chu, G. and Sharp, P.A. (1981). A gene chimaera of SV40 and mouse ß-globin is transcribed and properly spliced. *Nature.* **289**. 378-382.

Clements, D., Taylor, H.C., Herrmann, B.G. and Stott, D. (1996). Distinct regulatory control of the Brachyury gene in axial and non-axial mesoderm suggests separation of mesoderm lineages early in mouse gastrulation. *Mech Dev.* **56**. 139-149.

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

Conlon, F.L., Lyons, K.M., Takaesu, N., Barth, K.S., Kispert, A., Herrmann, B. and Robertson, E.J. (1994). A primary requirement for *nodal* in the formation and maintenance of the primitive streak in the mouse. *Development.* **120**. 1919-1928.

Conlon, F.L., Wright, C.V. and Robertson, E.J. (1995). Effects of the *TWis* mutation on notochord formation and mesodermal patterning. *Mech Dev.* **49**. 201-209.

Conlon, R.A. and Rossant, J. (1992). Exogenous retinoic acid rapidly induces anterior ectopic expression of murine Hox-2 genes in vivo. *Development*. **116**. 357-368.

Constantini, F., Radice, G., Lee, J.J., Chada, K.K., Perry, W. and Son, H.J. (1989). Insertional mutation in transgenic mice. *Prog in NAR and Mol Biol*. **36**. 159-169.

Cordes, S.P. and Barsh, G.S. (1994). The mouse segmentation gene kr encodes a novel basic domain-leucine zipper transcription factor. *Cell*. **79**. 1025-1034.

Cserjesi, P., Brown, D., Ligon, K.L., Lyons, G.E., Copeland, N.G., Gilbert, D.J., Jenkins, N.A. and Olson, E.N. (1995). Scleraxis: a basic helix-loop-helix protein that prefigures skeletal formation during mouse embryogenesis. *Development*. **121**. 1099-1110.

Cuénot, L. (1902). La loi de Mendel et l'hérédité de la pigmentation chez la souris. *Arch Zool exp gén 3e sér*. **10**. xxvii-xxx.

Danforth, C.H. (1930). Developmental anomalies in a special strain of mice. *Amer J Anat*. **45**. 275-287.

Day, I.N. (1992). Analysis of the 5'-AAUAAA motif and its flanking sequence in human RNA: relevance to cDNA library sorting. *Gene*. **110**. 245-249.

De Robertis, E.M., Fainsod, A., Gont, L.K. and Steinbeisser, H. (1994). The evolution of vertebrate gastrulation. *Dev Suppl*. 117-124.

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

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

Dietrich, S., Schubert, F.R. and Gruss, P. (1993). Altered Pax gene expression in murine notochord mutants: the notochord is required to initiate and maintain ventral identity in the somite. *Mech Dev*. **44**. 189-207.

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

Ding, Q., Motoyama, J., Gasca, S., Mo, R., Sasaki, H., Rossant, J. and Hui, C.-c. The *Gli2* zinc finger protein is necessary for floor plate differentiation. *Submitted*.

Dobrovolskaïa-Zavadskaïa, N. (1927). Sur la mortification spontanée de la queue chez la souris nouveau-née et sur l'existence d'un charactère héréditaire "non-viable". *C R Soc Biol*. **97**. 114-116.

Doetschman, T.G., Eistetter, M., Katz, M., Schmidt, W. and Kemler, R. (1985). In vitro development of blastocyst-derived embryonic stem cell lines: Formation of visceral yolk sac, blood islands and myocardium. *J Embryol exp Morph.* **87**, 27-45.

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

Dunn, L.C., Gluecksohn-Schoenheimer, S. and Bryson, V. (1940). A new mutation in the mouse affecting spinal column and urogenital system. *J Hered.* **31**, 343-348.

Dymecki, S.M. (1996a). Flp recombinase promotes site-specific DNA recombination in embryonic stem cells and transgenic mice. *Proc Natl Acad Sci USA.* **93**, 6191-6196.

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

Ekker, S.C., Ungar, A.R., Greenstein, P., von Kessler, D.P., Porter, J.A., Moon, R.T. and Beachy, P.A. (1995). Patterning activities of vertebrate hedgehog proteins in the developing eye and brain. *Curr Biol.* **5**, 944-955.

Epstein, D.J., Vekemans, M. and Gros, P. (1991). *Splotch* (Sp^{2H}), a mutation affecting development of the mouse neural tube, shows a deletion within the paired homeodomain of *Pax-3*. *Cell.* **67**, 767-774.

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

Fan, C.-M. and Tessier-Lavigne, M. (1994). Patterning of mammalian somites by surface ectoderm and notochord: Evidence of sclerotome induction by a hedgehog homolog. *Cell.* **79**, 1175-1186.

Fan, C.M., Porter, J.A., Chiang, C., Chang, D.T., Beachy, P.A. and Tessier-Lavigne, M. (1995). Long-range sclerotome induction by sonic hedgehog: direct role of the amino-terminal cleavage product and modulation by the cyclic AMP signaling pathway. *Cell.* **81**, 457-465.

Faust, C. and Magnuson, T. (1993). Genetic control of gastrulation in the mouse. *Curr Opin Genet Dev.* **3**, 491-498.

Faust, C., Schumacher, A., Holdener, B. and Magnuson, T. (1995). The eed mutation disrupts anterior mesoderm production in mice. *Development.* **121**, 273-285.

Feinberg, A. and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction fragments to high specific activity. *Anal Biochem.* **132**, 6-13.

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

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

Frohman, M.A., Dush, M.K. and Martin, G.R. (1988). Rapid amplification of cDNA ends (RACE): a new method for obtaining full-length cDNA clones. *Proc Natl Acad Sci USA*, **85**, 8998-9002.

