CLONING AND ANALYSIS OF MOUSE CHROMOSOMAL LOCI SPECIFICALLY ACTIVE IN EMBRYONAL CARCINOMA STEM CELLS

KRISHNA MOORTHI BHAT, S., M.Sc.

CLONING AND ANALYSIS OF MOUSE CHROMOSOMAL LOCI SPECIFICALLY ACTIVE IN EMBRYONAL CARCINOMA STEM CELLS

by

© Krishna Moorthi Bhat, S., M.Sc.

Faculty of Medicine

Memorial University of Newfoundland

July 1989

A thesis submitted to the School of Graduate Studies in partial fulfilment of the requirements for the degree of Doctor of Philosophy

St. John's Newfoundland

Whoever protects the Truth will be protected by the Truth.

From Bhagavadgita

This thesis is dedicated in memory of my late uncle

Sri Ramachandra Bhat, Sabbanakodi

ABSTRACT

Chromosomal loci that are specifically active in the mouse embryonal carcinoma stem cells were cloned by using a functional selection procedure. The pluripotent P19 embryonal carcinoma cells were transfected with an enhancer-trap plasmid containing an enhancerless, inactive neomycin resistance gene and NEO^+ transformant cell lines were isolated. When the cells were induced to differentiate, most of the cell lines continued to express the neomycin resistance gene, however, in some cell lines, the neomycin resistance gene became repressed. From the later group of cell lines, eight in total, the integrated transgene plus the flanking cellular DNA sequences were cloned. Three of the cloned fragments from the above eight cell lines possessed a high NEO^+-transforming enhancer activity in the undifferentiated P19 cells. Among these three, two were inactive in differentiated P19 cells and NIH 3T3 cells, while the remaining one was active in both these differentiated cell types. Further analysis of these stem cell specific enhancers revealed that they were derived from the stem-cell specific Early Transposon-like genes.

In order to search for the presence of genes in the above stem cell specific loci, a P19 genomic library was constructed and the preinsertion regions at the neomycin resistance gene-integration sites were cloned from these cell lines. The cloned DNA was analyzed for the presence of genes by Northern blotting analysis.

Messages were detected in the Northern blots against some of the loci, however, their identity as functional genes is yet to be established.

During the course of this investigation, I observed the presence of Early Transposon-like genes in three of the above loci. Restriction mapping of the preinsertion loci and the Southern blot analysis of the DNA from mouse testis, parent P19 cells, and the three NEO^+ cell lines with the locus-specific probes, provided direct evidence that the transposon was inserted into these loci during the experimental time-frame and therefore was movable in the mouse genome. Analysis of the cell extracts from the three embryonal carcinoma cell lines, P19, F9, and PCC3 with a transposon-specific probe detected extrachromosomal copies of this transposon only in the P19 cells. Southern blot analysis of the DNA from mouse germ cell and various somatic cell lineages with the ends-specific transposon probes indicated that there were no apparent differences in the transposon integration sites between the germ line and the soma, suggesting that transposition of these ETn-like genes is strictly stem cell specific and ceases to occur before allocation of founder cells to the germ cell lineage and somatic lineages during mouse embryogenesis. These results demonstrate that the early transposon-like genes can act as a powerful insertion mutagen in the founder cells of the mouse embryo.

ACKNOWLEDGEMENTS

I wish to express my profound gratitude to Dr. H. Banfield Younghusband for his supervision, unfailing help, and continued support. I also acknowledge help from Dr. H. Hamada during my study. I am indebted to Dr. Chris Triggle and Dr. Chet Michalski for their support. Thanks are also due to Dr. Peter Howard for his valuable comments on the work and for improving the writing of this thesis, and Prof. M. Muramatsu of the University of Tokyo for allowing me to do part of the work described in this thesis in his laboratory. Mr. Fred Swift's assistance is gratefully acknowledged.

I am indebted to Mrs. Bernadette Power for her very valuable help and sincere concern for my well being, and Mrs. Hamada for her encouragement during my program. The company and help of my friends, Dr. M. Sakai, Kondo san, Hashimoto san and his wife Kahori san, Hanada san, Okuda san, Inoue san, Imakado san, Nishimura san, Fujiko san, Moriyama san, and Mariko san, at the University of Tokyo Faculty of Medicine, is very much appreciated.

I am grateful to Trudy for her help. I also acknowledge the support of my friends Mark and Linda from Vancouver through their regular letters, despite my very irregular replies. Special thanks are extended to my friends, John from Toronto, and Prof. Patrick Barren from Scotland, without their help, it would have been very difficult for me in Japan. My thanks are also due to Hashim, Rakesh, Sean, Swamy

and Michelle. My sincere thanks are due to Kim for her careful typing of this thesis in a short time.

The blessings of my parents, my brother and sister-in-law, and my three sisters and brothers-in-law were the greatest encouragement during this study.

I also gratefully acknowledge the award of a Memorial University of Newfoundland graduate fellowship and a pre-doctoral fellowship from the Cancer Society of Canada. Without the financial supplement received from the Faculty of Medicine, Memorial University of Newfoundland, during my stay in Japan, it would not have been possible for me to complete my program. The apartment provided by the Chugai Pharmaceutical Company during my stay in Japan is gratefully acknowledged.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

TABLE OF ABBREVIATIONS

bp	=	base pair(s)
C	=	degrees Celsius
Cat	=	chloramphenicol acetyl transferase
D-	=	undifferentiated EC stem cells
D+	=	differentiated EC cells
dCTP	=	deoxycytidine triphosphate
DMSO	=	dimethylsulfoxide
EC cells	=	embryonal carcinoma cells
EDTA	=	ethylenediaminetetraacetic acid
ETn	=	early transposon-like element
G418	=	genetecin 418, neomycin analogue
h	=	hour(s)
ICM	=	inner cell mass
IAP	=	intracisternal A-particle ge
kb	=	kilo base pairs
LBM	=	lauria-bertani medium containing 10% MgSO4
LBMM	=	LBM containing 0.2% maltose
MEM	=	minimum eagle's medium
min	=	minute(s)

MOI	=	multiplicity of infection
MW	=	molecular weight
N	=	normal
NEO$^+$	=	EC clonal cell line expressing the stably integrated Neo-R transgene
Neo-R	=	neomycin resistance gene
ng	=	nanogram(s)
OD	=	optical density
PBS	=	phosphate buffered saline
RA	=	retinoic acid
RSB	=	reticulocyte standard buffer
Rx	=	reaction number
sec	=	second(s)
SDS	=	sodium dodecyl sulphate
SSC	=	sodium chloride- sodium citrate buffer
SV 40	=	simian virus type 40
TE	=	tris-EDTA buffer
UV	=	ultraviolet

CHAPTER I

GENERAL INTRODUCTION

In this thesis, data are presented for the existence of chromosomal domains that are specifically active in the mouse embryonal carcinoma stem cells but not in their differentiated derivatives. A functional selection procedure has been used to clone such stem cell specific loci. I also present evidence that the stem cell specific early transposon-like genes are movable in the mouse genome.

A fascinating problem attracting considerable attention from biologists is that of the programming of development of an animal from a single cell to an adult. One of the basic concepts of embryology is that ordered and precise changes in the patterns of expression of genes, directly or indirectly, control the process of development (see Davidson, 1976). *Caenorhabditis elegans*, sea urchin, *Xenopus* and *Drosophila* have been utilised extensively in such developmental studies. In the past five years, considerable attention has also been given to understand the molecular aspects of development, such as the patterns of expression of genes during development, in mouse, a more complex mammalian organism. The experiments described in this thesis are relevent to this basic question of gene expression during early mouse embryogenesis. In view of this, the events and some of the important aspects of embryogenesis of the mouse are briefly outlined here.

I.1) Early mouse development

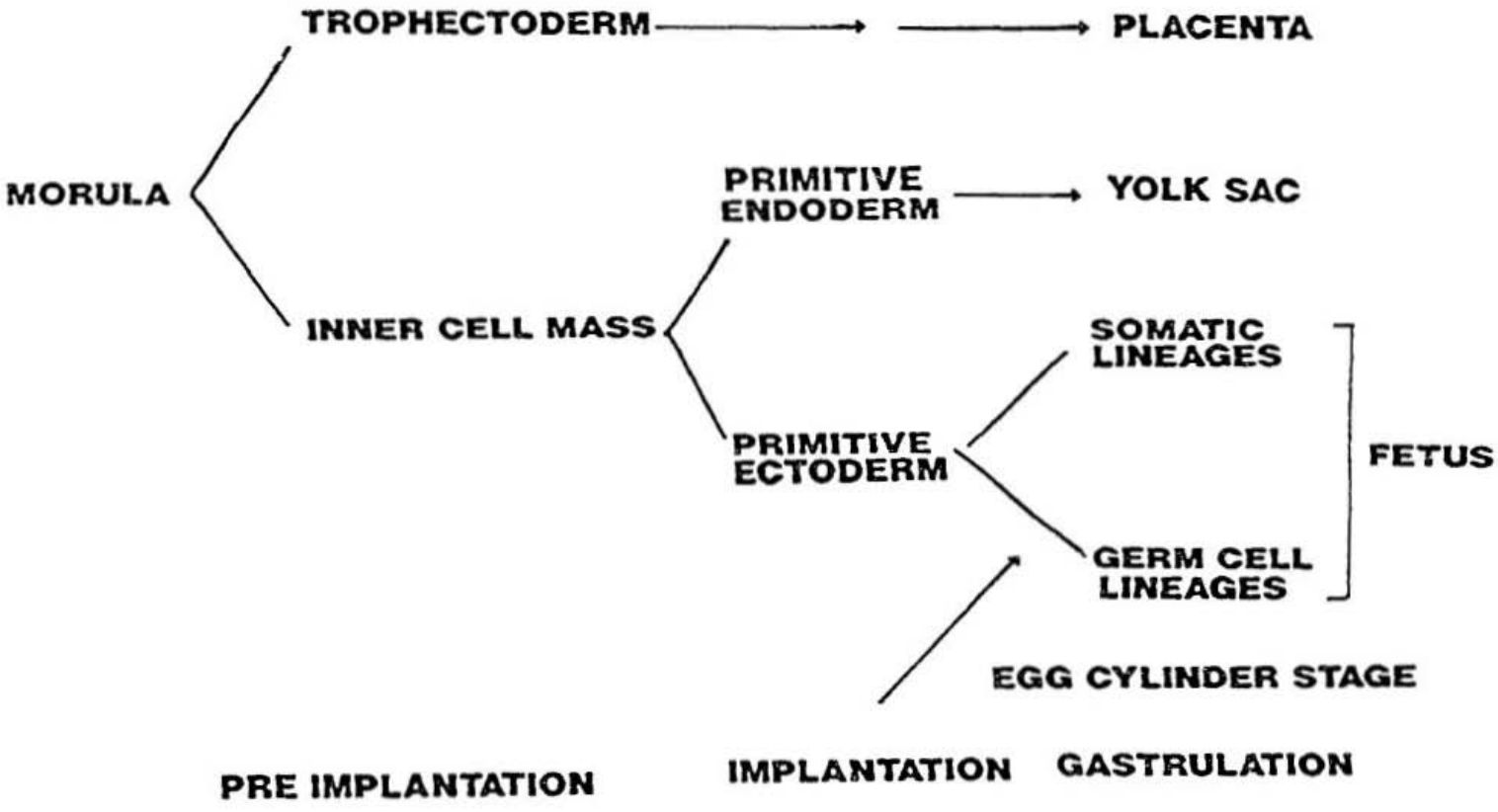
Early development of the mouse begins with the fusion of a sperm with the sperm receptor of an egg (see Wasserman, 1987) and the internalization of the sperm. Twenty hours after fertilization, pronuclear fusion takes place, and the zygote, surrounded by a transparent coat called the zona pellucida, begins its first mitosis (see Solter, 1987).

Figure I-1 depicts the simplified cell lineage relationships in early mouse development compiled from Gardner (1978) and Soriano and Jaenisch (1986). The zygote undergoes the first cell division within 24 h of fertilization to give rise to the two-cell embryo. As development proceeds, the interval between the successive cell divisions gets shorter (unlike in *Xenopus* and *Drosophila*, where the cell cycle slows down) and about 55 h after fertilization, the embryo, which now has 8 cells or "blastomeres", takes the form of a mulberry-shaped cluster. This 8 to 50 cell stage is called the morula. Between the 8 and 16 cell stages, the blastomeres change their cohesiveness and become compacted together with tight junctions forming between the outer cells. This in effect seals off the interior of the morula from the external medium. The surface of the morula becomes smoother and more nearly spherical. Soon afterwards, the internal intercellular spaces enlarge to create a fluid-filled central cavity - the blastocoel-, and the embryo becomes what is called the blastocyst (the 64 to 128 cell stage). This takes place 3.5 days post-coitum.

Figure I-1. Simplified cell lineage relations in early mouse development.

This figure is compiled from Gardner (1978) and Soriano and Jaenisch (1986). For simplicity, the contribution of cells from the primitive ectoderm lineage to extraembryonic structures namely yolk sac, placenta and allantois is not shown in the figure.

EARLY DEVELOPMENT IN MOUSE



The formation of the blastocyst represents the first overt differentiation of an embryo and results in the formation of an outer cell layer called the trophectoderm and the inner cell mass (ICM), which is an accumulation of cells at one pole on the inside of the trophectoderm. As development proceeds, the trophectoderm contributes to the placenta and the components of the extra-embryonic structures. These are required for implantation of the blastocyst and in the establishment of the fetal relationship with the mother in utero. The pluripotent cells of the ICM, on the other hand, contribute to the entire embryo proper.

At about 4.5 days post fertilization, the blastocyst is ready for implantation in the uterus. Before implantation can proceed, the protective zona pellucida of the blastocyst must be shed. By this time, the trophectoderm has differentiated into two cell lineages: a) the cyto-trophoblast, which goes on to encircle the cells of the ICM, and b) the syncytio-trophoblast, a group of large multinucleated cells which subsequently invade the lining of the uterus, in effect implanting the blastocyst into the uterine wall. Around this time, the ICM grows and gives rise to two further cell types; the primary endoderm, which is a differentiated and extraembryonic cell lineage that contributes to the extraembryonic structures, and the pluripotent primary ectoderm, which develops to form the entire fetus (this description of the preimplantation stages of mouse embryogenesis was compiled

from Gardner and Papaioannou, 1975; Johnson *et al.* 1977; Gardner, 1978; Martin, 1980; Soriano and Jaenisch,1986; and Darnell *et al.* 1986).

I.1.i) The allocation of the germ line

The next important aspect of embryogenesis is the allocation of cells to the germ cell lineage. The germ cell lineage is known to be derived from the primitive ectoderm (Gardner and Rossant, 1979). There is some confusion however, as to when the determination of cells of the primitive ectoderm to germ cell lineage occurs. Snow and Monk (1983) suggested that, up to the time of gastrulation, the cells of the primitive ectoderm are bipotential and may become either somatic or germ cells depending upon what position they occupy in the egg cylinder. Soriano and Jaenisch (1986), however, concluded that primordial germ cells are determined early, before determination of somatic cell lineages has occured. They infected 4-16 cell preimplantation stage mouse embryos with M-MuLV (a retrovirus), and analysed the correlation between the genetic transmission of copies of the provirus by the founder animal to the offspring and their presence in the somatic cells. A positive correlation should be observed if cells are the allocated to germ line late in the early-development (Soriano and Jaenisch, 1986). The results indicated however that there was no correlation between the genetic transmission of the proviruses and their presence in the somatic tissues in 17/38 proviruses. Ten

proviruses present in the soma of founder animals were not transmitted to the offspring, and seven proviruses which were not detected in the somatic tissues of the founder were transmitted to offspring. The failure of embryonal carcinoma stem cells to colonize the germ cells in chimaeric mice, discussed in section I.3.ii of this thesis, also supports the conclusion that primordial germ cells are determined early in embryogenesis, before the soma.

The final body plan of the fetus is established at gastrulation, when a few embryo founder cells proliferate to form the egg cylinder. These cells rapidly divide during the next 24 h of primitive streak development (Snow, 1977) followed by cell movements to generate the three primordial germ layers: the ectoderm, the mesoderm, and the endoderm. Between the 5th and the 10th day after fertilization the basic body plan of the mouse is established from the three germ layers and the interactions between them (see Gardner and Papaioannou, 1975; Darnell *et al.* 1986, for further details on the mouse postimplantation development). Depending upon the strain, the mouse is born after 19-20 days of gestation (Hogan *et al.* 1986).

The work described in this thesis is relevent to the differentiation of the ICM and/or embryonic ectodermal cells in mouse early development, since the EC cell line used in this study closely resembles the cells of these stages, as described in section I.3.

I.2) Some molecular aspects of early embryogenesis in the mouse

Embryonic development is generally viewed as a two step process; determination of precursor cells to specific cell lineages followed by their differentiation. "A cell is said to be determined once it has been instructed to - or once it has somehow decided for itself - to become a specific cell type at some future point" (Darnell *et al.* 1986). The work by Groudine and Weintraub (1982) provides some insight into the process of determination. These researchers reported that once induced, globin gene DNAase I-hypersensitive sites can be propagated to daughter cells in the absence of the original inducing factors. The process of determination was aptly summed up by Groudine and Weintraub (1982), who wrote, "certain specific determinative events are induced in precursor cells at one time in development and independent of the concurrent action of original inducing influence, the effects of these events are expressed in progeny cells that begin overt differentiation some time later".

There are many unanswered questions concerning the molecular mechanisms which underly the process of determination. Theoretically at least, one of the following phenomena could mediate the event of determination:

i) Gene specific DNA rearrangements. Examples include the intensively studied immunoglobulin and T-cell receptor gene rearrangements (Hozumi and Tonegawa,

1976), although similar gene specific rearrangements that might regulate development have not been demonstrated in eukaryotes.

ii) Chemical modifications such as DNA methylation. For example, the embryonic mouse fibroblast cell line C3H10T1/2 gives rise to myogenic, adipogenic, and chondrogenic clones following brief treatment with the hypomethylating agent 5-azacytidine (Taylor and Jones, 1979); DNA from 5-azacytidine-derived myogenic 10T1/2 clones convert normal 10T1/2 cells to myoblasts, while the DNA from normal 10T1/2 cells do not (Konieczny *et al.* 1986; Lassar *et al.* 1986). This would imply that upon hypomethylation, genes involved in the determination pathway are activated and this state is propogated to daughter cells, or to those cells transfected with the hypomethylated DNA. In fact, genes involved in myogenic lineage determination and differentiation have been isolated (MyoD1 gene, Davis *et al.* 1987; myd gene, Pinney *et al.* 1988) by functional selection procedures.

iii) Gene amplification, as seen in the case of the bidirectional gene amplification of the chorion gene locus in *Drosophila* (Spradling and Mahowald, 1980) resulted in the production of large amounts of eggshell protein which are required.

iv) Gene deletion. In the nematode *Ascaris megalocephala*, certain early determinative events discriminate between putative germ cell and somatic cell

progenitors and a portion of the genome is lost only in the somatic cell progenitors (see Browder, 1984).

v) Specific changes in chromatin conformation, such as the locus activation domains of the β-globin gene locus (Grosveld *et al.* 1987). This could thereby activate or suppress sets of genes involved in the determination pathway.

vi) Other determinative mechanisms, such as (a) maternal effects, (b) protein modification, and (c) translational control:

a) Maternal effect genes are those for which absence from the maternal parent of an embryo causes dramatic effects on the final body pattern of the fly *Drosophila*. The activity of this class of genes offers perhaps one of the most straightforward and simple ways of cell determination during embryogenesis. A maternal effect gene is transcribed during oogenesis and either the mRNA or the translated protein product of the message is deposited in the egg. In *Drosophila*, the establishment of the dorsal-ventral and the posterior-anterior body axes of the embryo are already determined in the egg by the expression of sets of maternal effect genes (Ingham, 1988; Anderson, 1987). Maternal determination of embryogenesis is also seen in amphibians, sea urchin and in a variety of other organisms (Davidson, 1976). A similar type of extensive spatial and temporal organization of development of

an embryo by maternal determinants, however, does not appear to take place in mammals. A maternal inheritance of one or a few regulatory proteins, however, which might be involved in the initial activation of the embryonic genome, cannot be ruled out. For instance, circumstantial evidence to support this view comes from the work of Levey *et al.* (1978). They detected a significant amount of mRNA from the maternal source (see section I.2.i of this thesis) and it is possible that the products of these messages could be activators of the embryonic genome. The need for an early activation of the embryonic genome for development to proceed in mammals (see section I.2.i), suggests that most embryonic determination and differentiation in mammals is under the control of the embryonic genome itself. Self reliance in the development of mammals has perhaps been selected during evolution to avoid the greater risk of error which exists in the maternal effect pathway.

b) Protein modification could also have an important role in the process of cell determination. In the process of cell determination, at least two types of protein modifications could take place:

(1) Protein phosphorylation. For instance, the yeast heat shock gene promoter binding factor, HSF, is present in cells irrespective of whether the cells are heat shocked or not. However, the HSF protein is believed to be in an inactive form

when the cells are not heat shocked (Sorger *et al.* 1987; Sorger and Pelham, 1988). Sorger and Peham (1988) reported that HSF protein becomes phosphorylated at elevated temperatures. As a result, the protein appears to be converted to an active form with the consequence of enhanced heat shock gene transcription (Sorger and Pelham, 1988).