Fujimoto, H. and Yanagisawa, K.O. (1983). Defects in the archenteron of mouse embryos homozygous for the T- mutation. *Differentiation*, **25**, 44-47.

Garcia-Bustos, J., Heitman, J. and Hall, M.N. (1991). Nuclear protein localization. *Biochim Biophys Acta*, **1071**, 83-101.

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

Gaunt, S.J., Blum, M. and De Robertis, E.M. (1993). Expression of the mouse goosecoid gene during mid-embryogenesis may mark mesenchymal cell lineages in the developing head, limbs and body wall. *Development*, **117(2)**, 769-778.

Geissler, E.N., Ryan, M.A. and Housman, D.E. (1988). The dominant-white spotting (W) locus of the mouse encodes the c-kit proto-oncogene. *Cell*, **55**, 185-192.

Gilbert, S.F. (1994). *Developmental biology*. Sinauer Associates, Sunderland.

Gluecksohn-Schoenheimer, S. (1943). The morphological manifestation of a dominant mutation affecting tail and urogenital system. *Genetics*, **28**, 341-348.

Goebl, M.G. and Petes, T.D. (1986). Most of the yeast genomic sequences are not essential for cell growth and division. *Cell*, **46**, 983-992.

Gont, L.K., Steinbeisser, H., Blumberg, B. and de Robertis, E.M. (1993). Tail formation as a continuation of gastrulation: the multiple cell populations of the Xenopus tailbud derive from the late blastopore lip. *Development*, **119**, 991-1004.

Goring, D.R., Rossant, J., Clapoff, S., Breitman, M.L. and Tsui, L.C. (1987). In situ detection of beta-galactosidase in lenses of transgenic mice with a gamma-crystallin/lacZ gene. *Science*, **235**, 456-458.

Gossler, A., Doetschman, T., Korn, R., Serfling, E. and Kemler, R. (1986). Transgenesis by means of blastocyst-derived embryonic stem cell lines. *Proc Natl Acad Sci USA*, **83**, 9065-9069.

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

Gossler, A. and Zachgo, J. (1993). Gene and enhancer trap screens in ES cell chimeras. In *Gene Targeting: A practical approach.* (Ed. A.J. Joyner). 181-227. Oxford University Press, Oxford, UK.

Green, M.C. (1989). Catalog of mutants genes and polymorphic loci. In *Genetic variants and strains of the laboratory mouse.* (Ed. M.F.L.a.A.G. Searle). 12-403. Oxford University Press, Oxford, UK.

Gridley, T. (1991). Insertional versus targeted mutagenesis in mice. *New Biol.* **3**, 1025-1034.

Griffith, C.M., Wiley, M.J. and Sanders, E.J. (1992). The vertebrate tail bud: three germ layers from one tissue. *Anat Embryol (Berl).* **185**, 101-113.

Grüneberg, H. (1956). A ventral ectodermal ridge of the tail in mouse embryos. *Nature.* **177**, 787-788.

Grüneberg, H. (1958). Genetical studies on the skeleton of the mouse - XXII. The development of Danforth's short-tail. *J Embryol exp Morph.* **6**, 124-148.

Grunwald, D.J. and Streisinger, G. (1992). Induction of recessive lethal and specific locus mutations in the zebrafish with ethyl nitrosourea. *Genet Res.* **59**, 103-116.

Gu, H., Marth, J.D., Orban, P.C., Mossmann, H. and Rajewsky, K. (1994). Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting [see comments]. *Science.* **265**, 103-106.

Gubbay, J., Koopman, P., Collignon, J., Burgoyne, P. and Lovell-Badge, R. (1990). Normal structure and expression of Zfy genes in XY female mice mutant in Tdy. *Development.* **109**, 647-653.

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

Haldane, J.B.S., Spurt, A.D. and Haldane, N.M. (1915). Reduplication in mice. *J Genet.* **5**, 133-135.

Halpern, M.E., Ho, R.K., Walker, C. and Kimmel, C.B. (1993). Induction of muscle pioneers and floor plate is distinguished by the zebrafish no tail mutation. *Cell.* **75**, 99-111.

Hamada, H. (1986). Random isolation of gene activator elements from the human genome. *Mol Cell Biol.* **6**, 4185-4194.

Henkemeyer, M., Rossi, D.J., Holmyard, D.P., Puri, M.C., Mbamalu, G., Harpal, K., Shih, T.S., Jacks, T. and Pawson, T. (1995). Vascular system defects and neuronal apoptosis in mice lacking ras GTPase-activating protein. *Nature*. **377**, 695-701.

Herrmann, B.G. (1991). Expression pattern of the Brachyury gene in whole-mount TWis/TWis mutant embryos. *Development*. **113**, 913-917.

Herrmann, B.G. and Kispert, A. (1994). The T genes in embryogenesis. *Trends Genet*. **10**, 280-286.

Herrmann, B.G., Labeit, S., Poustka, A., King, T.R. and Lehrach, H. (1990). Cloning of the *T* gene required in mesoderm formation in the mouse. *Nature*. **343**, 617-622.

Hill, D.P. and Wurst, W. (1993). Gene and enhancer trapping: mutagenic strategies for developmental studies. *Curr Top Dev Biol*. **28**, 181-206.

Hill, R.E., Favor, J., Hogan, B.L., Ton, C.C., Saunders, G.F., Hanson, I.M., Prosser, J., Jordan, T., Hastie, N.D. and van Heyningen, V. (1991). 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**, 522-525.

Hitotsumachi, S., Carpenter, D.A. and Russell, W.L. (1985). Dose-repetition increases the mutagenic effectiveness of N-ethyl-N- nitrosourea in mouse spermatogonia. *Proc Natl Acad Sci USA*. **82**, 6619-6621.

Holmdahl, D.E. (1925). Experimentelle untersuchungen über die lage der grenze swischen primärer und sekundärer körperentwicklung beim huhn. *Anat Anz*. **59**, 393-396.

Holmdahl, D.E. (1939). Die morphogenese des vertebratenorganismus vom formalen und experimentellen gesichtspunkt. *Wilhem Roux Arch Entwmech Orgs*. **139**, 191-226.

Hope, I.A. (1991). 'Promoter trapping' in Caenorhabditis elegans. *Development*. **113**, 399-408.

Hope, I.A. (1994). PES-1 is expressed during early embryogenesis in Caenorhabditis elegans and has homology to the fork head family of transcription factors. *Development*. **120**, 505-514.

Hui, C.-c. and Joyner, A.L. (1993). A mouse model of Greig cephalopolysyndactyly syndrome: the extra-toes mutation contains an intragenic deletion of the *Gli-3* gene. *Nat Genet*. **3**, 241-246.

Huszar, D., Balling, R., Kothary, R., Magli, M.C., Hozumi, N., Rossant, J. and Bernstein, A. (1985). Insertion of a bacterial gene into the mouse germ line using an infectious retrovirus vector. *Proc Natl Acad Sci USA*. **82**, 8587-8591.

Imai, Y., Suzuki, Y., Matsui, T., Tohyama, M., Wanaka, A. and Takagi, T. (1995). Cloning of a retinoic acid-induced gene, GT1, in the embryonal carcinoma cell line P19: neuron-specific expression in the mouse brain. *Brain Res Mol Brain Res.* **31**, 1-9.

Iseki, S., Araga, A., Ohuchi, H., Nohno, T., Yoshioka, H., Hayashi, F. and Noji, S. (1996). Sonic hedgehog is expressed in epithelial cells during development of whisker, hair, and tooth. *Biochem Biophys Res Commun.* **218**, 688-693.

Jaenisch, R. (1988). Transgenic animals. *Science.* **240**, 1468-1474.

Jaenisch, R. and Soriano, P. (1986). Retroviruses as tools for mammalian development. *Symp Fundam Cancer Res.* **39**, 59-65.

Jeannotte, L., Ruiz, J.C. and Robertson, E.J. (1991). Low level of Hox1.3 gene expression does not preclude the use of promoterless vectors to generate a targeted gene disruption. *Mol Cell Biol.* **11**, 5578-5585.

Jessell, T.M. and Dodd, J. (1992). Floor-plate derived signals and the control of neural cell pattern in vertebrates. *Harvey Lect.* **86**, 87-128.

Johnson, R.L., Laufer, E., Riddle, R.D. and Tabin, C. (1994). Ectopic expression of *Sonic hedgehog* alters dorsal-ventral patterning of somites. *Cell.* **79**, 1165-1173.

Joyner, A.L., Kornberg, T., Coleman, K.G., Cox, D.R. and Martin, G.R. (1985). Expression during embryogenesis of a mouse gene with sequence homology to the Drosophila engrailed gene. *Cell.* **43**, 29-37.

Joyner, A.L. and Martin, G.R. (1987). En-1 and En-2, two mouse genes with sequence homology to the Drosophila engrailed gene: expression during embryogenesis [published erratum appears in Genes Dev 1987 Jul;1(5):521]. *Genes Dev.* **1**, 29-38.

Kelley, M.R., Kidd, S., Berg, R.L. and Young, M.W. (1987). Restriction of P-element insertions at the Notch locus of Drosophila melanogaster. *Mol Cell Biol.* **7**, 1545-1548.

Kerr, W.G., Nolan, G.P. and Herzenberg, L.A. (1989). In situ detection of transcriptionally active chromatin and genetic regulatory elements in individual viable mammalian cells. *Immunol Suppl.* **2**, 74-78.

Kispert, A. and Herrmann, B.G. (1993). The Brachyury gene encodes a novel DNA binding protein. *EMBO J.* **12**, 3211-3220.

Klimyuk, V.I., Nussaume, L., Harrison, K. and Jones, J.D. (1995). Novel GUS expression patterns following transposition of an enhancer trap Ds element in Arabidopsis. *Mol Gen Genet.* **249**, 357-365.

Kohrman, D.C., Plummer, N.W., Schuster, T., Jones, J.M., Jang, W., Burgess, D.L., Galt, J., Spear, B.T. and Meisler, M.H. (1995). Insertional mutation of the motor endplate disease (med) locus on mouse chromosome 15. *Genomics.* **26**, 171-177.

Koller, B.H., Hagemann, L.J., Doetschman, T., Hagaman, J.R., Huang, S., Williams, P.J., First, N.L., Maeda, N. and Smithies, O. (1989). Germline transmission of a planned alteration made in a hypoxanthine phosphoribosyltransferase gene by homologous recombination in embryonic stem cells. *Proc Natl Acad Sci USA.* **86**, 8927-8931.

Koller, B.H. and Smithies, O. (1992). Altering genes in animals by gene targeting. *Annu Rev Immunol.* **10**, 705-730.

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

Koseki, H., Wallin, J., Wilting, J., Mizutani, Y., Kispert, A., Ebensperger, C., Herrmann, B.G., Christ, B. and Balling, R. (1993). A role for Pax-1 as a mediator of notochordal signals during the dorsoventral specification of vertebrae. *Development.* **119**, 649-660.

Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15**, 8125-8148.

Krumlauf, R. (1994). Hox genes in vertebrate development. *Cell.* **70**, 191-201.

Lawson, K.A., Meneses, J.J. and Pedersen, R.A. (1986). Cell fate and cell lineage in the endoderm of the presomite mouse embryo, studied with an intracellular tracer. *Dev Biol.* **115**, 325-339.