(2) Proteolytic processing of proteins. Conversion of proteins from the inactive to the active form by proteolytic cleavage is an important regulatory mechanism in the activation of some enzymes. It has been suggested that proteolytic modification of certain developmentally important proteins occurs in *Drosophila* (DeLotto and Spierer, 1986; Chasan and Anderson, 1989). For instance, dorsal-ventral pattern formation in *Drosophila* appears to require a cascade of proteolytic cleavages (DeLotto and Spierer, 1986). This is exemplified by the maternal effect genes, *Easter*, and *Toll*. *Easter* protein is a serine protease (Chasan and Anderson, 1989) involved in dorsal-ventral formation during *Drosophila* embryogenesis. The *Toll* protein on the other hand, specifies the polarity of the embryonic dorsal-ventral pattern. It has been suggested that the Easter protein may proteolytically activate the Toll protein and affect the distribution of Toll protein in the embryo during dorsal-ventral pattern formation (Chasan and Anderson, 1989).

c) Translational control. The control of determination and differentiation during embryogenesis can also occur at the translational level. For example, it has been reported that in the sea clam, oocyte mRNAs are translated only after fertilization (Darnell *et al.* 1986).

The next prominent step in the developmental history of a cell after determination is the decision to enter a differentiation pathway, committment, as it is called. For instance, when P19S18, an embryonal carcinoma cell line (McBurney *et al.* 1982), is treated with DMSO for a period of 40 h, the cells become committed to differentiate into muscle cells regardless of the presence or absence of the inducing agent (Edwards *et al.* 1983). While committment is a non-visible event, differentiation is an overt phenomenon, where a determined cell acquires its specialised phenotype (Darnell *et al.* 1986). Sets of cell type specific genes are activated and the phenotype and functional properties are maintained thereafter and propagated to the progeny cells. In principle, determination, committment, and differentiation may simply represent the overlapping activities of a series of different genes in time and space, to produce a continuous set of events (Darnell *et al.* 1986).

These terms and explanations are important because they are central to our understanding of the molecular aspects of embryogenesis, some of which are discussed below.

I.2.i) Loss of totipotency and early determinative events during mammalian embryogenesis

Embryogenesis is an unique process. In *Xenopus* (an amphibian) for instance, the celebrated nuclear transplantation experiments performed by John Gurdon (Gurdon, 1976) demonstrated that even nuclei from adult cells support complete and normal development, although the ability to promote normal development of an enucleated zygote decreases as the age of the donor nuclei increases. In the mouse, however, the nuclear transplantation experiments performed by McGrath and Solter (1984a), Surani *et al.* (1986) and Robl *et al.* (1986) showed that even nuclei from the four-cell stage, when transplanted into an enucleated zygote, fail to support development. The early determinative events in the nucleus therefore seem to have already taken place at this stage, with the consequence of loss of totipotency (ability to form an entire organism, both embryonic and extraembryonic; while pluripotency refers to the ability of cells to form a limited number of lineages) and committment of embryonic cells to go forward in the developmental pathway.

Attempts to explain this loss of totipotency have included important experiments performed by McGrath and Solter (1986). They introduced nuclei from 8-cell mouse embryos into oocytes which had been activated (i.e., initiated development by artificial stimuli such as pricking with a needle) 3 h earlier. The

female pronucleus was removed a few hours later. Control experiments were also done in which the oocytes received pronuclei from the zygote. In the control group, over 70 % developed to the blastocyst stage while in the experimental embryo, 3% reached morula-blastula stage. Although the success rate was low, this result suggests that given sufficient time, the determinative events of early stages can be reinitiated by an latet stage nucleus. In other words, the loss of totipotency in mammals is not permanent. It should be noted however that this reversibility in mouse is different from that of Xenopus. In the mouse, the reinitiation of development by a later stage nuclei or the reversibility of early determinative events is an artificial situation, whereas, in *Xenopus*, there is in fact no loss of totipotency of the nuclei. The above results also rule out the possibility that irreversible, permanent modifications of the genome such as rearrangement, deletions etc. form the basis for the early determinative events and the loss of totipotency.

Further work done on mouse development by Robl *et al.* (1986) and Solter (1986) gives some explanation as to why the nuclei from the later stages fail to reinitiate developmental programs. They enucleated 2-cell blastomeres and fused them with 8-cell stage or ICM nuclei. In the Robl *et al.* (1986) study, where 2-cell enucleated embryos were fused with 8-cell nuclei, 51% formed blastocysts, 42% implanted normally, and out of 11 examined, 2 contained normally developing embryos. In Solter's experiment, over 50% developed to the blastocyst stage *in vitro*.

The above results of McGrath and Solter (1986), Solter (1986), and Robl et al., (1986) indicate that the loss of totipotency in mammals reflects a requirement for a very precise temporal interactions between the embryonic genome and the cytoplasm. Furthermore, the results of Robl et al., (1986) and Solter (1986) indicate that in the mouse, there exists a mechanism such as the expression of genes specific to the 2-cell stage (Solter, 1987) or 2-cell stage specific modifications of crucial proteins (such as transacting regulatory proteins), which are essential to support further embryonic development.

Of relevence here is the question of when the embryonic genome begins its activity. Several attempts have been made to answer this question. Most of the experiments were performed by measuring the incorporation of [^{3}H]-Uridine, by cell free translation of RNA, by monitoring the expression of paternal isozyme variants, or by using transcriptional inhibitors such as alpha-amanitin. By analysing the incorporation of [^{3}H]-Uridine, Mintz (1964) showed that RNA synthesis occurs in mouse embryos at the 2-cell stage. This observation is supported by studies using α-amanitin, which blocks the cleavage of 2-cell embryos (Braude *et al.* 1979). Furthermore, Levey *et al.* (1978), using affinity chromatography of [^{3}H]Uridine-labelled embryonic RNA on oligo(dT)-cellulose, detected newly synthesized polyadenylated RNA in 2-cell mouse embryos, although a significant amount of maternal mRNA was also present. Also, Wudl and Chapman's (1976) study indicates that the paternal allele for β-glucoronidase is synthesized in the 4-cell

stage. These studies therefore strongly suggest that zygotic transcription in the mouse begins at the early phase of the 2-cell stage. In addition to mRNA synthesis, there is an indication that rRNA synthesis occurs in 2-cell mouse embryos (Knowland and Graham, 1972; Hilman and Tasca, 1969).

From the above observations, it is possible to discern that there is a correlation between the loss of totipotency or the failure to reprogram development by later stage nuclei, and the onset of transcriptional activity of the embryonic genome. This indicates that in the mouse the embryonic genome most likely intervenes in development as early as the 2-cell stage. This is consistent with the idea that the transcriptional activity of the embryonic genome is part of 2-cell stage specific determinative events that are required for further development of the embryo.

Further support for the relationship between the onset of transcriptional activity of the embryonic genome and the loss of totipotency is provided by the work of Calarco and McLaren (1976), Crosby *et al.* (1988) and Willadsen (1987) with sheep embryos. In sheep, embryonic transcriptional activity begins only at the 16-cell stage (Crosby *et al.* 1988). Ultrastructural changes in nucleoli, a measure of active rRNA synthesis, are observed in 16-cell sheep embryos (Calarco and McLaren, 1976) (as opposed to 2-cell mouse embryos). Consistent with this observation, Willadsen (1987) obtained results which show that in sheep, nuclei from eight or 16 cell stage embryos, when transplanted to enucleated eggs, were capable of complete

development. It is reasonable therefore to speculate that, in sheep, since the loss of totipotency occurs later than in the mouse, the initial determinative events also occur later.

In terms of transcriptional activity of the embryonic genome, the early determinative events and the loss of totipotency in mammals can be explained as follows; in the later stage donor nuclei transplanted to an enucleated zygote, transcriptional factors are bound to their responder genes in stable transcriptional complexes and an active gene expression program of a later stage is going on. In contrast, those genes that are required for the earlier stages of development are in an inactive conformation in the later stages of development, and are thus unavailable for the effector transcriptional factors from the acceptor cytoplasm when the genome is transplanted to an enucleated zygote. This would result in the failure of the reprogramming of the donor nuclei required for the reinitiation of development.

In general, therefore, the loss of totipotency in mammals, which accompanies the onset of embryonic transcriptional activity, cell determination, commitment, and differentiation, indicates that compatibility in interactions between the nucleus and cytoplasm and among the cells of the embryo is absolutely required for normal embryogenesis to occur. These are stage specific interactions, and they form the basis for early determinative events during embryogenesis. The experiments

described in this thesis were designed to identify and isolate the gene(s) involved in these interactions.

I.2.ii) Oncogenes in Development

The idea that oncogenes, which cause or contribute to cancer, are genes that regulate normal growth, embryogenesis and cell differentiation is of recent origin (Verma and co-workers, 1982, 1983; see the review by Adamson, 1987). There are several lines of evidence, both circumstantial and direct, to support this view, as discussed below.

A) The first line of evidence comes from the properties of protooncogene products:

(1) some protooncogene products are nuclear in localization and have DNA binding properties; examples include Myc, Myb, and Fos, (Adamson, 1987). Furthermore, the Fos protein has been shown to be a transcriptional regulator (Distel *et al.* 1987).

(2) some protooncoproteins are protein kinases (examples include c-Fms, c-Src, c-Abl, c-Mos). Protein kinases are generally believed to play a role in developmental decisions because they can either directly regulate genes (e.g., the catalytic region of cAMP-dependent protein kinase transactivates genes containing cAMP-responsive enhancers, Riabowol *et al.* 1988), or modify existing proteins.

(3) a few of the protooncogenes code for growth factors and growth factor receptors (such as c-erb-b, c-fms, see the review by Adamson, 1987).

Proteins possessing the above properties are generally thought to be important components of the circuitry of developmental processes.

B) The second line of evidence relating oncogenes to embryogenesis comes from the well defined spatial and temporal patterns of expression of protooncogenes during embryogenesis and cell differentiation. Mouse embryos showed persistant expression of c-Ha-ras, c-Ki-ras, c-fms, c-myc, c-fos, and c-sis (Muller *et al.* 1983; Slamon and Cline, 1984). N-Myc and p53 are expressed in the embryo proper at high levels from implantation to the mid-gestation period (Jacobovits *et al.* 1985; Rogel *et al.* 1985), c-fos and c-fms are highly expressed in the extra-embryonic membranes (Muller *et al.* 1983; Muller *et al.* 1982), int-1 during mid-gestation period and int-2 in the preimplantation stages (Jacobovits *et al.* 1986). The above results suggest that these protooncogenes play an important role during embryogenesis.

C) The third line of evidence comes from the fact that abnormal expression of the protooncogenes during embryogenesis leads to malformation of tissues or organs. For example, the protooncogene fos is normally expressed in adult bone tissues, however, deregulated c-fos expression (qualitatively normal, but high level expression) during mouse embryogenesis leads to malformation of bone tissue (Ulrich *et al.* 1986). There is some criticism however that interference of normal

embryogenesis by deregulated expression of protooncogenes may not indicate that they are principal embryogenesis genes, as this type of experiment often fails to distinguish direct from indirect effects.

D) The fourth line of evidence comes from similarity studies. That is, some of the protooncogenes and developmental genes have similarity at the nucleic acid as well as amino acid sequence levels. For instance, Int-1, a mouse mammary oncogene product (Nusse and Varmus, 1982) is similar to the *Drosophila* developmental gene, Wingless (Rijsewijk *et al.* 1987). Another oncogene product, c-Rel is 50% similar to dorsal protein. Dorsal gene is an embryonic polarity gene of *Drosophila* (Steward, 1987).

E) There is further direct evidence to support the oncogene-embryogenesis connection. In *Drosophila* for example, "lethal giant larvae" gene mutation has been isolated. The normal allele controls growth and differentiation of cells of the optic centres. Heterozygous flies are normal but homozygous mutants develop invasive, lethal neuroblastic tumors in the putative optic centres (Gardner *et al.* 1982). Another very interesting example is a heritable renal carcinoma of rats (Knudson, 1986). Here the heterozygotes show oncogenesis, but the homozygotes die at the embryonic stage (Eker *et al.* 1981). In mouse, mutation at the white locus leads to developmental abnormalities such as sterility, macrocytic anemia, and lack of hair

pigmentation. Chabot et al (1988) showed that the white locus gene is the protooncogene, c-Kit, which encodes a transmembrane tyrosine kinase receptor.

In light of the evidence described above, it is reasonable to assume that developmental genes include at least some of the oncogenes and that their correct expression (quantitative, temporal and spatial) is necessary for normal development. In fact, the chance is very high that the type of regulatory genes we hoped to clone i.e., active only in undifferentiated, malignant embryonal carcinoma (EC) stem cells but not in their differentiated derivatives (which are nonmalignant), would be protooncogenes (see Introduction to Chapter III).

I.2.iii) Homeobox-containing genes in development

Homeobox-containing genes are control genes which are active in the establishment of the segmentation pattern and in the specification of segment identity (Gehring, 1985; Ouweneel, 1986). These were originally isolated and studied in *Drosophila* (McGinnis *et al.* 1984; Scott and Weiner, 1984). The homeobox of all homeobox-containing genes has the same open reading frame and codes for a highly basic domain of a protein, the homeodomain (Gehring, 1985). There is convincing evidence based on mutational analysis (Laughon and Scott, 1984) and NMR studies, that the homeodomain contains a DNA-binding helix-turn-helix motif. This suggests that proteins having a homeodomain would be good candidates for

transcriptional regulators (reviewed recently by Levine and Hoey, 1988). For instance, the *fushi tarazu* (ftz) gene product is a site-dependent transcriptional activator and in addition, regulates it own transcription (autoregulation; Hiromi and Gehring, 1987). Recently, Jaynes and O'Farell (1988) reported that the engrailed gene product, another homeodomain protein, counteracts the activation of *fushi tarazu* by competing for homeodomain binding sites. They concluded that homeodomain-containing proteins can bind to a common site and either activate or repress transcription. Their results therefore provide experimental support for the generally accepted notion that a homeodomain allows a protein to bind to DNA; however the specific effect exerted by the protein (activation or repression of other genes) is a function of the remaining part of the protein, possibly mediated by protein-protein interaction. With variations in binding affinities to the target DNA sequences being due to variability in the amino acids of the homeodomains among the proteins (Jaynes and O'Farell, 1988; Levine and Hoey, 1988), the parent proteins may control the activities of a range of overlapping genes in a fine-tuned circuitry and effectively organize embryogenesis.

Information on the developmental role of homeobox genes in mi e (as well as in amphibians) has lagged behind. It is generally speculated that they regulate development and differentiation by mechanisms similar to that of the homeotic genes in *Drosophila* (Gehring, 1987; Ingham, 1988). So far a number of mouse homeobox genes that share striking similarities with the homeotic genes of

Drosophila have been cloned and their expression patterns have been monitored (Hart *et al.* 1985; Gaunt *et al.* 1986; Joyner and Martin, 1987; Holland and Hogan 1988). Not surprisingly, these homeobox genes exhibit spatially and temporally specific patterns of expression in the mouse embryo and adult structures. Recently, Wolgemuth *et al.* (1989) showed for the first time that expression of a mammalian homeobox gene, Hox-1.4, affects a developmental process. This gene is expressed during the mid-gestation period of mouse development, but interestingly, its highest level of expression is restricted to developing male germ cells in the adult (Wolgemuth *et al.* 1987; Rubin *et al.* 1986; Wolgemuth *et al.* 1986). When the Hox-1.4 gene was overexpressed in the embryonic gut of transgenic animals, this resulted in the abnormal gut development known as megacolon, which was inheritable (Wolgemuth *et al.* 1989).

An important advance in the study of mammalian homeobox-containing genes comes from the recent finding that the lymphoid specific, octamer motif-binding nuclear factor NF2 contains a homeodomain (Ko et al., 1988). This finding reinforces the notion that in mammalians also the homeodomain-containing proteins are gene regulators. Thus the current emphasis on these genes, may soon lead to a better understanding of their role in mammalian embryogenesis.

I.2.iv) Parental imprinting and Embryogenesis

This particular topic has no direct relevence to the work described in this thesis. I include it, however, because of its contribution to our understanding of gene expression programs during embryogenesis.

Parthenogenesis, i.e., development of an egg without fertilization by a sperm, is not uncommon among teleosts or birds. It was originally thought that parthenogenesis could also occur in mammalians. However, by using nuclear transfer methods, McGrath and Solter (1984) and subsequently others (Barton *et al.* 1984; Surani, 1985; Surani *et al.* 1986) demonstrated that embryos containing two male or two female pronuclei always fail to develop normally beyond the blastocyst stage. Embryos containing two female pronuclei can implant and develop into small embryos, but will not fully develop extraembryonic structures and ultimately abort (Surani *et al.* 1986). Those embryos containing two male pronuclei have normal extraembryonic structures but poorly developed embryonic parts (Surani, 1985). These results lead to the conclusion that the maternal genome is important in the development of the embryo proper and the paternal genome is involved in the development of the extraembryonic membranes. This phenomenon has come to be known as gametic or parental imprinting.

Results obtained with genetic experiments (Searle and Beechy, 1978; Cuttanach and Kirk,1985; Cuttanach, 1986) indicate that portions of several

chromosomes are either active or inactive during development depending upon the parental origin. This means that firstly only one of the two parental alleles is active, and secondly that the level of expression of a given gene is dependent on parental origin (see also Solter, 1987). One compelling evidence to support the first possibility comes from the behaviour of the T-hp mutation of mouse (Hairpin tail, see McLaren, 1979). The effects of this nuclear defect (McGrath and Solter, 1984b) are determined by the sex of the parent from which it is inherited. When inherited from the female parent, it is lethal at the embryonic stage, whereas embryos which inherit the mutation from the male parent survive (Johnson, 1974). The second possibility is supported, as Solter (1987) pointed out, by the reciprocal influence of chromosome 11 on the growth of the embryo, in that, paternal disomy and maternal nullisomy (for chromosome 11) offspring are bigger and paternal nullisomy and maternal disomy newborns are smaller (Cuttanach and Kirk, 1985). Furthermore, in nuclear transplantation studies, those embryos containing two female pronuclei, and therefore having maternal disomy for chromosome 11, develop into small embryos (these genes can thus be called maternal genes).

These findings lead to the question of what mechanism could account for parental imprinting. There is an indication that differential methylation of specific regions of chromosomes may be the mechanism of parental imprinting (methylation is believed to play a role in gene transcription). The restriction enzymes *Hpa*II or *Hha*I are sensitive to methylation interference. Provided a particular gene or

a transgene has one of these enzyme sites, and there is a probe available for its detection by Southern blot analysis, one can evaluate the methylation status of the gene. A correlation can thus be established between the methylation state of a gene and its parental derivation. Using this strategy, several laboratories have reported that certain transgenes are hypomethylated when inherited from the father and hypermethylated when inherited from the mother (Reik *et al.* 1987; Sapienza *et al.* 1987; Swain *et al.* 1987). This can be called a type I locus. Sapienza *et al.* (1987) also found that in one of the transgenic lines, the reverse is observed, i.e., sequences become more methylated after transmission through the male germ line, a type II locus. Based on the frequency of integration of transgenes into these two types of loci in several experiments, it is likely that there are many fewer type II loci than type I loci in the mouse genome. Interestingly, Swain *et al.* (1987) found that their transgene (Myc) was expressed only when it was paternally inherited and only in the heart, although the gene was in the hypomethylated form in all tissues. Thus there appear to be several levels of regulation of these genes, one of which appears to be methylation.

I.3) Embryonal Carcinoma cells and embryogenesis

The study described in this thesis employed an embryonal carcinoma cell line. In view of this, I have described in the following some of the important aspects of EC cells and why these cells are useful in embryogenesis studies.

When an early mouse embryo is transplanted to extrauterine sites such as the kidney or testis of an adult, these embryos become disorganised and form a tumor (Damjanov *et al.* 1987; Martin, 1980) (this result possibly supports the contention that oncogenesis and embryogenesis are linked). The proliferating, malignant, undifferentiated stem cells of these tumors (teratocarcinomas) are known as embryonal carcinoma cells. EC cell-tumors can arise naturally either when oocytes undergo parthenogenetic development in situ or when the primordial germ cells abnormally proliferate in the testis (Stevens, 1975). In recent years, increasing numbers of investigators are using EC cells to study embryonic development. There are several reasons for this, some of which are described below:

a) Pluripotency: an EC tumor consists of a wide spectrum of cell types such as teeth, fingers and hair (Martin, 1980) which have differentiated from the proliferating EC cells. In other words, EC cells are pluripotent and can differentiate into many tissue types.

b) Culture *in vitro*: EC cells can be cultured *in vitro* without loss of their pluripotency, or their normal chromosome constitution (see Martin, 1980).

c) Differentiation *in vitro*: EC cells can be made to differentiate *in vitro* into a wide variety of tissue types, such as muscle cells, neurons, endoderm-like cells, parietal endoderm, visceral endoderm, etc. using various drugs such as DMSO, RA, or RA plus cAMP (see McBurney *et al.* 1982; Edwards *et al.* 1983; Martin, 1980; Strickland and Mahdavi, 1978; Strickland *et al.* 1980). This ability to respond to chemical agents and differentiate into a particular lineage *in vitro* can be used to study cellular determination and differentiation and to clone genes which are important in development.

d) Formation of chimaeras: in addition to the above embryonic properties, Brinster (1974) demonstated for the first time that EC cells microinjected into blastocysts contributed to the formation of chimaeric mice. In fact, EC cells can contribute to every tissue except germ cells in the chimeric mouse (see Martin, 1980). Of relevence to explain the lack of germ cell colonization by EC cells is the recent finding by Soriano and Jaenisch (1986). They concluded that the germ cells are set aside before the soma. It is likely that EC cells are equivalent to those embryonic cells which are committed to somatic lineages and therefore represent later cells in the developmental heirarchy (see Figure I-1). While the lack of germ line colonization is a limitation preventing the use of EC cells to create germ line mutations, their pluripotential capacity to differentiate *in vitro* has been widely

utilised to understand the molecular mechanisms of embryogenesis and cell differentiation.