Lawson, K.A., Meneses, J.J. and Pedersen, R.A. (1991). Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development.* **113**, 891-911.

Le Douarin, N.M. (1969). Particularités du noyau interphasique chez la caille japonaise (*Coturnix coturnix japonica*). Utilisation de ces particularités comme "marquage biologique" dans les recherches sur les interactions tissulaires et les migrations cellulaires au cours de l'ontogénèse. *Bull Biol Fr Belg.* **103**, 435-452.

Le Douarin, N.M. (1973). A biological cell labelling technique and its use in experimental embryology. *Dev Biol.* **30**, 217-222.

Leder, A., Swan, D., Ruddle, F., D'Eustachio, P. and Leder, P. (1981). Dispersion of alpha-like globin genes of the mouse to three different chromosomes. *Nature.* **293**, 196-200.

Lemaire, P. and Kodjabachian, L. (1996). The vertebrate organizer: structure and molecules. *Trends Genet.* **12**, 1-7.

Lyon, M.F. and Searle, A.G. (1989). *Genetic variants and strains of the laboratory mouse.* Oxford University Press, Oxford.

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

Mahon, K.A., Overbeek, P.A. and Westphal, H. (1988). Prenatal lethality in a transgenic mouse line is the result of a chromosomal translocation. *Proc Natl Acad Sci USA.* **85**. 1165-1168.

Marti, E., Takada, R., Bumcrot, D.A., Sasaki, H. and McMahon, A.P. (1995). Distribution of Sonic hedgehog peptides in the developing chick and mouse embryo. *Development.* **121**. 2537-2547.

Martin, G.R. (1981). Isolation of a pluripotent stem cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci USA.* **78**. 7634-7638.

McCarthy, L., Hunter, K., Schalkwyk, L., Riba, L., Anson, S., Mott, R., Newell, W., Bruley, C., Bar, I., Ramu, E. *et al.* (1995). Efficient high-resolution genetic mapping of mouse interspersed repetitive sequence PCR products, toward integrated genetic and physical mapping of the mouse genome. *Proc Natl Acad Sci USA.* **92**. 5302-5306.

McClintock, B. (1950). The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci USA.* **36**. 344-355.

McMahon, A.P. and Bradley, A. (1990). The Wnt-1 (int-1) proto-oncogene is required for development of a large region of the mouse brain. *Cell.* **62**. 1073-1085.

Meneely, P.M. and Herman, R.K. (1979). Lethals, steriles and deficiencies in a region of the X chromosome of *Caenorhabditis elegans*. *Genetics.* **92**. 99-115.

Michaud, E.J., Bultman, S.J., Stubbs, L.J. and Woychik, R.P. (1993). The embryonic lethality of homozygous lethal yellow mice (Ay/Ay) is associated with the disruption of a novel RNA-binding protein. *Genes Dev.* **7**. 1203-1213.

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

Mullins, M.C., Hammerschmidt, M., Haffter, P. and Nusslein, V.C. (1994). Large-scale mutagenesis in the zebrafish: in search of genes controlling development in a vertebrate. *Curr Biol.* **4**. 189-202.

Nagy, A., Gocza, E., Diaz, E.M., Prideaux, V.R., Ivanyi, E., Markkula, M. and Rossant, J. (1990). Embryonic stem cells alone are able to support fetal development in the mouse. *Development.* **110**. 815-821.

Nagy, A. and Rossant, J. (1993). Production of completely ES cell-derived fetuses. In *Gene targeting: a practical approach.* (Ed. A.L. Joyner). 147-179. Oxford University Press, Oxford, UK.

Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. and Roder, J.C. (1993). Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc Natl Acad Sci USA*, **90**, 8424-8428.

Natarajan, D. and Boulter, C.A. (1995). A lacZ-hygromycin fusion gene and its use in a gene trap vector for marking embryonic stem cells. *Nucleic Acids Res.* **23**, 4003-4004.

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

Niswander, L., Yee, D., Rinchik, E.M., Russell, L.B. and Magnuson, T. (1988). The *albino* deletion complex and early postimplantation survival in the mouse. *Development.* **102**, 45-53.

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

Nusslein-Volhard, C., Frohnhofer, H.G. and Lehmann, R. (1987). Determination of anteroposterior polarity in Drosophila. *Science.* **238**, 1675-1681.

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

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

Okazaki, S., Tanase, S., Choudhury, B.K., Setoyama, K., Miura, R., Ogawa, M. and Setoyama, C. (1994). A novel nuclear protein with zinc fingers down-regulated during early mammalian cell differentiation. *J Biol Chem.* **269**, 6900-6907.

Paavola, L.G., Wilson, D.B. and Center, E.B. (1980). Histochemistry of the developing notochord, perichordal sheath and vertebrae in Danforth's short-tail (Sd) and normal C57BL/6 mice. *J Embryol exp Morph.* **55**, 227-245.

Pasteels, J. (1939). La formation de la queue chez les vertébrés. *Ann Soc Royale Zool Belgique.* **70**, 33-51.

Pasteels, J. (1943). Proliférations et croissance dans la gastrulation et la formation de la queue des vertébrés. *Arch Biol.* **54**, 1-51.

Pearson, W. and Lipman, D.J. (1988). Improved tools for biological sequence comparison. *Proc Natl Acad Sci USA.* **85**, 2444-2448.

Phelps, D.E. and Dressler, G.R. (1993). Aberrant expression of Pax-2 in Danforth's short tail (Sd) mice. *Dev Biol.* **157**, 251-258.

Placzeck, M. (1995). The role of the notochord and floor plate in inductive interactions. *Curr Opin Genet Dev.* **5**, 499-506.

Placzek, M., Tessier-Lavigne, M., Yamada, T., Jessel, T. and Dodd, J. (1990). Mesodermal control of neural cell identity: floor plate induction by the notochord. *Science.* **250**, 985-988.

Pourquié, O., Coltey, M., Teillet, M.A., Ordahl, C. and Le Douarin, N.M. (1993). Control of dorsoventral patterning of somitic derivatives by notochord and floor plate. *Proc Natl Acad Sci USA.* **90**, 5242-5246.

Pownall, M.E., Strunk, K.E. and Emerson Jr , C.P. (1996). Notochord signals control the transcriptional cascade of myogenic bHLH genes in somites of quail embryos. *Development.* **122**, 1475-1488.

Pretsch, W. and Charles, D.J. (1984). Detection of dominant enzyme mutants in mice: model studies for mutations in man. *IARC Sci Publ.* 361-369.

Rashbass, P., Cooke, L.A., Herrmann, B.G. and Beddington, R.S. (1991). A cell autonomous function of Brachyury in *T/T* embryonic stem cell chimaeras. *Nature.* **353**, 348-351.

Rashbass, P., Wilson, V., Rosen, B. and Beddington, R.S. (1994). Alterations in gene expression during mesoderm formation and axial patterning in *Brachyury (T)* embryos. *Int J Dev Biol.* **38**, 35-44.

Reddy, S., DeGregori, J.V., von Melchner, H. and Ruley, H.E. (1991). Retrovirus promoter-trap vector to induce lacZ gene fusions in mammalian cells. *J Virology.* **65**, 1507-1515.

Reddy, S., Rayburn, H., von Melchner, H. and Ruley, H.E. (1992). Fluorescence-activated sorting of totipotent embryonic stem cells expressing developmentally regulated lacZ fusion genes. *Proc Natl Acad Sci USA.* **89**, 6721-6725.

Riddle, R.D., Johnson, R.L., Laufer, E. and Tabin, C. (1993). *Sonic hedgehog* mediates the polarizing activity of the ZPA. *Cell.* **75**, 1401-1416.

Riesgo-Escovar, J., Woodard, C., Gaines, P. and Carlson, J. (1992). Development and organization of the Drosophila olfactory system: an analysis using enhancer traps. *J Neurobiol.* **23**, 947-964.

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

Rinchik, E.M., Bell, J.A., Hunsicker, P.R., Friedman, J.M., Jackson, I.J. and Russell, L.B. (1994). Molecular genetics of the brown (b)-locus region of mouse chromosome 4. I. Origin and molecular mapping of radiation- and chemical-induced lethal brown deletions. *Genetics.* **137**, 845-854.

Rinchik, E.M., Carpenter, D.A. and Selby, P.B. (1990). A strategy for fine-structure functional analysis of a 6- to 11- centimorgan region of mouse chromosome 7 by high-efficiency mutagenesis. *Proc Natl Acad Sci USA.* **87**, 896-900.

Rivera-Perez, J.A., Mallo, M., Gendron-Maguire, M., Gridley, T. and Behringer, R.R. (1995). Goosecoid is not an essential component of the mouse gastrula organizer but is required for craniofacial and rib development. *Development.* **121**. 3005-3012.

Roelink, H., Augsburger, A., Heemskerk, J., Korzh, V., Norlin, S., Ruiz i Altaba, A., Tanabe, Y., Placzek, M., Edlund, T. and Jessell, T.M. (1994). Floor plate and motor neuron induction by vhh-1, a vertebrate homolog of hedgehog expressed by the notochord. *Cell.* **76**. 761-775.

Rohdewohld, H., Weiher, H., Reik, W., Jaenisch, R. and Breindl, M. (1987). Retrovirus integration and chromatin structure: Moloney murine leukemia proviral integration sites map near DNase I-hypersensitive sites. *J Virol.* **61**. 336-343.

Rong, P.M., Teillet, M.A., Ziller, C. and Le Douarin, N.M. (1992). The neural tube/notochord complex is necessary for vertebral but not limb and body wall striated muscle differentiation. *Development.* **115**. 657-672.

Rossant, J. and Hopkins, N. (1992). Of fin and fur: mutational analysis of vertebrate embryonic development. *Genes Dev.* **6**. 1-13.

Rowe, L.B., Nadeau, J.H., Turner, R., Frankel, W.N., Letts, V.A., Eppig, J.T., Ko, M.S.H., Thurston, S.J. and Birkenmeier, E.H. (1994). Maps from two interspecific backcross DNA panels available as a community genetic mapping resource. *Mamm Genome.* **5**. 253-274.

Rubenstein, J.L., Nicolas, J.F. and Jacob, F. (1984). Construction of a retrovirus capable of transducing and expressing genes in multipotential embryonic cells. *Proc Natl Acad Sci USA.* **81**. 7137-7140.

Rudnicki, M.A., Schnegelsberg, P., Stead, R., Braun, T., Arnold, H. and Jaenisch, R. (1993). *MyoD* or *myf-5* is required for the formation of skeletal muscle. *Cell.* **75**. 1351-1359.

Ruiz i Altaba, A., Cox, C., Jessell, T.M. and Klar, A. (1993a). Ectopic neural expression of a floor plate marker in frog embryos injected with the midline transcription factor Pintallavis. *Proc Natl Acad Sci USA.* **90**. 8268-8272.

Ruiz i Altaba, A. and Jessell, T.M. (1993). Midline cells and the organization of the vertebrate neuraxis. *Curr Opin Genet Dev.* **3**. 633-640.