I.3.i) Equivalence of EC stem cells to embryonic cell types

When EC stem cells are used to study the events of development, an important question that needs to be addressed is the equivalence of EC cells to various embryonic cell types. This question can be answered in the following way. EC cell lines such as F9 monolayer cells, when treated with RA and dibutyryl cAMP, differentiate into parietal endoderm (Strickland and Mahdavi, 1978). When small aggregates of cells are treated with RA alone, they differentiate into visceral endoderm (Hogan *et al.* 1981). It is believed that both parietal and visceral endoderm are derived from the common bipotential precursor, primary endoderm (see Figure I-1) (see Gardner, 1978). Strickland and Mahdavi (1978) however, showed that addition of a low concentration of RA to F9 monolayers induces them to differentiate into primary endoderm-like cells. If we take these three sets of results, it is likely that F9 cells resemble the cells of the ICM. This contention is supported by the *in vitro* development pattern of another EC cell line, PSA-1, which closely resemble cultured ICM cells in that both form embryoid bodies (an embryoid body consists of an outer differentiated endoderm cell layer and inner

pluripotential embryonal carcinoma cells, in the form of an aggregate with resemblance to the fetal portion of a 5-day old embryo)(Martin *et al.* 1977).

On the other hand, three lines of evidence argue that EC cells represent a late phase of primitive ectoderm, after the allocation of cells to the germ line (see Figure I-1). First, a comparison of proteins synthesized by early embryonic cells and EC cells indicate that EC cells are most similar to primary ectoderm (Martin *et al.* 1978). Second, Diwan and Stevens (1976) reported that isolated primary ectoderm grafted to an extrauterine site will give rise to a teratomatous tumor. Third, EC cells lack the ability to colonize the germ line. Furthermore, Stevens (1982; and references therein) concluded that EC cells are morphologically, antigenically, biochemically, and in developmental potential equivalent to embryonic ectoderm. If one is to conclude that EC cells are indeed equivalent to embryonic cells of the primary ectoderm, it is difficult to explain the results with F9 cells discussed above. Since there is a great heterogeneity among the various EC cell lines (see Chapter II), it is possible that F9 represents cells of the ICM. Alternatively, it is possible that chromosomal rearrangements or other types of events have occurred in such a way that F9 cells are able to respond to chemical inducers under different culture conditions (see above) and mimic the behaviour of ICM cells. In this view, the failure of EC cells to colonize the germ line would not be surprising as cells carrying chromosomal aberrations are generally selected against during the formation of the germ line.

I.3.ii) Kinetics of differentiation of EC cells

As stated above, EC cells are called pluripotent because they can be differentiated *in vitro* into a spectrum of cell types, either spontaneously (cell lines such as PSA-1) or with chemical inducers (cell lines such as F9, P19). However, not all EC cell lines differentiate into a wide variety of cell types. Some can differentiate into only one or two cell types (F9 cells for instance). There are cell lines which are called nullipotent, i.e., incapable of differentiation *in vivo* or *in vitro* (Martin, 1975), and cell lines which form embryoid bodies first and then differentiate into several lineages (Martin and Evans, 1975). Another pluripotent cell line, P19 (McBurney and Rogers, 1982), the cell line used in this study, can differentiate into (mainly) neurons and astroglia in the presence of retinoic acid (Jones-Villeneue *et al.* 1982) and into fibroblasts and cardiac and skeletal muscle cells in the presence of DMSO (McBurney *et al.* 1982). Furthermore, when P19 cells are treated with very high doses of DMSO, the cells differentiate into neurons (Edwards *et al.* 1983) and low doses of RA cause differentiation to muscle cells (Edwards and McBurney, 1983). It thus appears there is a fine balance of interactions among proteins encoded by genes which are responsive to these inducers. It is also interesting to note that P19 cell aggregates treated with both RA and DMSO differentiate into neurons (McBurney *et al.* 1982). This indicates that neurogenic genes are probably dominant. It is generally thought that P19 cells are less

determined and represent earlier stages of embryogenesis compared to other EC cell lines. However, I would like to point out that one could also view P19 cells as having greater plasticity in terms of developmental potential, akin to embryonic cells of the organogenesis stage.

It appears that cell aggregation is an important factor in the differentiation of EC stem cells. The cells have to be cultured at high local density either as dense monolayers or as aggregates in order to differentiate. For example, Martin and Evans (1975) described some EC clonal cell lines which formed embryoid bodies (see Martin and Evans, 1975). If these embryoid bodies are allowed to attach to tissue culture plates, they differentiate into cartilage, muscle, and pigmented epithelium.

When P19 cells are cultured as monolayers in the presence of DMSO (which induces aggregates of P19 cells to differentiate into fibroblasts and muscle cells), no differentiation is observed even after 20 days in culture. If these 20 day old cells are then aggregated, however, they form fibroblasts and muscle cells (McBurney *et al.* 1982). These findings seem to reflect some kind of position effect or cell-to-cell interaction while the cells are in three dimensional multicellular aggregates. This may have a fundamental role in the processes of cell determination and committment. Also, this situation is perhaps similar to the requirement for cell-to-cell interaction and communication through gap junctions, which occurs during normal embryogenesis and without which embryonic

development could not take place (see for example, Lee *et al.* 1987, and references therein).

I.4) Transposable genetic elements and DNA transposition

In simple terms, transposons can be defined as mobile genetic elements, which move from one location to another in the genome. Barbara McClintock was the first person to formulate a hypothesis to explain the transposition event and test it in maize plants. She was able to conclude that there are genetic units, controlling elements, as she called them, which are mobile within the cell (McClintock, 1952). She observed that they are associated with controlled chromosomal breakages, regulation of gene expression, insertional mutagenicity, etc. Since then, much work has been done in this field and transposons have been discovered in almost in every class of organisms (see Cold Spring Harbour symposium on the movable genetic elements, part I and II), including humans (Paulson *et al.* 1985).

In recent years, transposons have been classified into two categories, which applies to both eukaryotic and prokaryotic elements, based on the mechanism of transposition (Kuff *et al.* 1983; Finnegan, 1989). Type I elements are those for which the transposition mechanism involves cycles of reverse transcription and insertion of the copy DNA. Generally, linear extrachromosomal DNA copies of the reverse-transcribed RNA will be present in this type of transposition. The Type

II transposition events do not involve an RNA intermediate. Instead, they involve a direct transposition from DNA to DNA. The extrachromosomal copies here include linear, circular, and supercoiled forms. There are two recognised sub-mechanisms for the Type II event. In one mechanism, the transposon is exised from one location in the genome and the same DNA is integrated elsewhere in the genome in a conservative process (examples include the P-element of *Drosophila*, and Tc1 in *Caenorhabditis elegans*, see Finnegan, 1989; Shapiro, 1979; Calos and Miller, 1980). In the second type, which is called a replicative or duplicative transposition event, the transposon is replicated and the replicated copy is inserted somewhere else in the genome (see Federoff, 1983) (examples include the IS, Tn, fold-back and TU elements, see Finnegan, 1989). Because of this mobility within the genome, transposons cause chromosomal rearrangements such as insertion rearrangement, deletions, inversions, translocations, duplications, etc.

Below is a brief description of transposable genetic units that have been well characterized.

I.4.i) The Type I elements

The transposition of the elements listed below is via an RNA intermediate: (a) Retrovirus-like elements. Akin to retroviral LTRs, the elements of this category have direct repeats at the ends and a putative primer binding site at the left LTR.

The DNA between the LTRs has open reading frames which encode genes for group specific antigen (gag), reverse transcriptase, and another protein (see Finnegan, 1989). The members of this family are generally about 6-kb long, and the sequences of different copies of the same element are found to be variable (see Lueders and Kuff, 1980; Ono *et al.* 1980; Brulet *et al.* 1983). This heterogeneity may be due to an error-prone reverse transcriptase enzyme. Members of this class include the ETn and IAP in mouse, the copia family in *Drosophila*, the Ty element of yeast, and THE elements in humans (see Flavell and Ish-Horowicz, 1981; Temin, 1980; Paulson *et al.* 1985).

(b) Non-viral type I elements. The term non-viral refers to the absence of terminal direct repeats, unlike the retrovirus-like elements. These are also about 6-kb long, but varying lengths of DNA from the 5' end is often found to be deleted in these elements (Finnegan, 1989). These elements have an open reading frame for the gag protein and another for reverse transcriptase. The members of this class include the jockey, I, F, and G elements in *Drosophila*, the Cin4 element in maize, and the L1 elements in mammals (see Finnegan, 1989).

I.4.ii) The Type II elements

The transposition of members of this class does not involve an RNA intermediate; instead these transpose directly from DNA to DNA. The exact

mechanism of transposition is not known. This Type of element is subdivided as follows:

(a) Elements with short terminal repeats. The type II elements of this class have characteristically short inverted repeats at their ends. The internal DNA codes for at least one protein, the transposase. Members of this class include the hobo and P- elements of *Drosophila*, the Tc1 element in *Caenorabditis elegans*, Ac/Ds and Spm/En elements in maize (see Finnegan, 1989).

(b) Elements with long terminal repeats. Members of this class have long inverted repeats at their ends. For example, the Mu element in maize has 215 bp terminal repeats, the DIRS element of *Dictyostelium discoideum* has 330 bp repeats. Members of this class are of varying lengths. Some encode a putative reverse transcriptase in the terminal repeat (e.g. Mu element) and therefore do not fit exactly into any of the above classes. Examples of this type include the TU element of sea urchin, and the Fold Back element of *Drosophila*.

I.4.iii) Transposons and gene regulation

The transposons were discovered because of their effects on the expression patterns of genes. Over the past years, data have been obtained to shed light on the mechanism by which these mobile genetic units influence gene regulation. In general, a transposon can affect a gene by a) insertion into the gene itself,

usually destroying the preexisting gene function, or by b) insertion next to a gene into a place from where it can alter transcription, but not the function of the gene product.

In the literature, several genetically well characterised examples have been reported for insertion mutagenesis of a gene by a transposon. Two examples are the dilute locus insertion mutation (Copeland *et al.* 1983) and the hairless mutation (Stoye *et al.* 1988) in the mouse. These mutations have been shown to be caused by the insertion of an endogenous provirus into the genes (endogenous proviruses are also grouped with transposable genetic units, see Varmus, 1983). Both destroy the gene function by inserting into the protein coding sequence.

Another example is the insertion of the intracisternal A-particle gene (IAP), a mobile genetic unit (Kuff *et al.* 1983), into the mouse immunoglobin k-light chain gene at the intervening sequences (Hawley *et al.* 1982). This resulted in a defect in k-light chain production.

Modification of gene regulation by insertion next to the gene, without destroying the function of the gene product, entails complicated molecular mechanisms. An example is the insertion mutagenesis in *Drosophila* caused by a mobile genetic unit called gypsy (a copia family member) in yellow and hairy wing suppressor mutations. The gypsy element was found to be inserted upstream of the promoter of the yellow gene (Parkhurst and Corces, 1986). This results in the decreased expression of the yellow gene with the consequence of altered cuticle

colouration. One is tempted to speculate that the gypsy transcriptional activity in the vicinity of the yellow promoter interferes with the yellow gene transcription by co.npeting for transcriptional factors, particularly if the gypsy element contains a "short-distance" transcriptional enhancer.

Parkhurst et al. (1988) provided another explanation for the original mutation. The yellow mutation can be suppressed by gypsy element insertion mutagenesis at another locus, called suppressor of hairy wing, su(Hw). Parkhurst et al. (1988) showed that the su(Hw) protein is a gypsy enhancer binding protein. The insertion of gypsy at su(Hw) interferes with the gypsy transcription. This effectively reduces the yellow gene transcription and restores the original mutation. Parkhurst et al. (1988) suggested that yellow gene enhancer-su(Hw) protein interaction changes the chromatin organization and reduces transcription from the yellow promoter.

One important point that should be mentioned here is the consequence of transposons having transcriptional enhancers within their genetic units (see Chapters II and IV of this thesis). In particular, enhancers can activate a heterologous gene from a distance, in an orientation-independent manner, from a position either 3' or 5' of the gene. Insertion of a member of the class of transposons which have enhancers, next to a gene can have a profound effect on the expression of that gene. For example, the insertion of an IAP genome next to the Ren-2 gene in mouse resulted in the enhanced expression of the Ren-2 gene (Burt *et al.* 1984).

It is likely that these IAPs have an enhancer within their genetic unit, although this has yet to be uncovered. Another example is the provirus-conferred androgen regulation of the sex-limited protein (Slp) gene in mouse (Stavenhagen and Robins, 1988). An ancient proviral 5' LTR which contains an androgen-responsive enhancer was found to be inserted 2-kb upstream of the slp gene, and in effect regulated Slp expression (Stavenhagen and Robins, 1988).

I.4.iv) Transposons in evolution and development

There are three aspects of transposons which are of particular significance to evolution and development. The first is the ability of transposable elements to affect genes without destroying the coding sequences and therefore the preexisting function of the protein. The second is the transposition from and into a gene (see for example Schwarz-Sommer *et al.* (1985). The third is the DNA rearrangements such as deletions, duplications, and translocations, which are caused by the transposition itself. However, any DNA changes caused by transposons are subjected to natural selection. That is, those which have deleterious effects on the organism, would be subjected to negative selection. On the other hand, any changes that are advantageous to the organism would be positively selected. Transpositions which affected gene expression in a less drastic way, so that the changes were not subjected to strong selective forces, would create diversity within the species.

I will discuss the example of P-M hybrid dysgenesis here. This phenomenon, which is under genetic control (see below), is important from both the evolutionary and developmental point of view. P-M hybrid dysgenesis in *Drosophila melanogaster* is brought about by a family of transposons called P elements (Engels, 1979; O'Hare and Rubin, 1983). Strains of flies which do not have P-elements are referred to as the M strains. When P elements are introduced into an M strain by crossing P-strain males to M-strain females, P-element transposition is stimulated and hybrid dysgenesis occurs in the germ line of the offspring (see Rubin *et al.* 1985). These dysgenic offspring have a high incidence of sterility and show chromosomal rearrangements only in their germ line.

The most important point here is that mutagenesis of the germline occurs. This can introduce favorable as well as harmful mutations that might influence development in the offspring. At the same time, it is reasonable to hypothesize that mutations that have minor effects can also occur during dysgenesis. In fact, lines of *Drosophila* subjected to P-M hybrid dysgenesis showing more variability in abdominal bristle numbers compared to normal flies have been reported (Mackay, 1985). This type of mutation, which may not be subjected to strong selective pressure, may have a role in species diversity and genetic variability. The direct participation of transposons or DNA transpositions in eukaryotic embryonic development however, does not appear to take place, as judged from nuclear transfer studies.

An important aspect of transposons that has been utilised in developmental studies is that they can be used as gene transfer vectors. Spradling and Rubin (1982a, b), using the P-element transposon of *Drosophila* as a vector, were able to introduce cloned genes into the germ line at a high frequency. The cloned gene was inserted between the short inverted repeats of the P-element and then microinjected into the posterior pole of an early embryo. This procedure has been widely used in studying *Drosophila* embryogenesis (see for example, Kuziora and McGinnis, 1988). It can also be used to transfer genes in higher eukaryotes using other transposons, except that the introduced gene will not be specific to the germ line as in the case of P-element mediated gene introduction. An attractive system in the mouse would be to use the ETn to introduce genes to embryos. Since the expression of the ETn is specific to early embryonic cells (see Chapter IV of this thesis), the expression of a gene introduced under ETn regulation would also be stem cell specific. This would allow one to manipulate gene expression programs specifically in the founder cells of the embryo.

I.5) Statement on the research problem

EC cells, as described above, closely resemble stem cells of the early mouse embryo, and can differentiate into cells similar to definitive embryonic cells. Therefore, EC cells have provided a useful system for studying embryogenesis

(Martin, 1980). Understanding the first stages of embryogenesis at a molecular level and the mechanism that regulates gene expression during early development requires (a) the isolation of genes that are differentially expressed when the earliest decisions for the differentiation are made, and (b) the cis acting and trans acting proteins that control their expression. In recent years, several genes have been isolated that are differentially expressed, such as laminin and type IV collagen for parietal endoderm and α-fetoprotein for visceral endoderm. These genes are expressed during later stages of development (Cooper *et al.* 1983; Dziadek and Adamson, 1978) and are unlikely to be involved in the initial process of determination. Thus it is of interest to isolate and study genes whose expression is regulated in the earliest stages of development. Except for one unknown gene (Stacey and Evans, 1984), which was not characterized in detail, no genes or regulatory elements which are specifically active in the pluripotent embryonic stem cells of an earlier stage have been isolated.

Stem cell specificity of gene expression would mean that the genes are expressed only in undifferentiated stem cells (due to their regulatory systems) but would be shut off at the onset of differentiation. This type of gene may include regulatory genes which are involved in the expression of other genes and might be required to maintain the undifferentiated pluripotent state of stem cells. Inactivation of such genes may be necessary for the cells to undergo differentiation. In the past, Levine *et al.* (1984) and others (Ikuma *et al.* 1986 and references therein) have

attempted to isolate such stem-cell specific genes. Their approach was to isolate specific cDNA clones from EC cDNA libraries by a differential plaque hybridization procedure. However, these attempts have not been successful and candidates for regulatory genes were not identified.

Another approach adopted by Walter Gehring and co-workers in Drosophila (O'Kane and Gehring, 1987) and Surani and colleagues in mouse (Allen *et al.* 1988) to identify differentially expressed genes was to mark the chromosomal loci with the LacZ gene and follow the expression pattern of this transgene during embryogenesis. Those chromosomal loci which support the transgene in a temporally and/or spatially interesting manner can be cloned using the transgene as probe and a full molecular analysis at the integration sites can be made. Again, neither of these studies identified any chromosomal loci or genes that are specifically active in embryonic stem cells.

I.5.i) Specific objectives

The specific objectives of this work were:

1) to obtain evidence for the existence of regions of the mouse genome that are active in undifferentiated stem cells but not in their differentiated derivatives.

2) to clone such loci using recombinant DNA technology and to:

   a. characterise them by searching for the presence of transcriptional enhancers,

   b. characterize the enhancers in terms of stem cell stage specificity, and

   c. search for the presence at these loci of stem cell specific genes.

CHAPTER II

CLONING OF MOUSE CHROMOSOMAL LOCI SPECIFICALLY ACTIVE IN EMBRYONAL CARCINOMA STEM CELLS

II.1) INTRODUCTION

Determination and differentiation during embryogenesis is believed to be directed by a well defined programme of gene expression. There are reasons to believe that this may be influenced by active chromosomal domains. For instance, using a transgene (an exogenous gene) as a probe, O'Kane and Gehring (1987) in *Drosophila* and Surani and co-workers in mouse (Allen *et al.* 1988) showed that expression of an exogenously introduced transgene during development is influenced by active chromosomal domains. In addition, Jaenisch and co-workers (Jaenisch *et al.* 1981; Barklis *et al.* 1986) had reported the presence of chromosomal loci in pre-implantation mouse embryos and in undifferentiated EC cells; these loci were special in that they allowed the expression of exogenously introduced Moloney leukaemia virus, despite the fact that both preimplantation embryos and undifferentiated EC cells are refractive to viral genome expression (Jaenisch and Berns, 1977; Stewart *et al.* 1982). The concept of an active (or an inactive) chromosomal domain as a dominant force in regulating gene expression may be

open to criticism, (effect as opposed to cause) but the finding by Grosveld and co-workers that β-globin gene expression is regulated by 'locus activation domains' (Grosveld *et al.* 1987; Talbot et al., 1989) demonstrated such a possibility.

Whether it is the state of a chromosomal domain that determines the gene activity or conversely, the activity of a gene which determines the state of the chromosomal domain, it is likely that in such domains or in the vicinity of such domains, genes exist. These chromosomal loci can be detected using transgenes which integrate randomly throughout the genome, since their chromosomal position can influence their expression pattern (Palmiter and Brinster, 1986; Jaenisch *et al.* 1981; Lacy *et al.* 1983; O'Kane and Gehring, 1987; Allen *et al.* 1988; Grosveld *et al.* 1987 and the references therein). When such domains are identified, the transgene can be used as a genetic marker from which a full molecular analysis of the integration sites could be made. This would then lead to the identification of endogenous genes or regulatory sequences that were responsible for the transgene position-effects observed.

In this work, we sought to explore the above phenomenon and to isolate genes whose expression is specific to stem cells. Stem cell specific genes would be active in undifferentiated stem cells but would be inactivated at the onset of differentiation. Such genes may be important for maintaining the pluripotency of embryonic cells. Inactivation of these genes might be necessary in the cell determination and differentiation into lineages. The chromosomal regions that

contain such genes will therefore be inactive following differentiation, and it would be possible to isolate such regions using a selectable marker transgene as a probe. As a first step, I wished to isolate mouse chromosomal domains that are specifically active in the stem cells with a functional selection procedure using the cis-activation dependent Neomycin resistance gene (Neo-R) construct as the dominant selectable transgene marker. Associated genes can subsequently be identified. The selection procedure is based on the observation that an "inactive" enhancerless gene can be activated if it integrates near an endogenous enhancer or, by other cis activation mechanisms (Hamada, 1986a). Previously, this method was applied to HeLa cells and two enhancer elements were isolated (Hamada, 1986b; Swift *et al.* 1987). In the present study, by applying the same procedure to a murine embryonal carcinoma cell line, chromosomal loci specifically active in the EC stem cells have been successfully isolated.

II.2) MATERIALS AND METHODS

In the following paragraphs, the help received from H. Hamada has been acknowledged wherever applicable.

II.2.i) Cells, Plasmids and Phage Vectors

The cell line P19S18 (McBurney and Rogers, 1982), a pluripotent EC cell line used in this study to isolate stem cell specific loci, was provided by Dr. McBurney of the University of Ottawa, Canada. The undifferentiated P19 cells were maintained in α-Minimum Eagle's Medium (MEM) containing 10% fetal calf serum as described previously (Rudniki and McBurney, 1987).

The plasmid pA10neo used in this study as a trap to select stem cell specific loci was constructed by ligating the *Hind*III-*Bam*HI 2.2 kilobase (kb) fragment from pSV2neo (Southern and Berg., 1982) to the *Hind*III-*Bam*HI 4kb-fragment from pA10cat (Laimins *et al.* 1984). The promoterless plasmid pOneo was constructed by deleting the SV 40 early promoter from pA10neo, by subcloning *Bam*HI-*Hind*III 4kb-fragment from pA10neo into puC 12 vector.

The phage vectors used in this study were λEMBL-3A and λEMBL-12. EMBL-3A has cloning sites for *Sal*I, *Bam*HI and *Eco*RI and EMBL-12 has cloning sites for *Sal*I, *Bam*HI, *Sst*I, *Xba*27I and *Eco*RI (Natt and Scherer, 1986). Both vectors accommodate DNA fragments of sizes between 8 kb and 23 kb. Phage were propagated either in NM 538 or VCS 257 strains of E. Coli.

The plasmid vectors used in subcloning were pUC 12 and pUC 18. Plasmids were propagated in *E. coli* C600.

II.2.ii) Isolation of NEO$^+$ transformants

P19 cells (5 x 10^5) were plated in a dish and 10 ug of pA10neo was precipitated with calcium phosphate and transfected to cells as described previously (Gorman *et al.* 1982). At 36 h after the transfection, nonselective medium was replaced by medium containing 200 μg of the Neomycin analogue Genetecin 418 (G 418, Genetech). P19 cells appeared to be very sensitive to G418, therefore, the transfected cells were not replated. After 10 to 12 days, NEO$^+$ colonies were isolated and maintained in the selective medium. Those established as stable cell lines were used for further study.

II.2.iii) Induction of Differentiation

P19 cell lines were induced to differentiate as follows. Undifferentiated cells were trypsinized and seeded in a petri dish and incubated for four days in the presence of chemical inducers, i.e. retinoic acid (RA) (Sigma) or Dimethyl Sulfoxide (DMSO). To induce cells to neurons, 300 nM RA was used and for muscle cell induction 1% DMSO was used. The medium was changed once after two days.

During the four-day incubation, the cells formed aggregates. These aggregates were subsequently plated on a tissue culture dish and incubated for an additional four days in the absence of the chemical inducer. Differentiated cell

types such as neurons and muscle cells become obvious at day seven or eight after the induction. Phase contrast pictures were taken at day eight. When the cells were induced to differentiate in the presence of G418, 200 μg of G418 per ml was present throughout.

II.2.iv) Genomic DNA extraction

High molecular weight genomic DNA (100-200 kb) was prepared from cells as follows: The cells were washed with ice cold PBS and scraped into a 50 ml tube, and centrifuged for 10 min at 3000 rpm and the supernatant was discarded. Cells were resuspended in ice cold PBS and spun again for 10 min at 3000 rpm and supernatant was discarded. This was repeated once more. The cells were resuspended in 1 volume lysis buffer (1 ml/10^8 cells) (Lysis buffer: 0.1 M NaCl, 10 mM Tris HCl, pH 8, 25 mM EDTA, pH 8, 0.5% SDS, 0.1 mg/ml proteinase K added fresh with each use) and incubated with shaking at 50 °C for 12-18 h. After incubation, the samples were gently extracted twice with an equal volume of phenol:chloroform. The DNA was dialysed overnight against 50 mM Tris, pH 8, 10 mM EDTA and 10 mM NaCl. The samples were then treated with 100 μg/ml DNase-free RNase at 37 °C for 3 h. After 3 h, they were extracted gently with phenol:chloroform and dialysed extensively against TE (10 mM Tris, 1 mM EDTA, pH 8).

II.2.v) RNA extraction

Cells were washed several times with ice cold PBS to remove all media, which contains Rnase, and scrapped off the plates with a rubber policeman. Cells were transferred to 50 ml tubes and washed again with cold PBS two times (5 min centrifugation at 3000 rpm). The cells were then suspended in hypotonic reticulocyte standard buffer (10 mM Tris-HCl, pH 7.4, 10 mM KCl, 1.5 mM $MgCl_2$) at a concentration of 2.5 x 10^7 cells/ml and transferred to a dounce homogenizer. The cells were kept on ice for 15 min and 0.3% final (v/v) NP 40 was added. The cells were homogenized on ice using a motor driven pestle with 12-15 strokes. The homogenized cells were centrifuged at 3000 rpm for 5 min at 4°C to remove cell debris and extracted with a final concentration of 1 volume of phenol:chloroform, 10 mM Tris-HCl, pH 9.6, 0.5% SDS, 0.5 M NaCl. After centrifugation at 3000 rpm for 10 min at 4C, the aqueous phase containing RNA was ethanol precipitated at -70 °C.

Most of the above work, listed in II.2.i to II.2.v was done by H. Hamada.

II.2.vi) Polyadenylated (poly A+) RNA selection

Poly A+ RNA selection from the total RNA prepared from undifferentiated (D-) and differentiated (D+) P19 cells was done by chromatography on oligo (dT)-cellulose (Maniatis *et al.* 1982).

II.2.vii) Restriction mapping of the chromosomal loci

For cloning the cellular sequences flanking the Neo-R gene from the Neo-R-transformed cell lines, restriction mapping of the Neo-R integration site was performed by genomic Southern blotting. The *Bam*HI-*Hind*III 2.2 kb fragment containing the Neo-R coding sequences of pA10neo was used as a probe. The probe was labelled with α-32P dCTP by Nick translation (Rigby *et al.* 1978) or by the Random primer method (Feinberg and Vogelstein, 1983) (Amersham/BRL kit).

A detailed restriction mapping of these loci was determined by: 1) a combination of enzyme digestions (eg. *Bam*HI, *Bam*HI+*Eco*RI, and *Eco*RI) and analysis by agarose gel electrophoresis, 2) by subcloning smaller fragments in pUC and analyzing them by enzyme digestions and gel electrophoresis, 3) by Southern blotting analysis using single-copy probes from the various loci (see Materials and Methods, Chapter III for the mapping and the isolation of single-copy probes).

II.2.viii) Southern blot analysis

Southern blotting analysis of DNA was done essentially as described in Maniatis *et al.* (1982) with minor modifications. For the analysis of cloned DNA, 1 μg of the DNA was used. The DNA was digested with appropriate enzymes followed by Phenol:Chloroform (1:1) extraction, and concentration by ethanol

precipitation. (For cloned DNA, the digestion mixture was directly loaded on to a gel). The DNA was loaded onto a 0.7% agarose gel and electrophoresed overnight. The gel was stained with ethidium bromide (10 mg/ml), and photographed. The DNA was partially hydrolysed by soaking the gel in 0.5 M HCl for 20 min, and then denatured by soaking the gel in several volumes of 1.5 M NaCl and 0.5 M NaOH for 90 min followed by neutralising the gel by soaking in 0.5 M Tris (pH 8.0) and 1.5 M NaCl for 2 h. The Southern transfer of DNA from gels to nitrocellulose filters (S & S) was done in 10 x standard sodium citrate (SSC) buffer (Maniatis *et al.* 1982) for more than 16 h. Following transfer, the filter was soaked briefly in 5 x SSC, dried at room temperature and baked for 3 h at 80 °C under vacuum.

Pre-hybridization and hybridization were done according to Maniatis *et al.* (1982) with minor modification. The pre-hybridization solution included 50% deionised formamide, 5 x SSC, 50 mM sodium phosphate (pH 7.5), 5 x Denhardt's solution (Maniatis *et al.* 1982), 0.5% SDS and 200 μg/ml sonicated, heat-denatured salmon sperm DNA (Sigma). Pre-hybridization was done overnight at 42 °C. Hybridization conditions were exactly the same except that the solution contained heat-denatured labelled probe and hybridization was done for two days. The filters were washed in 1 x SSC-0.1% SDS for 1.5 h at 60 °C and 0.1 x SSC-0.1% SDS for an additional 1.5 h at 55 °C. The filters were dried and exposed for autoradiography at -70 °C.

II.2.ix) Northern blot analysis

About 10 μg of poly A+ RNA was electrophoresed on a formaldehyde-agarose gel (the procedures used were from BRL, published in their Focus, Vol. 8, No. 2, 1986). After electrophoresis, the gel was soaked for 15 min in distilled water (the gel was not stained) and proceeded directly for blotting onto a nitrocellulose filter in 20 x SSC overnight. Pre-hybridization, hybridization and washing conditions were exactly the same as used for the Southern filters.

II.2.x) Molecular Cloning procedures

For cloning the cellular DNA flanking the Neo-R gene from the NEO^+ cell lines, the following molecular cloning procedures were adapted.

II.2.x.a) Construction of restricted genomic library

In order to enrich the sequences of interest, restricted genomic libraries were constructed as follows: Genomic DNA (about 400 μg) from each cell line was digested with an appropriate restriction enzyme and fractionated on a 0.55% agarose gel (electrophoresis was done for two days). Gel slices were excised from the appropriate size-region of the gel (appropriate size-region: region of the gel where

the DNA fragment to be cloned is present, determined by genomic Southern blot) and the DNA from the gel slices were electroeluted to Bio-Gel hydroxylapatite. The DNA was purified from hydroxylapatite (Bio-Rad) by eluting with 1 M potassium phosphate buffer (pH 6.8). The potassium phosphate buffer was removed from the DNA by DEAE Sepharose chromatography (the DNA was eluted with 0.6 M NaCl in TE). The fragments were extracted once with Phenol:Chloroform followed by extraction with chloroform and precipitation twice with ethanol. The DNA was washed twice with 80% ethanol, lyophilised and suspended in 5 mM Tris (pH 7.5). About 100-200 ng of this DNA was analyzed by Southern hybridization for the presence of the sequence of interest before ligating them to phage arms (see Figure II-1, A and B).

II.2.x.b) Preparation of the phage arms

Phage DNA (about 20 μg), either λEMBL 12 or λEMBL 3A, was digested with the appropriate restriction enzymes (e.g. for cloning *Bam*HI fragments, the phage DNA was digested with *Bam*HI and *Eco*RI; the digestion of phage arms with *Eco*RI will inactivate the phage middle fragment and it will not compete with the insert DNA for ligation into the arms) and the resulting small linker fragment was removed by a quick precipitation with 0.6 volume of isopropanol (Frischauf *et al.* 1983), followed by quick precipitation with ethanol (quick precipitation: as soon as isopropanol/ethanol was added in the presence of 0.3 M sodium acetate, the

Figure II-1. Construction of restricted genomic library.

High molecular weight genomic DNA (400 μg) digested to completion with an appropriate enzyme (1500 Units) was size fractionated on a 0.55 % agarose gel. (A): fragments of appropriate size-range (which includes the sequence of interest) were exised in three fractions from the gel, purified, and an aliquot from each fraction was electrophoresed on a minigel, together with the vector phage arms. T, top fraction; M, middle fraction; B, bottom fraction. See text for the preparation of phage arms.

(B): the minigel in (A) is Southern blotted and probed with pA10neo. The blot was exposed for one week. The sequence of interest (Neo-R gene plus the flanking cellular sequences) was present in the middle fraction (the sharp band). The cross hybridization seen with the marker lane and the vector phage arms was due to some homology of pUC sequences of pA10neo to these phages. The middle fraction was ligated to phage arms in a molar ratio of 1:1 and subjected to *in vitro* packaging (Stratagene). The recombinant phages were screened with the Neo-R specific probe (*Bam*HI-*Hind*III 2.2-kb fragment of pA10neo) and the clone of interest was isolated by the plaque hybridization method (see text for more details).

A
Marker
T
M
B
Phage arms
Marker
23.3
9.4
6.6
4.3
2.3
2.0
X kb
B
Marker
T
M
B
Phage arms
Marker

phage arms precipitate; they were immediately microfuged for 10 min and the solution was discarded along with the unprecipitated linker sequences). The pellet was washed with 80% ethanol twice, lyophilised and suspended in 5 mM Tris (pH 7.5) (see Figure II-1 for gel pattern of the phage arms).

II.2.x.c) Ligation and *in vitro* packaging

The insert DNA was ligated to the phage arms in a molar ratio of 1:1. The total DNA concentration in the ligation reaction was usually > 500 μg/ml. The ligation reaction was carried out for about 20 h at 4°C (1 x ligation buffer: 20.5 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 2 mM DTT, 0.4 mM ATP). The ligated DNA was then subjected to *in vitro* packaging using Giga-pack gold packaging extract from Strategene, according to the protocol supplied by them. The resulting recombinant phages were screened with the Neo-specific probe by the plaque hybridization procedure (Maniatis *et al.* 1982).

II.2.x.d) Storage of phage

Phage were always stored in SM buffer (Maniatis *et al.* 1982) at 4°C. Dilution of phages were also done in SM buffer.

II.2.x.e) Host bacteria

Host bacteria (NM 538 or VCS 257) were grown at 37°C in Luria-Bertani (LB) medium (recipe is from Maniatis *et al.* 1982) containing 10 mM $MgSO_4$ (LBM) and 0.2% maltose (LBMM) until the OD_{600} was about 1.0. The bacteria were pelleted by centrifugation at 3000rpm for 20 min and suspended in 10 mM $MgSO_4$ at a concentration of 1.5 x 10^9 cells/ml.

II.2.x.f) Screening Recombinant plaques by Plaque hybridization

All the procedures for plaque hybridization were adopted from Maniatis *et al.* (1982). Briefly, about 5 x 10^4 phages were used to infect 4.5 x 10^8 freshly grown host bacteria and incubated at 37°C for 30 min. After incubation, 7.5 mL of soft agarose (0.7% agarose in LBM medium, top agarose) maintained at 48°C were added to the above infection tubes and plated on a 150 mm LBM plate containing 1 2% Agar. The plates were incubated for about 12 h for λEMBL 3A and about 9-10 h for λEMBL 12. The plates were chilled for 2-3 h at 4°C and replica filters (nylon filters (NEN) that were marked asymmetrically at four positions) were made from the plates. While making the replica filters, the marks from the filters were copied on to the plates. The filters were then dried at room temperature for 2 h and treated sequentially with 0.2 N NaOH - 1.5 M NaCl solution for 1 min, 0.5

M Tris (pH 7.5) - 1.5 M NaCl solution for 1 min, and 5 x SSC for 15 sec. The filters were then dried for about 2 h before proceeding to the pre-hybridization step.

For pre-hybridization, the filters were wetted in 6 x SSC and incubated in a solution containing 5 x SSC and 1% SDS at 65 °C overnight in a sealed bag. Hybridization was done in a solution containing 5 x SSC, 1% SDS, 5 x Denhardt's solution, 200 μg/ml sonicated salmon sperm DNA, 200 μg/ml sonicated *E. coli* DNA, and nick translated Neo-specific probe. The carrier DNAs and the probe were heat denaturated before adding to the hybridization bag. The hybridization was done for 24 h at 65 °C and washed as described for the Southern filters and exposed for autoradiography at

-70°C overnight.

In order to isolate the positive plaques the marks from the filters were first copied onto the autoradiogram. The plate was then aligned to these marks on the X-ray and using a sterile tooth-pick, a small circle of 5 mm diameter was cut around the positive signal and this piece of top agarose was transferred to 1 ml of SM buffer. The plaques were extracted from the top agarose by rotating the tube for 3-4 h. A drop of chloroform was added and the solution stored at 4°C. The purity of the positive plaque at this stage was usually about 1%.

For further purification of the positive plaque, about 1 x 10^4 phages were plated on a 150 mm plate and the positive plaque was selected by plaque

hybridization as above to yield a purity of about 10%. To obtain 100% purity of the positive plaque, one more round of plating was done. About 100-200 phages were plated on a 90 mm LBM plate and a single positive plaque, which is well separated in the plate, was picked. This plaque was amplified and used for large scale phage DNA preparation.

II.2.x.g) Phage amplification

Phages were amplified by the plate lysate method (Maniatis *et al.* 1982) to use in the large scale phage culture (it is necessary to obtain a phage solution of high titre, > 10^{10}/ml, for this purpose). Two rounds of amplifications were done to achieve this. First, about 1.5 x 10^3 phages were plated on a 90 mm LMB plate and the phages were recovered from the plate in about 4mL SM buffer. In the second round, about 3 x 10^5 phages were plated on a 150 mm plate and phages were recovered as before in about 20mL SM buffer. The titre of this solution was usually > 10^{10}/ml.

II.2.x.h) Phage culture

For large scale phage culture, the multiplicity of infection (MOI) used was 1.0. The bacteria were infected when the OD_{600} was between 0.2 and 0.3. The

remaining steps of the phage culture and phage DNA preparation were as described in Maniatis *et al.* (1982).

Small scale phage culture was done by the plate lysate method as described in Maniatis *et al.* (1982).

II.2.xi) Subcloning procedures

DNA which was cloned in phage was subcloned into plasmids. pUC 12 and pUC 18 were used as vectors for this purpose. For subcloning fragments that contained the Neo-R gene (such as the entire fragments cloned from the stem cell specific NEO^+ cell lines, or deletion fragments that retained the Neo-R gene), the recombinant plasmids were selected on kanamycin plates (15 μg/ml). In the remaining cases, ampicillin plates (40 μg/ml) were used for selection. All the procedures for subcloning, plasmid culture and plasmid DNA preparations were from Maniatis *et al.* (1982). While preparing plasmid DNA, the SDS lysis method was used. To prepare form I plasmid DNA, first the DNA preparation was chromatographed on a Sephacryl 4B column (60 x 1.5 cm) equilibrated with TE, to separate proteins and RNA from the plasmid DNA. The plasmid DNA was ethanol precipitated once and then subjected to cesium chloride-ethidium bromide centrifugation as described in Maniatis *et al.* (1982).

Rapid, small scale plasmid preparation was done as described in Maniatis *et al.* (1982).

II.2.xii) Assays

The following assays were done by H.Hamada.

II.2.xii.a) Transient expression assay

The chloramphenicol acetyl transferase assay (Gorman *et al.* 1985) was used to demonstrate the enhancer dependence of the SV40 early promoter in P19 cells. Briefly, 10 μg of pUC-cat (a CAT gene linked to the enhancerless SV40 early promoter) or pSV-cat (an enhancer-plus counterpart) was transfected to P19 cells. At 48 h after transfection, the cells were harvested, and the cell extracts were prepared by repeated freezing and thawing. Cell extracts containing 200 μg of protein were incubated with 1 μCi of [^{14}C]-chloramphenicol (NEN) at 37°C for 30 min. Acetylated and non-acetylated forms of chloramphenicol were separated by thin-layer chromatography. After the chromatography, the chromatograms were exposed to X-ray film at -70°C overnight.

II.2.xii.b) Stable transformation assay

A stable transformation assay was done to: a) demonstrate the enhancer dependence of the SV40 early promoter in P19 cells, b) to assay the enhancer

activity of the cloned DNA, and c) to demonstrate the stem cell specificity of cloned DNA.

About 5 x 10^5 cells (P19 or NIH 3T3 cells) plated on a dish were transfected with 2.5 pmol of plasmid (10 μg for pA10neo) by the calcium phosphate method. The cells were incubated in nonselective medium for 36 h. The medium was then replaced with the selective medium (200 μg of G418/ml for P19 cells and 400 μg/ml for NIH 3T3 cells). In the case of P19 cells, the cells were not replated. However the transfected NIH 3T3 cells were trypsinized 36 h after the transfection and 10^5 cells were replated. Cells were incubated for an additional 12 days and stained with Giemsa stain. Colonies consisting of more than 100 cells were counted.