Ruiz i Altaba, A., Prezioso, V.R., Darnell, J.E. and Jessell, T.M. (1993b). Sequential expression of HNF-3 beta and HNF-3 alpha by embryonic organizing centers: the dorsal lip/node, notochord and floor plate. *Mech Dev.* **44**. 91-108.

Russell, L.B., Montgomery, C.S. and Raymer, G.D. (1982). Analysis of the *albino*-locus region of the mouse: IV. Characterization of 34 deficiencies. *Genetics.* **100**. 427-453.

Russell, L.B., Selby, P.B., von Halle, E., Sheridan, W. and Valcovic, L. (1981). The mouse specific-locus test with agents other than radiations: interpretation of data and recommendations for future work. *Mutat Res.* **86**. 329-354.

Russell, W.L., Kelly, E.M., Hunsicker, P.R., Bangham, J.W., Maddux, S.C. and Phipps, E.L. (1979). Specific-locus test shows ethylnitrosourea to be the most potent mutagen in the mouse. *Proc Natl Acad Sci USA.* **76**. 5818-5819.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular cloning: A laboratory manual. Second Edition.* Cold Spring Harbor Laboratory Press. New York.

Sanes, J.R., Rubenstein, J.L. and Nicolas, J.F. (1986). Use of a recombinant retrovirus to study post-implantation cell lineage in mouse embryos. *EMBO J.* **5**. 3133-3142.

Sasai, Y., Lu, B., Steinbeisser, H., Geissert, D., Gont, L.K. and De Robertis, E.M. (1994). Xenopus chordin: a novel dorsalizing factor activated by organizer- specific homeobox genes. *Cell.* **79**. 779-790.

Sasaki, H. and Hogan, B.L. (1993). Differential expression of multiple fork head related genes during gastrulation and axial pattern formation in the mouse embryo. *Development.* **118**. 47-59.

Sasaki, H. and Hogan, B.L. (1994). HNF-3 beta as a regulator of floor plate development. *Cell.* **76**. 103-115.

Sauer, B. and Henderson, N. (1989). Cre-stimulated recombination at loxP-containing DNA sequences placed into the mammalian genome. *Nucleic Acids Res.* **17**. 147-161.

Saxton, T.M., Henkemeyer, M., Gasca, S., Shen, R., Rossi, D., Shalaby, F., Gen-Sheng, F. and Pawson, T. (1997). Abnormal mesoderm patterning in mouse embryos mutant for the SH2 tyrosine phosphatase Shp-2. *EMBO J.* **16**. 2352-2364.

Scherdin, U., Rhodes, K. and Breindl, M. (1990). Transcriptionally active genome regions are preferred targets for retroviral integration. *Mol Cell Biol.* **64**. 907-912.

Schierenberg, E., Miwa, J. and von Ehrenstein, G. (1980). Cell lineages and developmental defects of temperature-sensitive embryonic arrest mutants in *Caenorhabditis elegans*. *Dev Biol.* **76**. 141-159.

Schulte-Merker, S., van Eeden, F.J., Halpern, M.E., Kimmel, C.B. and Nusslein-Volhard, C. (1994). *No tail (ntl)* is the zebrafish homologue of the mouse *T (Brachyury)* gene. *Development.* **120**. 1009-1015.

Schumacher, A., Faust, C. and Magnuson, T. (1996). Positional cloning of a global regulator of anterior-posterior patterning in mice. *Nature.* **383**. 250-253.

Selleck, M.A.J. and Stern, C.D. (1991). Fate mapping and cell lineage analysis of Hensen's node in the chick embryo. *Development*. **112**. 615-626.

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

Sheets, M.D., Ogg, S.C. and Wickens, M.P. (1990). Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Res*. **18**. 5799-5805.

Shih, C.-C., Stoye, J.P. and Coffin, J.M. (1988). Highly preferred targets for retrovirus integration. *Cell*. **53**. 531-537.

Silvers, W.K. (1979). *The coat colors of mice. A model for mammalian gene action and interaction*. Springer-Verlag. New York.

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

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

Slack, J.M.W. (1991). *From Egg to Embryo : Regional Specification in Early Development*. Cambridge University Press. Cambridge. UK.

Smith, D., Wohlgemuth, J., Calvi, B.R., Franklin, I. and Gelbart, W.M. (1993a). hobo enhancer trapping mutagenesis in Drosophila reveals an insertion specificity different from P elements. *Genetics*. **135**. 1063-1076.

Smith, W.C., Knecht, A.K., Wu, M. and Harland, R.M. (1993b). Secreted *noggin* protein mimics the Spemann organizer in dorsalizing *Xenopus* mesoderm. *Nature*. **361**. 547-549.

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

Solnica-Krezel, L., Schier, A.F. and Driever, W. (1994). Efficient recovery of ENU-induced mutations from the zebrafish germline. *Genetics*. **136**. 1401-1420.

Soriano, P., Cone, R.D., Mulligan, R.C. and Jaenisch, R. (1986). Tissue-specific and ectopic expression of genes introduced into transgenic mice by retroviruses. *Science*. **234**. 1409-1413.

Soriano, P., Gridley, T. and Jaenisch, R. (1987). Retroviruses and insertional mutagenesis in mice: proviral integration at the Mov 34 locus leads to early embryonic death. *Genes Dev*. **1**. 366-375.

Spemann, H. and Mangold, H. (1924). Über induktion von embryonenanlagen durch implantation artfremder organisatoren. *Arch für microscopische Anat und Entwmech.* **100**. 599-638.

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

Springer, P.S., McCombie, W.R., Sundaresan, V. and Martienssen, R.A. (1995). Gene trap tagging of PROLIFERA, an essential MCM2-3-5-like gene in Arabidopsis. *Science.* **268**. 877-880.

Stern, H.M. and Hauschka, S.D. (1995). Neural tube and notochord promote in vitro myogenesis in single somite explants. *Dev Biol.* **167**. 87-103.