II.3) RESULTS

II.3.i) Strategy for selection of embryonic stem cell specific loci

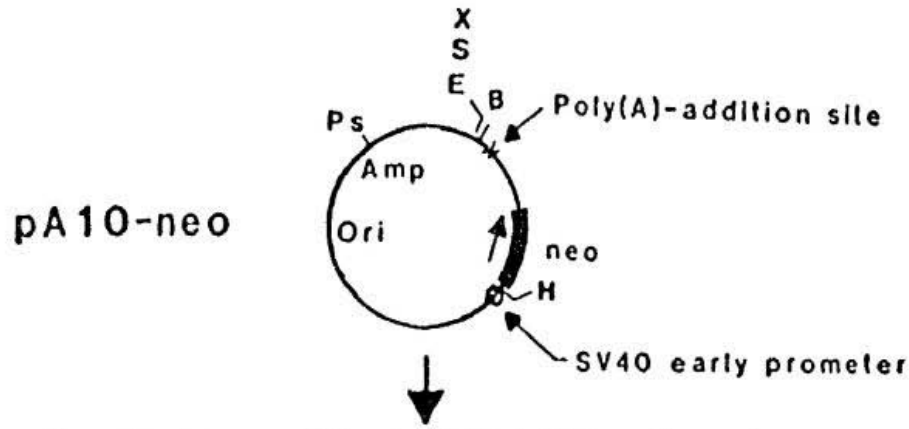
The present strategy is based on the previous observation that an enhancerless gene, which is transcriptionally inactive by itself, can be activated upon transfection if the gene is stably integrated near an endogenous enhancer (Hamada 1986a, b). The enhancer can be rescued by molecular cloning using the integrated marker gene as a probe. This procedure was previously applied to HeLa cells and two distinct enhancers were successfully isolated from the HeLa genome (Hamada

1986b; Swift *et al.* 1987). In this study, the same procedure was applied to P19 cells, a murine pluripotential EC cell line (McBurney and Rogers, 1982). This cell line can be induced to differentiate into a variety of cell types, such as neurons, astroglia, fibroblasts, and muscle cells (Edwards *et al.* 1983; Jones-Villeneuve *et al.* 1983). When the cells are injected into a blastocyst, they contribute to somatic tissues of the resulting chimaeric mouse. Therefore, P19 cells appear to represent the early embryonic cells (inner cell mass or primitive ectoderm) of the mouse development.

The strategy used for selecting chromosomal loci specifically active in the EC stem cells is summarised in Figure II-2. First, P19 cells were transfected with pA10neo, an enhancer-trap plasmid containing the Neo-R gene linked to the enhancerless SV40 early promoter. NEO^+ cells were selected and a number of transformants were established. Each cell line was assayed to determine 1) the copy number of the integrated Neo-R plasmid, 2) the ability to differentiate in the presence or absence of the neomycin analogue G418, and 3) the level of Neo-R gene expression before and after differentiation. Only those cell lines that a) contained a single copy of the Neo-R gene, b) could differentiate into a wide variety of tissue types similar to the parental cell line, and c) showed a greatly reduced Neo-R gene expression following differentiation (as judged from the failure to differentiate normally in the presence of G418 and greatly reduced Neo-R mRNA after differentiation indicated by the Northern blot analysis) were selected for further study. In such cell lines, the Neo-R gene is presumably activated through

Figure II-2. The strategy for cloning of the EC stem cell specific chromosomal loci.

The functional map and restriction sites of the enhancer- trap plasmid pA10neo is shown at the top. B, *Bam*HI; H, *Hind*III; E, *Eco*RI; Ps, *Pst*I; S, *Sst*I; X, *Xba*I. The SV40 early promoter (including a TATA box and GC boxes) is present within a 200 bp region upstream of the *Hind*III site (see text for details).



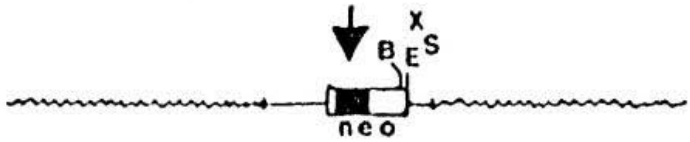
Transfect to undifferentiated P19 cells and select NEO$^+$ transformants

Select cell lines with the single copy

Select those cell lines which show reduced neo mRNA after differentiation by:

1. Inducing differentiation in the absence or presence of G418
2. Assaying the neo mRNA level before and after the differentiation

Determine restriction map



Cloning in phage

Subcloning in plasmid

Assay NEO$^+$ transforming activity

a stem cell stage specific regulatory pathway such as cis acting DNA sequences (enhancer or promoter). Cellular sequences flanking the integrated Neo-R gene were cloned from those cell lines and assayed for their stem cell specificity and searched for regulatory elements.

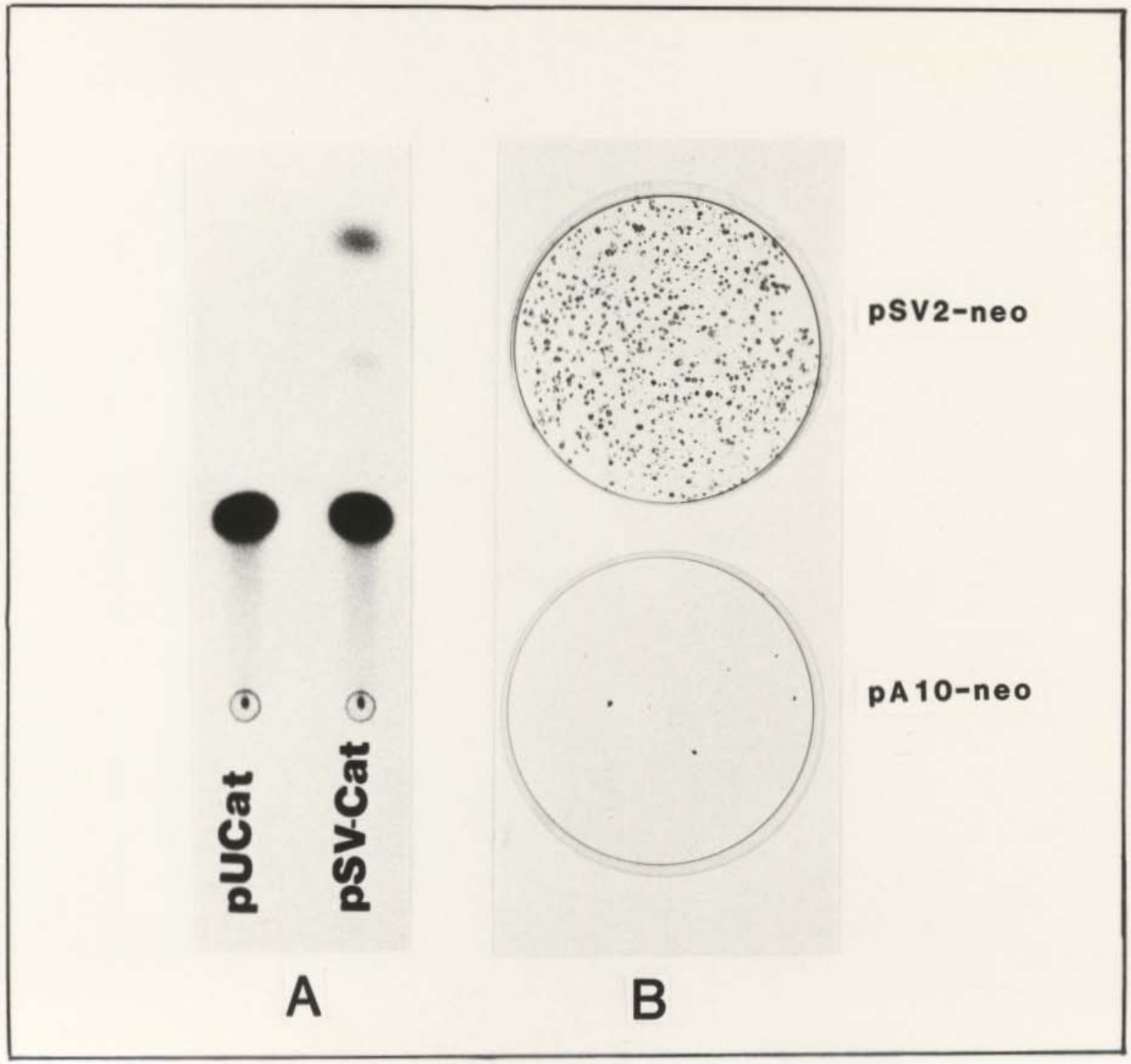
II.3.ii) The SV40 early promoter is enhancer dependent in P19 cells

To apply this selection procedure to a given cell line, one critical requirement is that expression of the marker gene must be enhancer dependent in the cells. It was important to determine whether P19 cells satisfy this criterion, because the SV40 early promoter has been reported to be enhancer independent in F9 cells, another EC cell line (Gorman *et al.* 1985). This was tested by two assays. When the activity of the enhancerless constructs (pUC-cat and pA10neo) and the enhancer plus counterparts (pSV-cat and pSV2neo) were examined by transfection, both the transient (cat) expression assay (Figure II-3, panel A) and the stable transformation assay (panel B) showed that the enhancerless promoter (pUC-cat and pA10neo) was much less active than the enhancer-linked promoter (pSV-cat and pSV2neo). These results demonstrated to us that the SV40 early promoter is enhancer dependent in P19 cells and that it can be used in the functional selection procedure.

Figure II-3. The SV40 early promoter is enhancer-dependent in P19 cells.

(A) Transient assay results. pUC-Cat (10 μg) containing a cat gene linked to the enhancerless SV40 early promoter, or pSV-Cat, an enhancer plus counterpart, was transfected to undifferentiated P19 cells. At 48 h after the transfection, the cells were harvested, and the expression of the cat gene was monitored by enzymatic conversion of chloramphenicol to its acetylated forms.

(B) Stable transformation assay results. The enhancerless construct, pA10neo (10 μg), or the enhancer plus counterpart, pSV2neo was transfected to P19 cells, and NEO^+ colonies were selected in the presence of G418 as described in Materials and Methods.


pUCat
pSV-Cat
pSV2-neo
pA10-neo
A
B

II.3.iii) Characterization of NEO$^+$ cell lines

When P19 cells were transfected with the enhancerless construct pA10neo, NEO$^+$ colonies appeared with a low frequency (see Figure II-3B). A total of 80 NEO$^+$ colonies were recovered, and 40 of them were established as stable cell lines. The results of dot blot hybridization analysis (Hamada 1986a) with the Neo-R-specific probe showed that 30 of the cell lines contained a single copy of the Neo-R gene (data not shown). These 30 cell lines were selected for further study.

When each of the 30 cell lines were induced to differentiate by retinoic acid in the absence of G418, all were able to differentiate into neurons and glial cells like the parental cell line (data not shown), indicating that all of the cell lines have retained their pluripotency. Next, we wished to select those cell lines (out of 30) in which the Neo-R gene had integrated into a stem cell specific locus. Hence, each of the 30 cell lines was subjected to selection assays.

First, each cell line was induced to differentiate in the presence of G418. If the Neo-R gene has integrated into a stem cell specific locus, such a cell line would be unable to complete normal differentiation, since it would then not be resistant to the antibiotic, G418. In this assay, a total of 11 cell lines transformed with pSV2neo were used as positive controls (the SV40 enhancer is active throughout differentiation), and the parental P19 cell line was used as a negative control. As expected, all of the pSV2neo-transformed cell lines were able to

complete normal differentiation, while the P19 cells died before forming differentiated colonies (data not shown).

When each of the 30 NEO^r cell lines selected above were induced to differentiate in the presence of G418, they fell into two groups (Figure II-4). The first group of cell lines (group I) differentiated normally in the presence (or in the absence) of G418; the retinoic acid-induced aggregates formed large colonies containing numerous neurons and glial cells (Figure II-4, a and b). Among the 30 cell lines, 22 showed this characteristic indicating that in these 22 loci, the Neo-R gene was integrated into non-stem cell specific loci. However, the second group (group II, 8 cell lines) showed an abnormal differentiation pattern in the presence of G418 (Figure II-4d). The aggregates did not grow well after being plated on tissue culture dishes, replicated glial cells were much less frequent than normal, and neurons were rarely detected (compare c and d in Figure II-4). This suggests that the Neo-R gene may have integrated into stem cell specific loci in these cell lines.

To confirm the above results, the expression of the Neo-R gene before and after differentiation was examined. Poly A+ RNA was prepared from undifferentiated and differentiated cells, and the Neo-R mRNA levels were determined by Northern blotting analysis (Figure II-5). The Neo-R message was examined in two pSV2neo-transformed P19 cell lines (cell lines 111 and 121, positive control), one cell line from the group I and all the eight cell lines from group II (which showed abnormal differentiation pattern in the presence of G418). In the pSV2neo

Figure II-4. Retinoic acid induced differentiation in the absence (-G418) or presence (+G418) of G418.

Each NEO^+ cell line was induced to differentiate by retinoic acid in the absence (a and c) or presence (b and d) of G418. A total of 22 cell lines were classified as Group I, where the cell lines could differentiate normally in the absence (a) or presence (b) of G418 (compare a and b). Eight cell lines were classified as Group II, where the cell lines could not differentiate normally in the presence of G418 (d)(compare c and d).

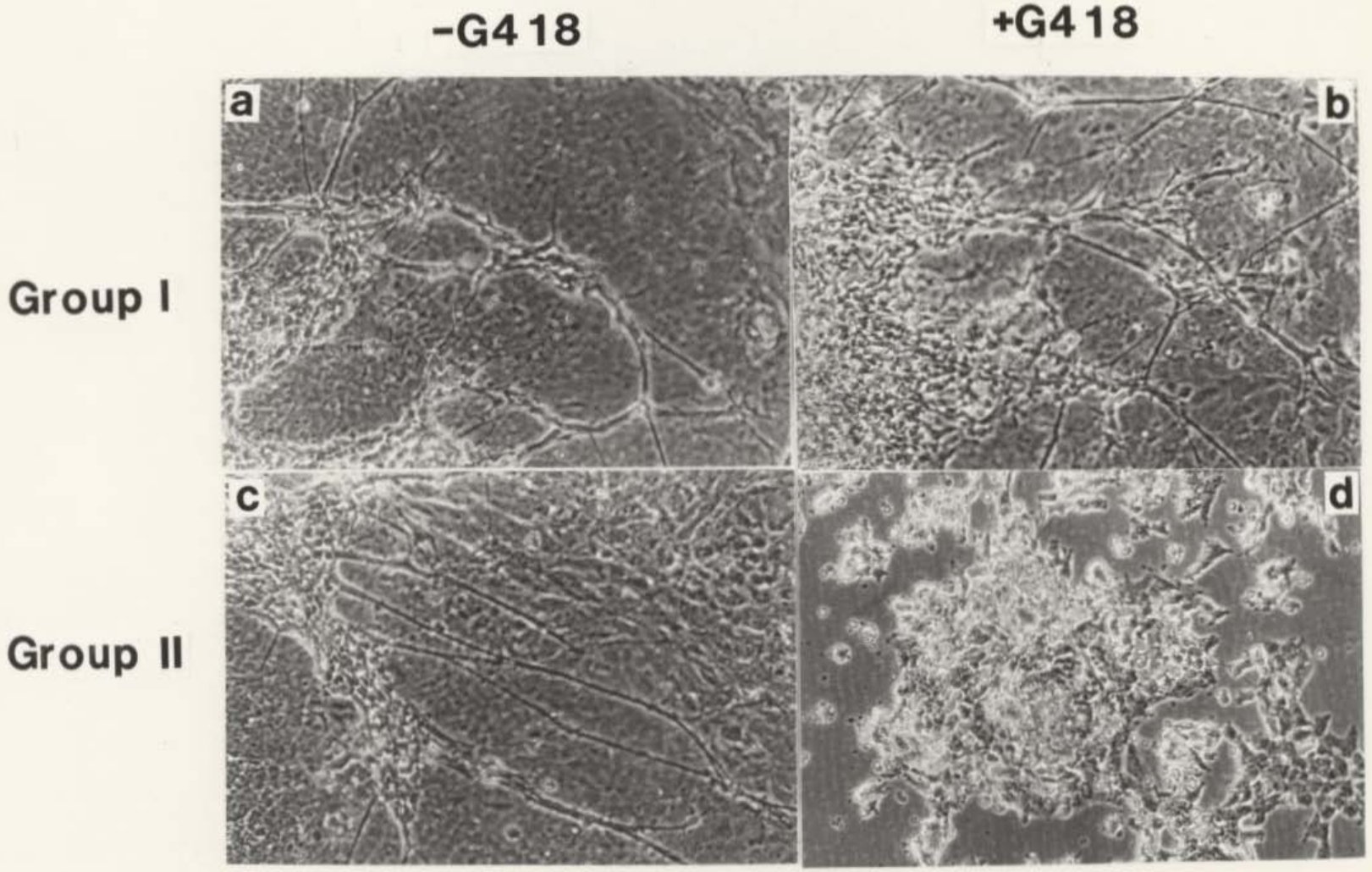
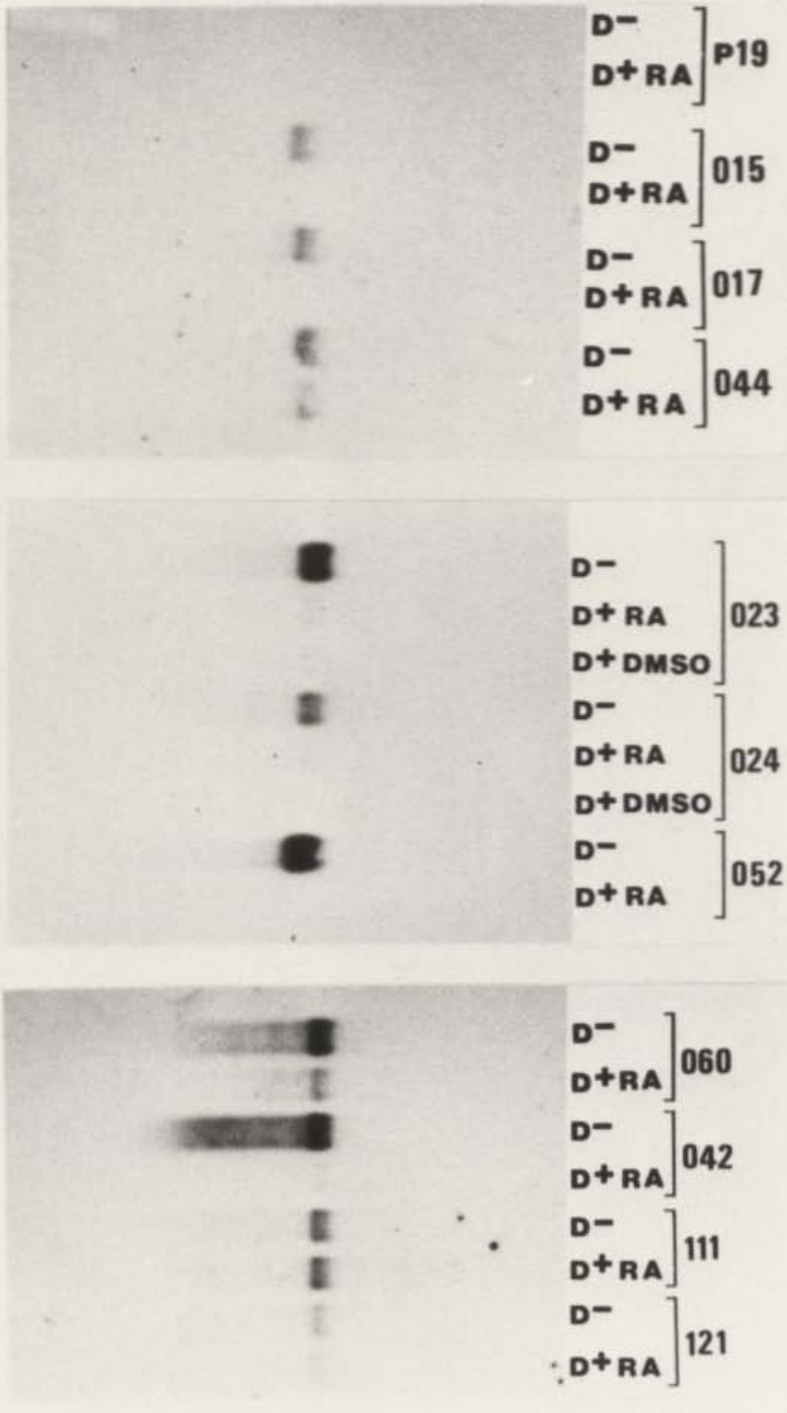

-G418
+G418
Group I
Group II
a
b
c
d

Figure II-5. Neo-R mRNA levels before and after differentiation of P19 cells.

Poly A+ RNA was prepared from undifferentiated (D-) cells grown in the medium without G418 for three to four days. Poly A+ RNA was also prepared from differentiated (D+) cells. Differentiation was induced by retinoic acid (RA) or in one cell line with dimethyl sulfoxide (DMSO). No G418 was present during differentiation. About 10 μg of poly A+ RNA was loaded on a formaldehyde-agarose gel, blotted, and hybridized to the Neo-R specific probe. The names of the cell lines are indicated at the top. P19 in the untransfected parental cell line. Cell lines 111 and 121 are transformed with pSV2neo. Cell line 044 belongs to group I, and cell lines 015, 017, 023, 024, 042, 052, and 060 belong to group II.


D−
D+RA
P19
D−
D+RA
015
D−
D+RA
017
D−
D+RA
044
D−
D+RA
D+DMSO
023
D−
D+RA
D+DMSO
024
D−
D+RA
052
D−
D+RA
060
D−
D+RA
042
D−
D+RA
111
D−
D+RA
121

transformed cell lines and the group I cell line (044, the levels of Neo-R mRNA were similar before (D-) or after (D+) differentiation. On the other hand, in most of the Group II cell lines (015, 017, 023, 042, 052), the Neo-R message was greatly reduced following cell differentiation. In the cell line 060, the decrease was only modest.

These results indicated that in most of the group II cell lines, the Neo-R gene is active when the cells are undifferentiated, but it is repressed following differentiation. Therefore, the Neo-R gene appears to be integrated into stem cell specific loci in these cell lines. The group II cell lines were hence subjected to molecular cloning.

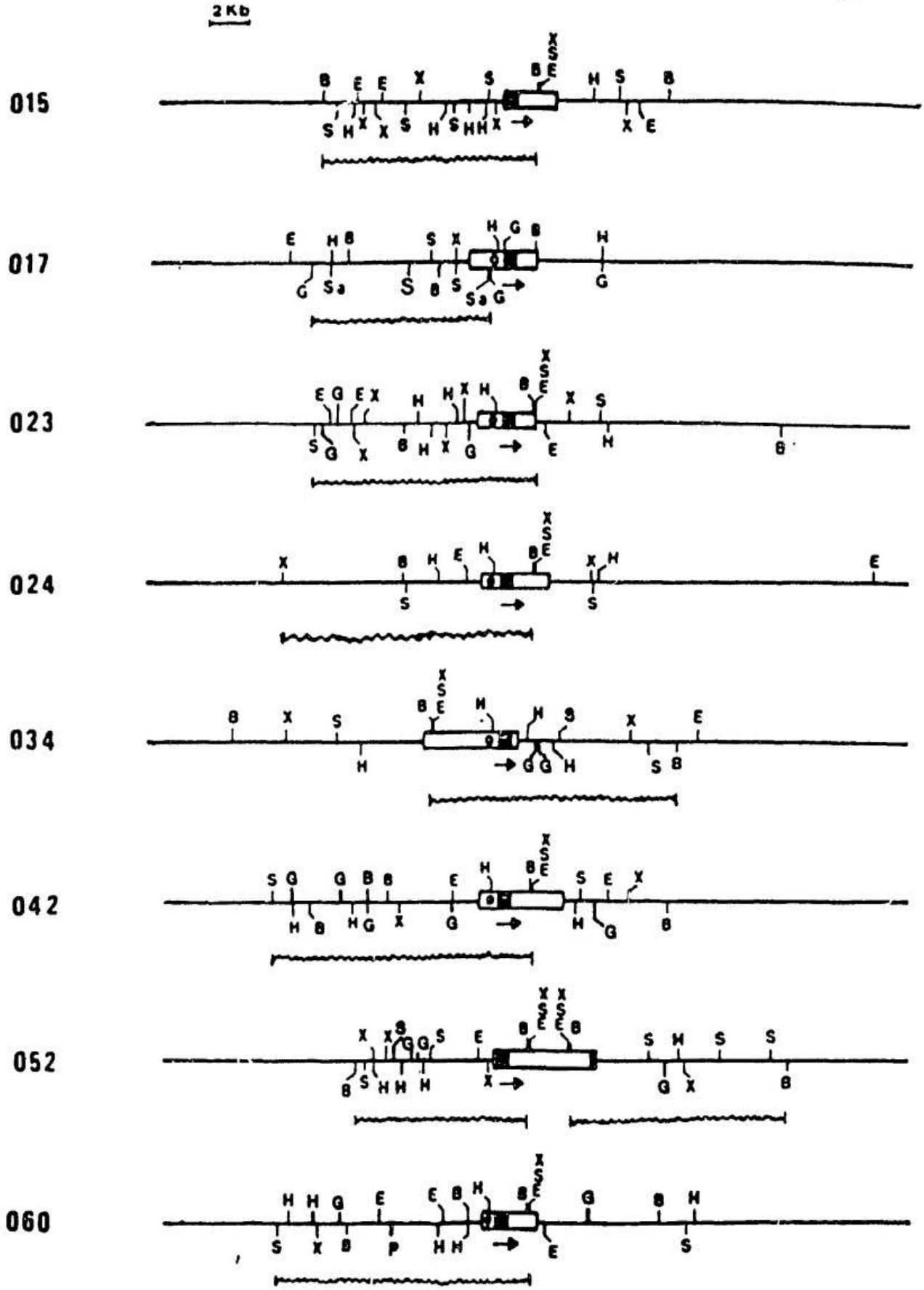
II.3.iv) Cloning of mouse DNA sequences flanking the Neo-R gene

In order to clone the mouse DNA sequences flanking the Neo-R gene, the restriction maps of the Neo-R integration sites of group II cell lines were determined by genomic Southern blotting (Figure II-6). As predicted by the dot hybridization analysis, all the group II cell lines contained a single copy of the Neo-R gene. However, 052 had one intact Neo-R gene and a partial (hence a non-functional) Neo-R gene.

The portion of the enhancer trap plasmid sequence remaining in the cell lines varied, but the *Hind*III-*Bam*HI 2.2 kb region that includes the Neo-R coding sequence and poly (A) addition signal sequence remained intact in five cell lines

Figure II-6. Cloning of mouse DNA sequences flanking the integrated Neo-R gene from group II cell lines.

Restriction maps were obtained with genomic DNA by Southern blotting and probing with the Neo-R specific prob , and with the cloned DNA by restriction and gel analysis. The names of cell lines are shown on the left. Symbols used are: wavy line, the nine cloned DNA fragments; straight line flanking the box, mouse DNA; open box, pA10neo sequences; closed box, Neo-R coding sequences; open circle, SV40 early promoter of pA10neo; arrow, direction of transcription. B, *Bam*HI; E. *Eco*RI; H, *Hind*III; G, *Bgl*II; S *Sst*I; X, *Xba*I.


2Kb
015
017
023
024
034
042
052
060

(017, 023, 024, 042 and 060). However, in two cell lines, 015 and 052, the *Hind*III site was deleted. This would mean that the SV40 early promoter must have been deleted in these two cell lines since it is located approximately 100 base pairs (bp) upstream of the *Hind*III site. Therefore, Neo-R gene transcription in these two cell lines must be initiated by an endogenous promoter or similar element. This notion is supported by the observation that in these two cell lines, the Neo-R mRNA appears to be smaller (see Figure II-5) compared to the Neo-R message transcribed from the SV40 early promoter.

Mouse DNA sequences flanking the Neo-R gene were cloned using the standard cloning procedures (see Material and Methods) from each of the eight group II cell lines (the wavy lines in figure II-6). Only one side of the integration site was cloned from seven cell lines (015, 017, 023, 024, 034, 042, 060) and both sides of the Neo-R gene were cloned from 052. The cloned phages were accordingly designated as λ015, λ017, λ023, λ024, λ042, λ051-1 (left side of the Neo-R gene), λ050-2 (right side), and λ060. The restriction sites of the cloned DNA were consistent with those previously determined by genomic Southern blotting, indicating that the nucleotide structure of the DNA sequences was not altered during the cloning procedures. Also, as predicted, the SV40 early promoter was deleted in λ015 and λ052 and was preserved in λ017, λ023, λ024, λ034, λ042 and λ060.

Each of the nine cloned fragments showed a distinct restriction map indicating that the integration of the Neo-R gene had taken place at independent sites, although it is still possible that some of the cloned fragments were derived from the same chromosomal locus.

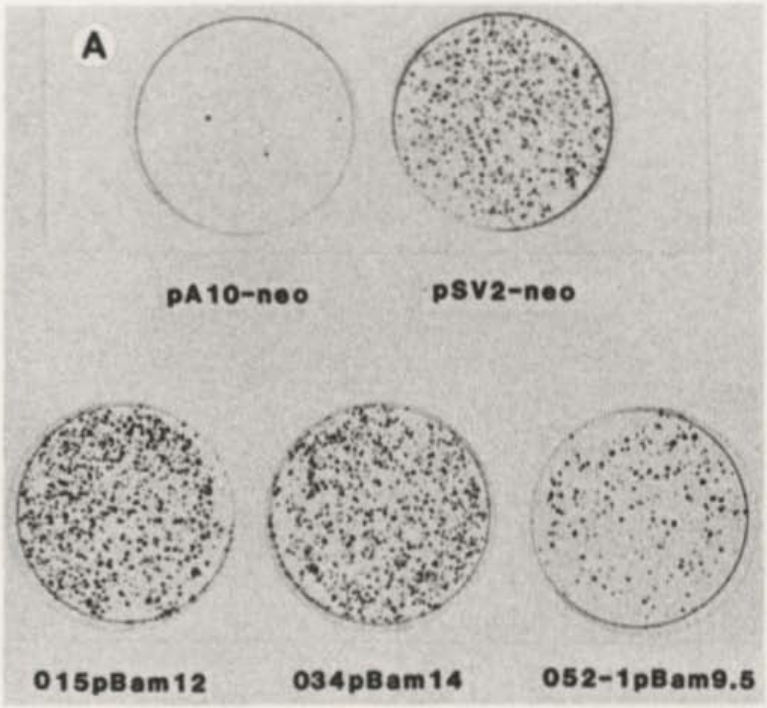
II.3.v) Cloned DNA fragments are transcriptionally active in the stem cells

Once the DNA was cloned from the group II cell lines, these DNA fragments were tested to determine whether or not they are transcriptionally active in stem cells. If the integrated Neo-R gene is activated by a regulatory element such as an enhancer, and if such an element is present in the cloned DNA, then the cloned fragments should possess a high transcriptional activity in the stem cells. Since the DNA fragments cloned in λ015, λ017, λ023, λ024, λ034, λ042, λ052-1, and λ060 contained the Neo-R gene, these fragments were conveniently subcloned in pUC (recombinant clones were selected on Kanamycin plates) for transfection. The plasmids were transfected to P19 cells and the activity was monitored by the stable transformation assay (Figure II-7A and Table 1, column 1). The enhancerless construct, pA10neo and the enhancerless-promoterless construct, p0-neo served as the negative controls and pSV2neo as the positive control. Both the negative controls showed very low transformation efficiencies. A few NEO^+ colonies were also observed with pUC 12 DNA, however they failed to grow after replating

Figure II-7. Activating enhancer elements are present in three of the cloned DNA sequences.

(A) The entire lengths of the inserts cloned in phage (the wavy line in Figure II-6) were subcloned at the corresponding sites of pUC12 and the plasmids were transfected to P19 cells. Their NEO^+ transforming efficiencies were determined (Materials and Methods). Representative Giemsa stained plates are shown here. pA10neo and pSV2neo are negative and positive controls, respectively. Three of the cloned DNAs, 015p*Bam*12, 034p*Bam*14 and 052-1p*Bam*9.5 (left of the enhancer-trap, see Figure II-6) showed a high NEO^+-transforming activity.

(B) Various deletion fragments of the 015, 034, and 052 cloned fragments were subcloned in pUC (shown as solid lines beneath the restriction maps), and the resulting deletion mutants were tested for NEO^+ transforming activity. The transformation activity is expressed as the number of NEO^+ colonies per 5×10^5 P19 cells. Symbols used are: closed circle, approximate locations of activating elements; open circle, SV40 early promoter or putative cellular promoter (in 015 and 052). For restriction site abbreviations, see legend to Figure 6.


A
pA10-neo
pSV2-neo
015pBam12
034pBam14
052-1pBam9.5

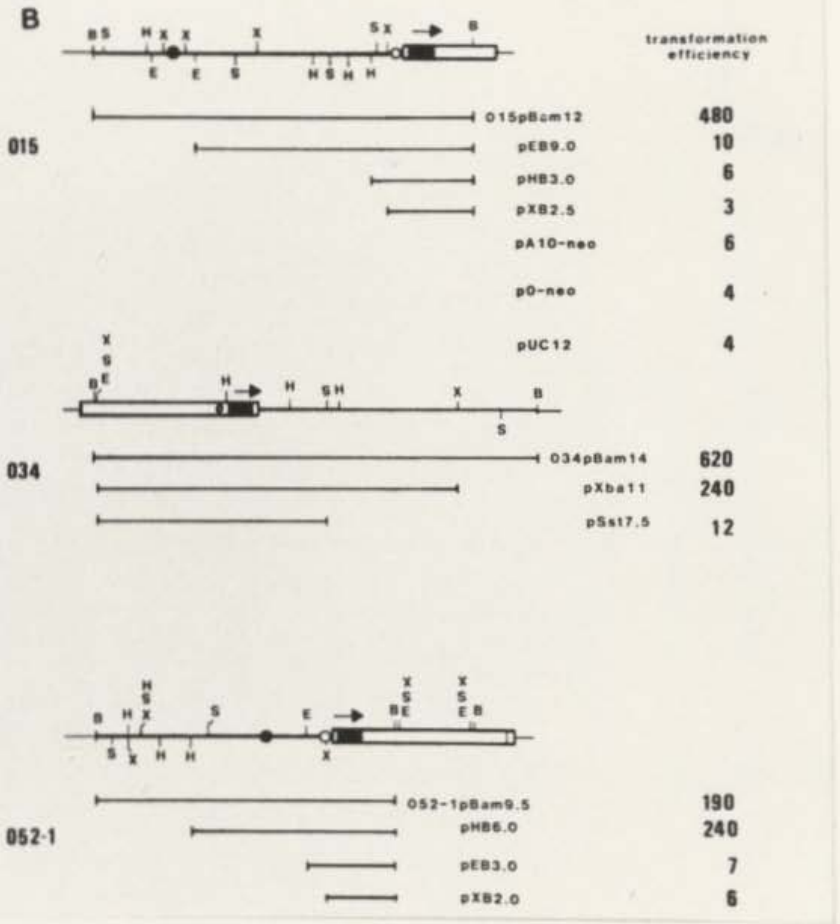

B
transformation efficiency
015
015pBam12 480
pEB9.0 10
pHB3.0 6
pXB2.5 3
pA10-neo 6
pO-neo 4
pUC12 4
034
034pBam14 620
pXba11 240
pSst7.5 12
052-1
052-1pBam9.5 190
pHB6.0 240
pEB3.0 7
pXB2.0 6

Table 1: NEO$^+$-transformation activity in P19 cells and NIH 3T3 cells

DNA	Transformation Frequency	
	P19	NIH 3T3
pOneo	4	5
pA10neo	6	18
015p*Bam*12	480	12
052-1p*Bam*9.5	240	15
034p*Bam*14	620	320
023p*Sst*13	5	N.D.
024p*Xba*I16	15	N.D
042p*Sst*14	20	N.D.
060p*Sst*13	8	N.D.
017p*Bgl*I10	N.D	N.D
pSV2neo	250	580

The various plasmids were transfected to P19 cells or NIH 3T3 cells and their NEO$^+$-transforming activities were determined as described in Materials and Methods. NEO$^+$-transformation frequency is expressed as the number of G418-resistant colonies / 5 x 10^5 cells. N.D., not determined.

(Hamada, personal communication); the pSV2neo, as expected, showed a high transformation frequency. Among the seven cloned DNA fragments tested, three, 015p*Bam*12, 034p*Bam*14, and 052-1p*Bam*9.5, showed very high transformation frequencies, implying that they contained regulatory elements that can activate the Neo-R gene expression. The remaining four fragments, 023p*Sst*13 (i.e., a 13-kb *Sst*I fragment from 023 NEO$^+$ cell line subcloned in a plasmid), 024p*Xba*16, 042p*Sst*14, and 060p*Sst*13 showed efficiencies similar to that of pA10neo.

To locate the regulatory element responsible for Neo-R gene activation, various deletion mutants were constructed from 015p*Bam*12, 034p*Bam*15, and 052-1p*Bam*9.5. Their activities were determined by the stable transformation assay (Figure II-7B). In the case of 015 deletion constructs, the activity was lost in pEB9.0, pHB3.0, and pXB2.5, indicating that the element is located in the left 3 kb region. As described above, the original SV40 early promoter was deleted in 015. Therefore, the Neo-R transcription must have been initiated by an endogenous promoter sequence located very close to the integration site. However, the results of the deletion-construct transformation experiments suggest that the high level Neo-R expression is due to an enhancer element located in the left 3 kb region.

The results of 052-1 deletion mutants suggest a similar conclusion. pHB6.0 retained full activity, however, pEB3.0 lost its activity. This suggests that the enhancer element in this locus is located in the *Hind*III-*Eco*RI 3 kb region. Also in this cell line, the SV40 early promoter is initiated by an endogenous promoter sequence.

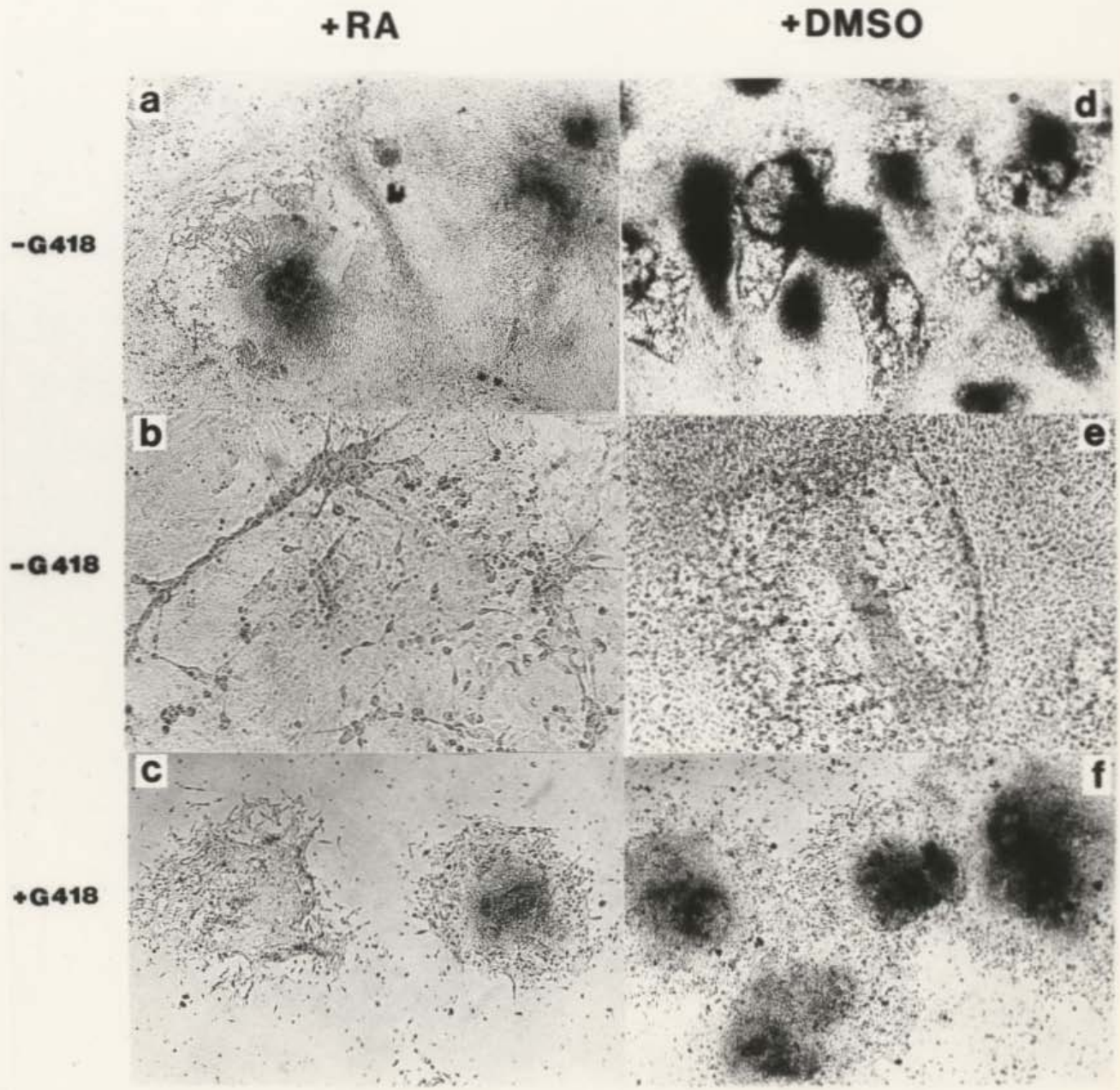
In the 034 locus, the SV40 early promoter was retained. It is of interest that, in the deletion analysis, the p*Xba*11 had retained 40% of the activity of 034p*Bam*14, while p*Sst*7.5 had completely lost the activity. These results suggest that one or multiple enhancers are present in the *Sst-Bam*HI 6.5 kb region.