Streit, A., Thery, C. and Stern, C.D. (1993). Of mice and frogs. *Trends Genet.* **10**. 181-183.

Sundaresan, V., Springer, P., Volpe, T., Haward, S., Jones, J.D., Dean, C., Ma, H. and Martienssen, R. (1995). Patterns of gene action in plant development revealed by enhancer trap and gene trap transposable elements. *Genes Dev.* **9**. 1797-1810.

Sutherland, H.F., Lovell-Badge, R.H. and Jackson, I.J. (1993). Characterisation of two identical independent non-homologous integration sites in mouse embryonic stem cells. *Gene.* **131**. 265-268.

Takeuchi, T., Yamazaki, Y., Katoh-Fukui, Y., Tsuchiya, R., Kondo, S., Motoyama, J. and Higashinakagawa, T. (1995). Gene trap capture of a novel mouse gene, jumonji, required for neural tube formation. *Genes Dev.* **9275**. 1211-1222.

Tam, P.P. and Tan, S.S. (1992). The somitogenetic potential of cells in the primitive streak and the tail bud of the organogenesis-stage mouse embryo. *Development.* **115**. 703-715.

Theiler, K. (1988). Vertebral malformations. In *Advances in anatomy, embryology, and cell biology.* 112. (Ed. F. Beck, W. Hild, W. Kriz, R. Ortmann, J.E. Pauly and T.H. Schiebler). 1-99. Springer-Verlag, Berlin Heidelberg.

Tsubota, S., Ashburner, M. and Schedl, P. (1985). P-element-induced control mutations at the r gene of Drosophila melanogaster. *Mol Cell Biol.* **5**. 2567-2574.

Tucker, A.S. and Slack, J.M.W. (1995). The *Xenopus laevis* tail-forming region. *Development.* **121**. 249-262.

van Straaten, H.W., Hekking, J.W., Wiertz-Hoessels, E.J., Thors, F. and Drukker, J. (1988). Effect of the notochord on the differentiation of a floor plate area in the neural tube of the chick embryo. *Anat Embryol (Berl).* **177**. 317-324.

von Melchner, H., DeGregori, J.V., Rayburn, H., Reddy, S., Friedel, C. and Ruley, H.E. (1992). Selective disruption of genes expressed in totipotent embryonal stem cells. *Genes Dev.* **6**, 919-927.

von Melchner, H., Reddy, S. and Ruley, H.E. (1990). Isolation of cellular promoters by using a retrovirus promoter trap. *Proc Natl Acad Sci USA.* **87**, 3733-3737.

Waddington, C.H. (1933). Induction of the primitive streak and its derivatives in the chick. *J Exp Biol.* **10**, 38-46.

Weinstein, D.C., Ruiz i Altaba, A., Chen, W.S., Hoodless, P., Prezioso, V.R., Jessell, T.M. and Darnell, J.E., Jr. (1994). The winged-helix transcription factor HNF-3 beta is required for notochord development in the mouse embryo. *Cell.* **78**, 575-588.

Wilkie, T.M., Braun, R.E., Ehrman, W.J., Palmiter, R.D. and Hammer, R.E. (1991). Germ-line intrachromosomal recombination restores fertility in transgenic MyK-103 male mice. *Genes Dev.* **5**, 38-48.

Wilkinson, D.G., Bhatt, S. and Herrmann, B.G. (1990). Expression pattern of the mouse T gene and its role in mesoderm formation [see comments]. *Nature.* **343**, 657-659.

Wilson, V. and Beddington, R.S.P. (1996). Cell fate and morphogenetic movement in the late mouse primitive streak. *Mech Dev.* **55**, 79-89.

Wilson, V., Manson, L., Skarnes, W.C. and Beddington, R.S. (1995). The T gene is necessary for normal mesodermal morphogenetic cell movements during gastrulation. *Development.* **121**, 877-886.

Wilson, V., Rashbass, P. and Beddington, R.S. (1993). Chimeric analysis of T (Brachyury) gene function. *Development.* **117**, 1321-1331.

Wurst, W. and Joyner, A.L. (1993). Production of targeted embryonic stem cell clones. In *Gene targeting: A practical approach.* (Ed. A.L. Joyner), 33-61. Oxford University Press, Oxford.

Wurst, W., Rossant, J., Prideaux, V., Kownacka, M., Joyner, A., Hill, D., Guillemot, F., Gasca, S., Cado, D., Auerbach, A. and Ang, S.-L. (1995). A large-scale gene-trap screen for insertional mutations in developmentally regulated genes in mice. *Genetics.* **139**, 889-899.

Yamada, G., Mansouri, A., Torres, M., Stuart, E.T., Blum, M., Schultz, M., De Robertis, E.M. and Gruss, P. (1995). Targeted mutation of the murine goosecoid gene results in craniofacial defects and neonatal death. *Development.* **121**, 2917-2922.

Yang, M.Y., Armstrong, J.D., Vilinsky, I., Strausfeld, N.J. and Kaiser, K. (1995). Subdivision of the Drosophila mushroom bodies by enhancer-trap expression patterns. *Neuron.* **15**, 45-54.

Yang, T.T., Cheng, L. and Kain, S.R. (1996). Optimized codon usage and chromophore mutations provide enhanced sensitivity with the green fluorescent protein. *Nucleic Acids Res.* **24**. 4592-4593.

Yeh, E., Gustafson, K. and Boulianne, G.L. (1995). Green fluorescent protein as a vital marker and reporter of gene expression in Drosophila. *Proc Natl Acad Sci USA.* **92**. 7036-7040.

Yoshida, M., Yagi, T., Furuta, Y., Takayanagi, K., Kominami, R., Takeda, N., Tokunaga, T., Chiba, J., Ikawa, Y. and Aizawa, S. (1995). A new strategy of gene trapping in ES cells using 3'RACE. *Transgenic Res.* **4**. 277-287.

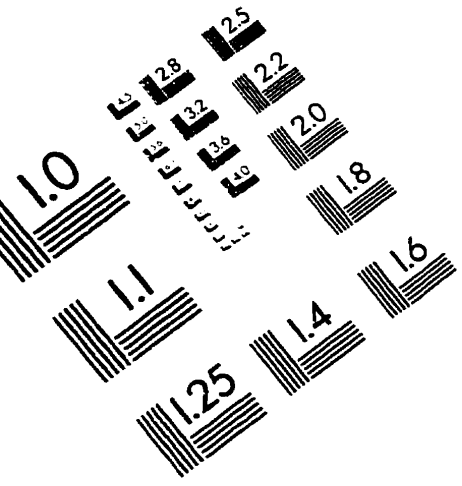
Young, J.M. and Hope, I.A. (1993). Molecular markers of differentiation in Caenorhabditis elegans obtained by promoter trapping. *Dev Dyn.* **196**. 124-132.

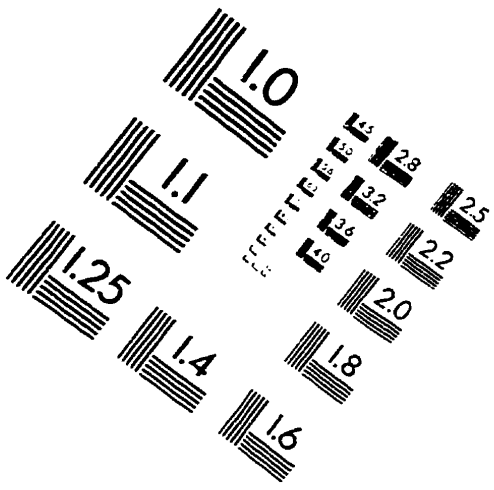
Zernicka-Goetz, M., Pines, J., McLean Hunter, S., Dixon, J.P.C., Siemering, K.R., Haseloff, J. and Evans, M.J. (1997). Following cell fate in the living mouse embryo. *Development.* **124**. 1133-1137.

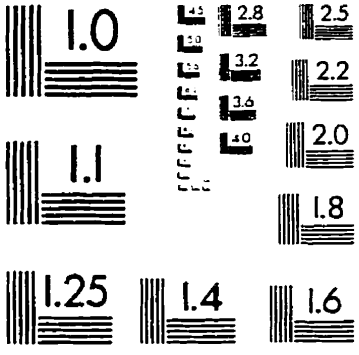
Zhang, G., Gurtu, V. and Kain, S.R. (1996). An enhanced green fluorescent protein allows sensitive detection of gene transfer in mammalian cells. *Biochem Biophys Res Commun.* **227**. 707-711.

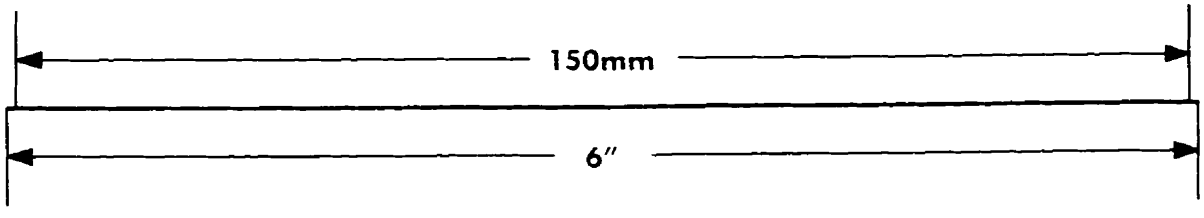
Zhou, X., Sasaki, H., Lowe, L., Hogan, B.L.M. and Kuehn, M.R. (1993). *Nodal* is a novel TGF-ß-like gene expressed in the mouse node during gastrulation. *Nature.* **361**. 543-547.

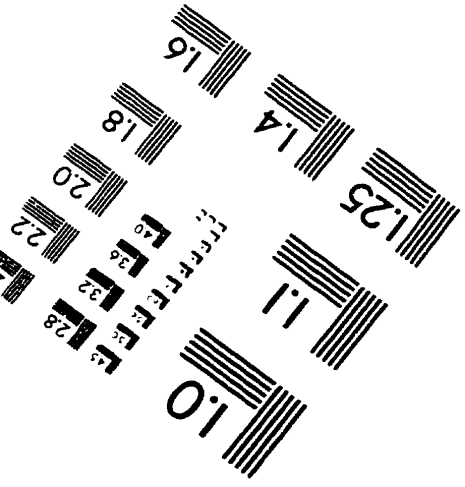
IMAGE EVALUATION
TEST TARGET (QA-3)

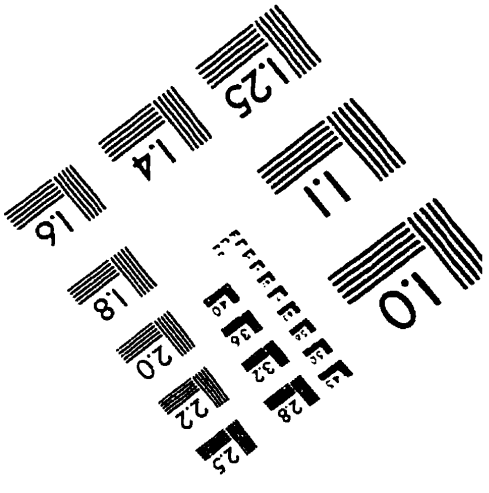












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989