II.3.vi) 015 and 052 cloned DNA fragments are inactive in differentiated cell types

Whether the activities of the 015, 034 and 052-1 clones are specific to the stem cells, was determined next. This was tested by two sets of experiments. The first was a differentiation assay. P19 cells were first transformed with 015p*Bam*12, 034p*Bam*14, and 052-1p*Bam*9.5 individually. The transformants were then induced to differentiate in the absence or presence of G418 (Figure II-8). Differentiation was induced by either RA or DMSO. In the absence of G418, 015p*Bam*12-transformed, 034p*Bam*14-transformed, and 052-1p*Bam*9.5-transformed cells could differentiate normally. They differentiated to neurons and glial cells with RA (Figure II-8, a and b), and to beating cardiac muscle cells with DMSO (d and e). However, in the presence of G418, 015 and 052-1 transformants could not complete differentiation. The aggregates did not grow large, and no obvious neurons (Figure II-8, c) or muscle cells (f) were observed. This observation suggests that the Neo-R gene is inactivated following differentiation. In contrast, the 034p*Bam*14-transformants differentiated normally in the presence of G418 (data not shown).

Figure II-8. The 015 and 052 enhancer elements are inactivated during differentiation.

NEO^+ colonies (200 to 500) transformed with 015p*Bam*12, 034p*Bam*14 or 052-1p*Bam*9.5 were pooled and maintained in the selective medium for four days. Each pool was then induced to differentiate by RA (a, b, and c) and by DMSO (d, e, and f) in the absence (a, b, d, and e) or presence (c and f) of G418. Phase-contrast photographs were taken eight days after induction. The photographs shown here are of 015p*Bam*12-transformed cells. The cells transformed with 052-1p*Bam*9.5 also showed a similar pattern. The cells transformed with 034p*Bam*14, however, differentiated normally in the absence as well as in the presence of G418 (data not shown, see Discussion).


+RA
+DMSO
-G418
-G418
+G418
a
b
c
d
e
f

In the second assay, the activities of the three cloned DNAs were determined in NIH 3T3 cells, a mouse fibroblast cell line (Table 1, column 2). The positive control pSV2neo showed a high NEO^+ transforming activity and pA10neo and p0-neo (negative controls) showed only background activities. Supporting the previous differentiation assay result, 015 and 052-1 were inactive in NIH 3T3 cells, while 034 was active.

These two lines of evidence indicate that the cloned DNA from the 015 and 052 loci were active only in EC stem cells and inactive in the differentiated cell types. Despite the fact that 034 locus was stem cell specific in supporting the Neo-R gene activity, the cloned DNA from this locus was active in both the stem cells as well as in the differentiated cell types (see Discussion).

II.3.vii) 015, 052 and 034 loci contain an early Transposon-like element

In 015 and 052, the SV40 early promoter was replaced by an endogenous promoter near the integration site. This observation was pleasing because, this would mean that there must be cellular genes associated with these promoters in the normal cells. Therefore, the nucleotide sequence of these promoters from 015 and 052 were determined (Hamada, personal communication). In both loci, the promoter sequence at the integration sites were the same. These sequences were next analysed by the Micro Genie computer program. The analysis showed that

these sequences belong to the early transposon-like elements (ETn). The ETn is a gene family of moderately repeated DNA sequences dispersed in the mouse genome (Brulet *et al.* 1983), the transcription of which is developmentally regulated (Brulet *et al.* 1983; Ikuma *et al.* 1986).

Whether the ETn genes in these loci were transposed DNA sequences, or endogenous to these loci, was then determined. This was done by cloning the pre-insertion region of the chromosome corresponding to these loci and determining the restriction map of the region. The results of these analyses showed that, in addition to 015 and 052 loci, 034 locus also contained a partial copy of the transposon extending up to the *Xba*I site (see 034 restriction map in Figure II-7B) indicating that the enhancer sequence found in this locus is only partly derived from the ETn and partly from the mouse sequence. The details of these experiments and results are given in Chapter IV of this thesis. In the remaining five loci (017, 023, 024, 042, and 060), however, transposition had not occured (see Results section, Chapter III).

II.4) DISCUSSION

The aim of this work was to isolate chromosomal loci that are specifically active in the EC stem cells, and to search those loci for the presence of genes and regulatory elements. The foregoing results demonstrate that several mouse

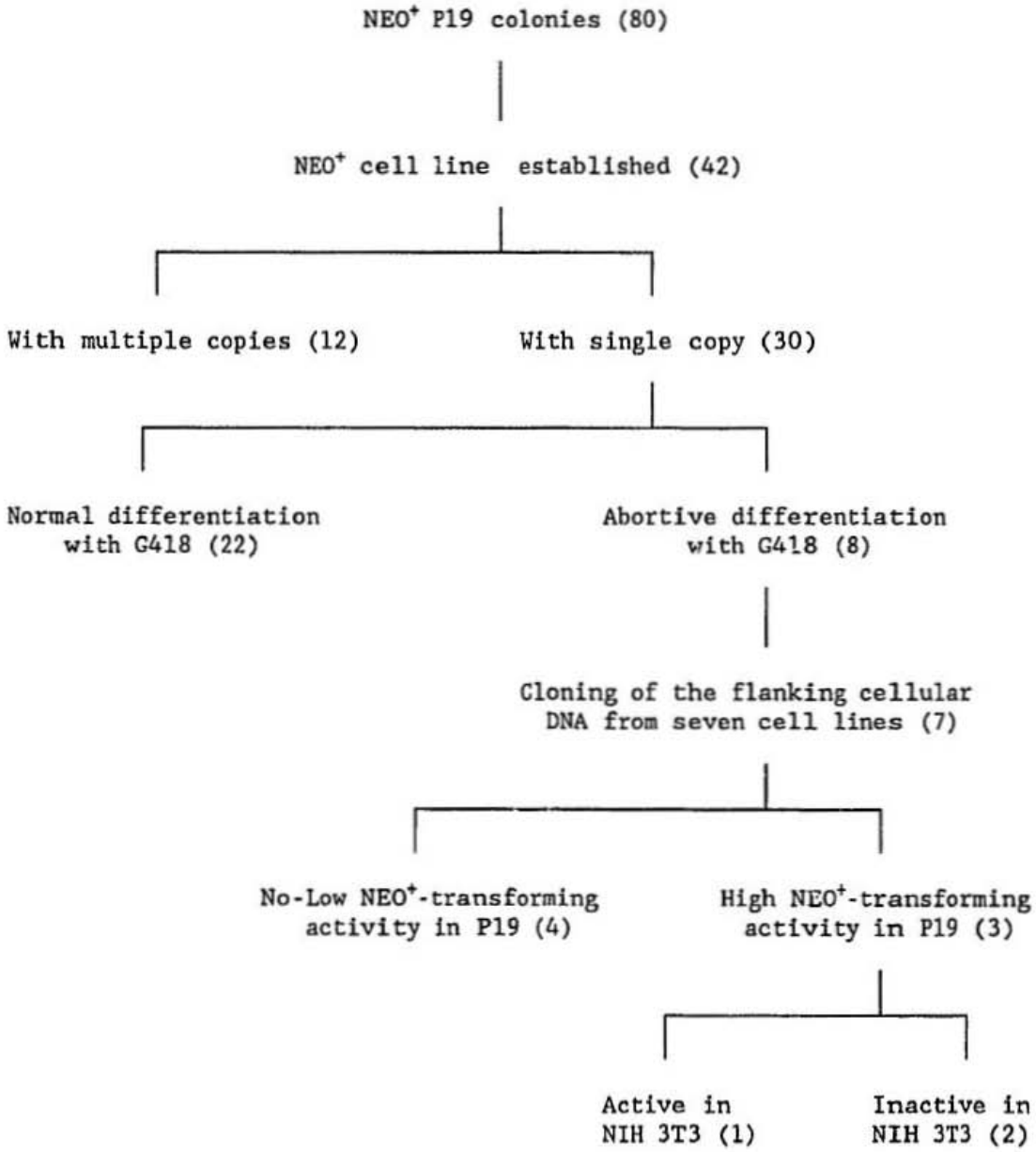
chromsomal loci and transcriptional regulatory elements (see below) that are specifically active in the EC stem cells have been cloned.

The most critical requirement in this procedure is that the linked promoter of the transgene should be enhancer dependent in a given cell line. Fortunately, the SV40 early promoter used in this study was enhancer dependent in P19 cells. This was unexpected because others have shown that the same promoter is enhancer independent in other EC cell lines such as F9 and PCC4 (Gorman *et al.* 1985). This difference could very well be due to the cell lines used. Heterogeneity between EC cell lines has been well documented and is thought to be reflected in the range of ability for differentiation both in culture (Rudnicki and McBurney, 1987) and in chimaeric association with a host embryo. Furthermore, some EC cell lines including F9, are reported to have adenovirus Ela-like transacting activity (Hen *et al.* 1986; Imperiale *et al.* 1984). As suggested by Gorman *et al.* (1985), it may be the same factor that renders the early promoter of SV40 enhancer independent in such cell lines and which may be absent in P19 cells. Another possibility could be the difference in the construction of enhancerless promoter, i.e., in this case (pA10neo), the deletion of the promoter is more extensive compared to Gorman's construct.

A summary of the results is given in Figure II-9. Among the 30 NEO^+ transformed cell lines, eight cell lines showed an abnormal differentiation pattern in the presence of G418 (Figures II-4 and II-5). The cloned DNA from three of these cell lines contained regulatory elements. The results also showed that the

Figure II-9. Summary of isolation and analysis of stem cell specific loci.

The number of cell lines obtained is shown in the parenthesis (see text for details).


NEO+ P19 colonies (80)
NEO+ cell line established (42)
With multiple copies (12)
With single copy (30)
Normal differentiation with G418 (22)
Abortive differentiation with G418 (8)
Cloning of the flanking cellular DNA from seven cell lines (7)
No-Low NEO+-transforming activity in P19 (4)
High NEO+-transforming activity in P19 (3)
Active in NIH 3T3 (1)
Inactive in NIH 3T3 (2)

stem cell specific enhancer sequences found in two loci (see below for explanation on the third one) were part of the ETn element and it was these enhancers that conferred stem cell specificity to the Neo-R expression in these loci. Although the biological role of the ETn gene product during early development remains to be determined, the uncovering of the stem cell specific enhancer in the ETn in this study will help to elucidate the mechanism of its developmental regulation.

Following the Northern blotting analysis of neo mRNA from the various NEO+ cell lines, the blot was not reprobed with a probe for some constitutively expressing gene; this to serve as an internal control for the quantity of poly A+ RNA loaded in D- and D+ lanes. While it would have been better if we had this data, it was not absolutely necessary. The reasons for this are as follows: first, the selection procedure used was stringent, i.e., if there is residual Neo-R gene expression in the cells G418 resistance is conferred. Therefore, inability of cells to complete differentiation in the presence of the drug in the differentiation assay acurately reflects the inactivation of the Neo-R gene following differentiation. The results of the differentiation assay was entirely supported by the Northern analysis and no conflicting data in the two assays were observed. Second, the difference in the Neo-R message levels before and after differentiation was very significant and this difference could not be explained by a difference in the amount of poly A+ RNA loaded.

The case of 034 was clearly different. That is, the original cell line selected was stem cell specific in supporting the Neo-R gene activity. However, the cloned DNA from the 034 locus was not stem cell specific; it was active in differentiated P19 cells as well as in NIH 3T3 cells (Table 1). This locus contained a partial copy of the ETn with its enhancer sequences (up to *Xba*I site, see Figure II-7B). The enhancer activity in this locus is contributed partly by the ETn enhancer and partly by enhancer sequences of cellular origin. Assuming that this ETn enhancer is responsible for the stem cell specificity of this locus (as in 015 and 052 loci), the question arises as to why the cloned DNA that included both enhancers is not stem cell specific. One obvious explanation is that the other half of the enhancer activity contributed by mouse sequence in this locus is not specific to stem cells only, but is also active in differentiated tissue types. Again, the question arises as to why then it did not support the activity of the Neo-R gene in the original locus. One possibility to reconcile these results is that, in addition to the activating elements (one cellular and the other ETn derived), there is a repressor element near the integration site which is not included in the cloned DNA. This repressor element would be inactive in stem cells, allowing the two enhancer elements to function. When the cells are differentiated, the ETn enhancer being stem cell specific is now inactive, but the now active repressor element interferes with the cellular enhancer. The transformation efficiencies of 034p*Bam*14 in P19 stem cells and in NIH 3T3 cells, and the 034p*Bam*14 deletion analysis indicates that the cellular enhancer is

active throughout differentiation (compare lines 1 and 2 in Table 1) and it augments the ETn enhancer in stem cells by 50% to 60%, (compare lines 1 and 2 in Table 1; see 034 deletion analysis results in Figure II-7B). If the cloned DNA did not contain this repressor element, then the stem cell specificity of the enhancer would automatically be lost.

In the remaining five loci, 017, 023, 024, 042, and 060, ETn gene transposition had not occured. This conclusion is based mainly on the basis of restriction maps of the above five loci, which are different from that of ETn. That is, the restriction sites - the closely spaced *Hind*III-*Bgl*II-*Bgl*II-*Sst*I sites which were conserved in ETn-like genes were absent in other loci (see Fig. II-6). The restriction maps of the above loci were also different from that of intracisternal A-particle genes (Ono *et al.* 1980), a group of endogenous proretrovirus-like elements (Chase and Piko, 1973). Secondly, the analysis of the pre-insertion regions of the chromosome corresponding to these five loci by restriction mapping revealed no transposition (see Chapter III).

There have been a few attempts by others to isolate stem cell specific loci or enhancers from EC cells (Barklis *et al.* 1986; Taketo and Tanaka, 1987). The approach was to use the Neo-R as a marker gene linked to the murine leukaemia virus LTR. Since the murine leukaemia virus LTR is repressed in EC cells (Gorman *et al.* 1985), such a gene can be used as a trap (like the pA10neo used in this study) for selecting elements that can reactivate the LTR. However, this

method cannot select stem-cell specific elements as the LTR is active in differentiated cells. In both the studies, the enhancer-like elements selected were able to reactivate transcription from the viral LTR in EC stem cells, but their specificity was not determined (see Barklis *et al.* 1986; Takeo and Tanaka, 1987). In this context, it should be noted that the loci selected in this study were different. Furthermore, the restriction maps of the loci isolated by Barklis et al (1987) and Takeo and Tanaka (1987) and those isolated in this study are different.

The cloned DNA from 023, 024, 042, and 060 did not possess high NEO^+ transforming activity. One obvious explanation for this observation is that the activating element is located downstream (only the upstream regions were cloned from these cell lines) or further upstream from the cloned DNA. Alternatively, the Neo-R activation could be due to other mechanisms. Allen *et al.* (1988) noted that an integrated transgene (in this case, the Neo-R gene) is subjected to at least three kinds of effect in the cell. Firstly, the expression can be regulated by cis-acting DNA sequences such as enhancers, exemplified by the 015, 034 and 052 loci. Secondly, the transgene could be integrated directly into a cellular gene, with the consequence of transcriptional read-through into the transgene. Another situation would be similar to the case of β-globin gene expression where the entire locus is flanked by the so called locus activation domains (Grosveld *et al.* 1987). Whatever mechanism it may be, it should be noted that the mechanism(s) regulating the Neo-R gene expression in the above five loci cloned in this study are specific to the stem cells.

CHAPTER III

GENES IN THE STEM-CELL SPECIFIC LOCI

III.1) INTRODUCTION

In an interesting paper, Rosenstraus and Levine (1979) reported that when nullipotent F9 EC cells are cultured in mixed aggregates with pluripotent PSA-1 cells, the nullipotent cells inhibited the ability of the pluripotent cells to differentiate beyond the endoderm stage. This observation was later extended by Littlefield and Felix (1982) who showed that somatic cell hybrids between EC stem cells and their retinoic acid induced derivatives have the stem cell phenotype. Similar results were also found with EC stem cell x rat hepatoma hybrid cells (Wray and Jaeckle, 1983) suggesting that certain factor(s) in the undifferentiated EC cells suppress the differentiation phenotype of the hybrids. Based on these studies, Levine et al (1984) proposed an interesting hypothesis, that the differentiation of stem cells may be a reductive process, namely certain gene(s) are expressed specifically in the stem cell stage of the EC cells, and their products are responsible for maintaining the undifferentiated state of the stem cells. When these genes are switched off, the cells are allowed to differentiate into lineages. Furthermore, all EC cell lines without exception, are malignant and many of them are pluripotent. However, their

differentiated derivatives are neither pluripotent nor malignant. Therefore, Levine *et al.* (1984) further proposed that the expression of that class of genes which might control the properties of both pluripotency and malignancy is suppressed during differentiation.

Previous work by others indicated that there are several different genes whose expression is limited to stem cells and not in their differentiated derivatives. Examples include heat shock genes (Bensaude and Morange, 1983; Levine *et al.* 1984), antigenic determinants (Solter *et al.* 1979), and growth factors (Gudas *et al.* 1983). None of the genes in the above categories, however, appear to be the candidates for genes in the reductive pathway. Interestingly, there has been some work on the kinetics of myc gene expression and cell differentiation. Down regulation of myc expression also accompanies RA-induced differentiation (Westin *et al.* 1982), and DMSO-induced murine MEL cell differentiation (Lachman and Skoultchi, 1984). For instance, in MEL cells, constitutive expression of myc by the use of surrogate promoters inhibits the ability of DMSO to induce differentiation (Coppola and Cole, 1986; Prochownik and Kukowska, 1986). Furthermore, in F9 cells, expression of antisense myc sequences induced differentiation (Griep and Westphal, 1988). These results suggest that down regulation of myc is necessary for differentiation. This, in fact, brings credence to the idea of a reductive mechanism in cell differentiation. The myc gene is possibly one of the genes of the reductive pathway.

Although the observation that RA induces differentiation of EC stem cells has stimulated research aimed at understanding the molecular mechanisms underlying differentiation, at present very little is known about the early events associated with RA induction. There is evidence that a complex of RA and its receptor protein is translocated to the nucleus where it indirectly or directly induces new gene activity (Jetten and Jetten, 1979). There is also evidence that DNA binding domains of the RA and thyroid hormone receptors are 62% identical in their amino acid sequences (Giguere *et al.* 1987) and that RA and thyroid hormone induce gene expression through a common responsive element (Umesono *et al.* 1988).

With regard to the role of RA in stem cell differentiation, however, there is also the question of competence. That is, F9 cells differentiate into parietal endoderm with RA and dibutyryl cAMP (Strickland *et al.* 1980) while P19 cells differentiate into neurons (McBurney *et al.* 1982). It appears that in these two cases, a different set of genes are competent to respond to RA, reflected in their differences in developmental potential or plasticity. However, one common phenomenon which may be happening in these cell lines is that a set of genes are switched off, allowing the cells to go through a determination pathway and differentiating into lineages. Therefore, isolation of this class of genes would represent a major advance in understanding the phenomenon of cell determination and differentiation.

Previously, there have been a few unsuccessful attempts to isolate genes from EC cells whose expression is entirely restricted to stem cell stage (Levine *et al.* 1984; Ikuma *et al.* 1986). Here an attempt was made to clone such genes. Using a pA10neo transgene as a selectable genetic marker, chromosomal loci that are specifically active in the undifferentiated EC stem cells were first isolated (Chapter II). The preinsertion chromosomal regions corresponding to these stem cell specific loci were cloned and the DNAs were searched for the presence of genes by the Northern blot analysis. The results of these experiments are given below.

III.2) MATERIALS AND METHODS

The following procedures were used for the mapping and cloning of single copy sequences from preinsertion regions corresponding to the stem cell specific loci.

III.2.i) Single Copy Mapping and Isolation

For single copy mapping, the Neo-R gene plus the flanking cellular sequences cloned previously in phage (Chapter II), were digested with a combination of enzymes, so that the cellular DNA was restricted into smaller fragments. These digests were electrophoresed on a 0.7% agarose gel followed by Southern blotting

as described in Chapter II. The blots were hybridized with nick translated P19 genomic DNA (the genomic DNA was digested with *Hind*III before nick translation). After washing, the blots were exposed for one week. Those restriction fragments which fail to light up in the autoradiogram are presumably present in one to a few copies in the genome. Such fragments are identified and they were gel purified and subcloned into plasmids. These were next confirmed with respect to their single copy nature by genomic Southerns as follows. P19 genomic DNA was restriction digested with an appropriate enzyme, blotted onto nitrocellulose and probed with the labelled "to be confirmed" single copy subclone. Only those which are present as a single copy per haploid genome were used in isolating the pre-insertion loci and in chromosomal walking.

III.2.ii)Construction of *Sau* 3A partically digested Genomic library

In order to isolate pre-insertion loci and for use in genome walking, a *Sau* 3A partially digested P19 genomic library was constructed in the phage vector λEMBL 3A.

Initially, conditions for partial digestion of high molecular weight P19 DNA were established in pilot reactions as described by Maniatis *et al.* (1982). From this digestion pattern, the amount of enzyme needed to produce the maximum number of fragments in the 16-20 kb range was ascertained (as judged from the

intensity of fluorescence in the gel of the pilot reactions). In large scale preparative digestions, 150 μg of P19 genomic DNA was digested with *Sau*3A using the optimised conditions (see Figure III-1A). After stopping the reaction by addition of EDTA to a final concentration of 20 mM, an aliquot from each reaction was electrophoresed (Figure III-1A). In the meantime, the DNA was phenol:chloroform extracted and ethanol precipitated. Those reactions, which produced fragments in the 16-20 kb range (see Figure III-1A, Rx 1 to 6), were pooled and electrophoresed on a 0.5% preparative agarose gel. The DNA from the 16-20 kb region was purified from the gel (Figure III-1B) and prepared for ligation into *Bam*HI-*Eco*RI digested λEMBL 3A phage arms as described in Chapter II (section II.2.x). For ligation, DNA fragments from M1, M2, and M3 in Figure III- 1B were pooled and used. The ligation and *in vitro* packaging (Stratagene) were done same as described for the restricted library in Chapter II, section II.2.x.

III.2.iii) Isolation of pre-insertion genomic loci

Genomic clones from the library were isolated by plaque hybridization (Chapter II). Single-copy probes from each of the loci were nick translated, pooled and about 1.5 million phages from the genomic library were screened with the pooled probes. The positive clones were purified individually and the clones were assigned to different loci by hybridizing them individually to probes from each loci.

Figure III-1. Construction of *Sau* 3A-partially digested genomic library.

High molecular weight P19 genomic DNA (150 μg per reaction) was digested with various amounts of *Sau* 3A restriction enzyme (10-times scaled-up of pilot reactions which produced maximum overlapping intensities from 16 kb to 22 kb) for 1 hr at 37 °C. The reactions were stopped by addition of EDTA to final 20 mM concentration. An aliquot from each reaction (Rx 1 to Rx 6 in A) was run on a gel. In the meantime, the remaining DNA was phenol:chloroform extracted and ethanol precipitated. Those reactions which produced maximum fragments between 16 kb and 22 kb were pooled (in this case, Rx 1 to Rx 6), and electrophoresed on a 0.5% preparative agarose gel. The amount of enzyme used was: Rx 1, 3.5 U; Rx 2, 5.24 U; Rx 3, 7 U; Rx 4, 8.74 U; Rx 5, 10.5 U; Rx 6, 12.23 U, per 150 ug of DNA in 10 mL of reaction volume.

(B) Fragments from the above preparative gel were exised in four fractions ranging from 16 kb to 23 kb. They were purified, ethanol precipitated twice and an aliquot (1/10th of the total) from each fraction was run on a gel along side with the *Bam*HI + *Eco*RI digested phage vector EMBL 3A. T, top fraction; M1, M2 and M3 are successive middle fractions; B, bottom fraction. For ligation, M1, M2, and M3 were pooled and 1/3rd of the pooled DNA (about 7 μg) was ligated to EMBL 3A phage arms in a 1:1 molar ratio. The remaining procedures were the same as given for the construction of the restricted genomic library (Chapter II, section II.2.x).

A
Marker
Rx.6
Rx.5
Rx.4
Rx.3
Rx.2
Rx.1
Marker
-23.3
-9.4
-6.6
-4.3
-2.3
-2.0
X kb
B
Phage arms
Marker
B
M3
M2
M1
T
Marker
-23.3
-9.4
-6.6
-4.3
-2.3
-2.0
X kb

Amplification, large scale phage culture, phage DNA preparation and restriction mapping was done as described in Chapter II, section II.2.x.

III.2.iv) Northern blot analysis

The total RNA extractions from the P19 D- and D+ cells were done by H. Hamada. Poly A+ selection was done as described in Maniatis et al. (1982).

In order to examine the presence of protein coding sequences (genes) in the stem cell specific loci (Neo-R pre-insertion loci), single-copy sequences were isolated from the cloned pre-insertion loci DNA as described in III.2.i. Next, poly A+ RNA from the undifferentiated P19 cells was prepared, electrophoresed on a formaldehyde containing 1% agarose gel and transferred to a nitrocellulose filter. All the single-copy sequences from one locus were combined and labelled by the random primer method and were hybridized to the above RNA biot. Hybridization and washing conditions were the same as described in Chapter II (section II.2.vii). The blots were exposed for three to four weeks.

III.3) RESULTS

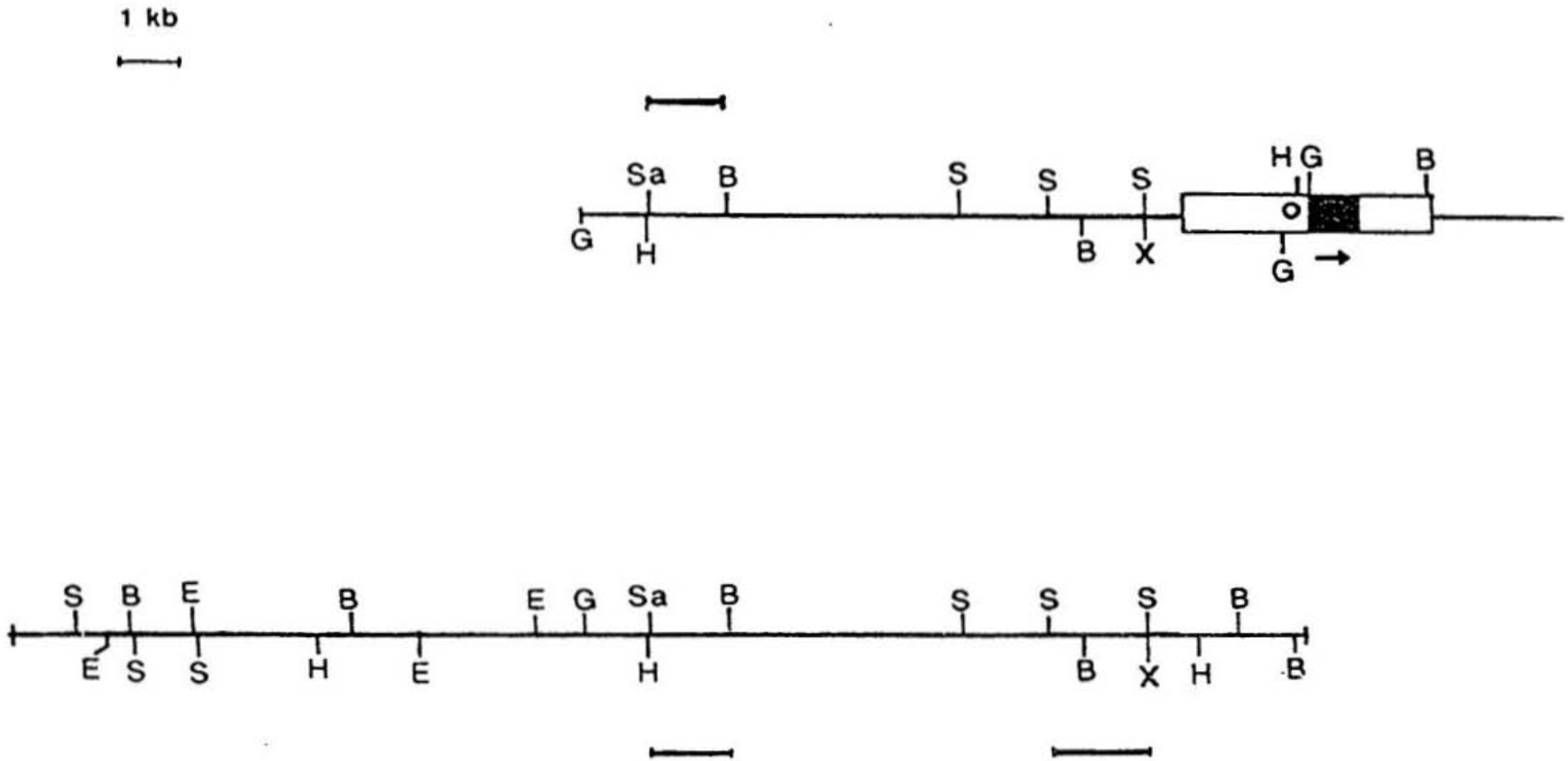
III.3.i) Isolation of the preinsertion chromosomal regions corresponding to the stem cell specific loci

Single-copy DNA probes subcloned from the previously cloned DNA (in phage) from the NEO$^+$-cell lines (Chapter II) were used to isolate the preinsertion regions corresponding to the NEO$^+$ stem cell specific loci. No preinsertion chromosomal region was cloned from the 024 locus due to the lack of a single-copy probe that could be used to screen a genomic library. From each of the seven loci, between 20 to 30 kb of DNA were cloned and subjected to restriction analysis. This revealed that in three of the loci, namely 015, 034, and 052, there was chromosomal rearrangement at the Neo-R integration site. This was due to the insertion of an ETn genome during the NEO$^+$ cell line selection (details of these loci and the ETn insertion are described in Chapter IV of this thesis). However, there were no chromosomal rearrangements in the remaining four loci.

From the 017 locus, about 21 kb of DNA was cloned (see Figure III-2). Single-copy mapping of this locus revealed two stretches of single-copy sequences (see Figure III-2, the two lines below the restriction map), a *Sal*I-*Bam*HI 1.3 kb fragment (this sequence was used as a probe to isolate the preinsertion region) and a *Sst*I 1.6 kb fragment. These two sequences from this locus were subsequently used in Northern blotting

Figure III-2. Cloning, restriction mapping and single-copy mapping of the preinsertion locus corresponding to the 017 locus.

Preinsertion regions corresponding to the 017 NEO^+-locus were cloned as overlapping genomic clones from a *Sau* 3A-digested P19 genomic library (see Figure III-1, Materials and Methods). The upper drawing is the NEO^+ (Neo-R gene integrated)-locus described in Chapter II; the bottom drawing is of the preinsertion region of the chromosome at the Neo-R integration site, isolated using a single-copy DNA probe from the NEO^+-locus (the straight line shown above the NEO^+-locus restriction map, see text for details of single-copy mapping). The straight lines below the preinsertion-locus restriction map are single-copy DNA sequences used in the Northern blotting analysis of poly A+ RNA from undifferentiated P19 cells. Symbols used: open box, pA10neo; closed box, Neo-R coding sequence; open circle, SV40 early promoter; arrow, direction of Neo-R gene transcription; B, *Bam*HI; E, *Eco*RI; H, *Hind*III; S, *Sst*I; G, RgiII; K, *K11pn*I; X, *Xba*I.



LOCUS 017

analysis of poly A+ RNA from undifferentiated P19 cells to examine whether there are any mRNA coding sequences in this locus.

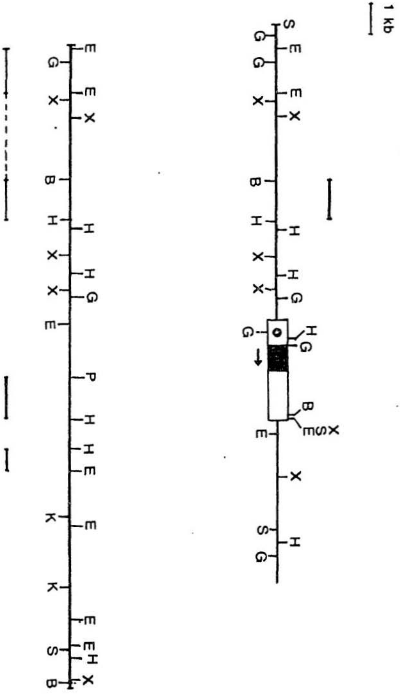
From the 023 locus, about 21 kb of DNA was cloned (Figure III-3). This locus had more single-copy sequences (the lines at the bottom of the restriction map in Figure III-3). The single-copy sequence *Bam*HI-*Hind*III 1.3 kb was used in the isolation of the preinsertion locus. In the Northern blotting analysis of this locus, all the single-copy sequences shown in the figure, as well as a probe containing repeated sequences (*Eco*RI-*Bam*HI 2.8 kb, dotted line in Figure III-3) were used.

The 034 locus is one of the three loci that contained the ETn genome (the thick line in Figure III-4). About 30 kb of DNA, corresponding to this locus, was cloned. Although this locus contained an enhancer element of the ETn origin, the presence of a cellular negative regulatory element which becomes functional upon differentiation of cells, has been speculated (see Discussion in Chapter II). Therefore, all the single-copy sequences, a total of 12.5 kb, were subcloned from this locus (shown as straight lines at the bottom of the restriction map, Figure III-4; the sequence used in isolating the preinsertion chromosome is the *Sst*I-*Bam*HI 1 kb fragment present in the middle region of the cloned DNA) and used in the Northern blot analysis.

From the 042 locus, about 27-kb of DNA was cloned using the single copy *Sst*I-*Hind*III 1.2-kb (Figure III-5) fragment as a probe. A total of 3-kb of single-

Figure III-3. Cloning, restriction mapping and single-copy mapping of the preinsertion locus corresponding to the 023 locus.

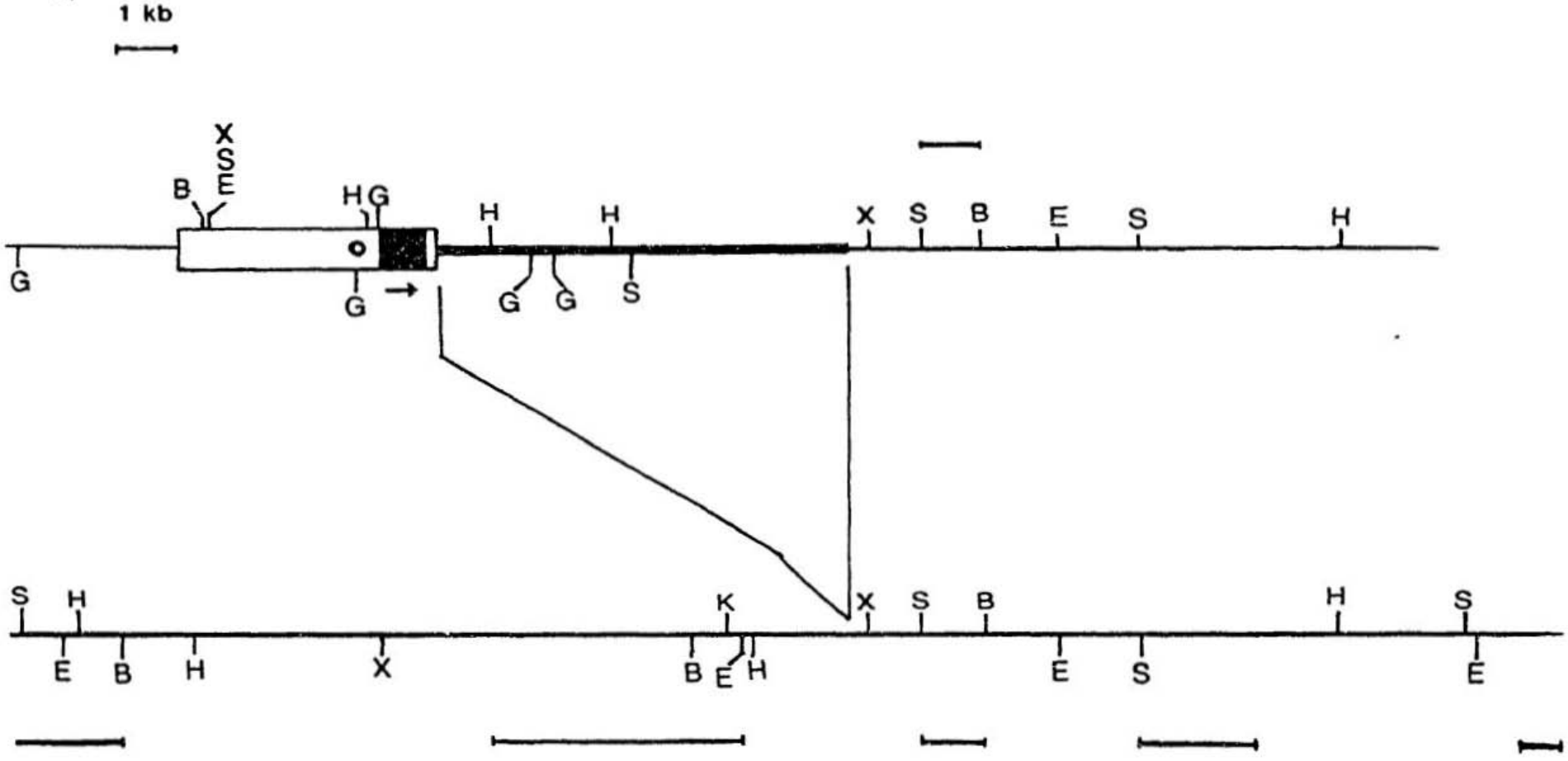
See legend to Figure III-2 for details.



LOCUS 023

Figure III-4. Cloning, restriction mapping and single-copy mapping of the preinsertion region corresponding to the 034 locus.

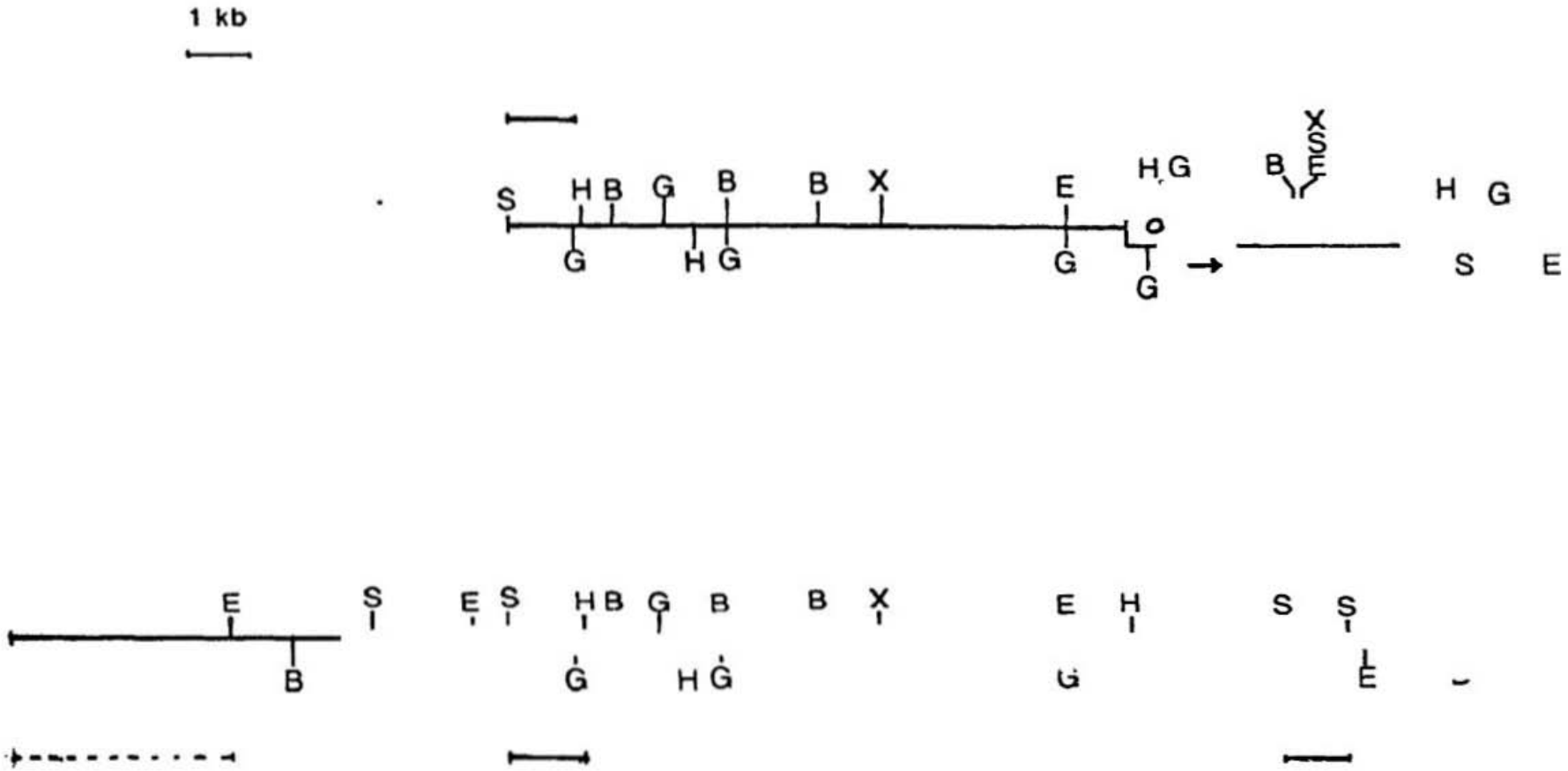
The thick line in the upper NEO^+-locus restriction map indicates an incomplete ETn genome, inserted into this locus along with the pA10neo (see Chapter IV for details). For symbols, restriction site abbreviations and other details, see legend to Figure III-2.



LOCUS 034

Figure III-5. Cloning, restriction mapping and single-copy mapping of the preinsertion region corresponding to the 042 locus.

See legend to Figure III-2 for details.



LOCUS 042

copy sequences, and 3.5-kb of DNA which contained repeated sequences in a few copies (dotted line in Figure III-5) from this locus were subcloned and used in the Northern blotting analysis.

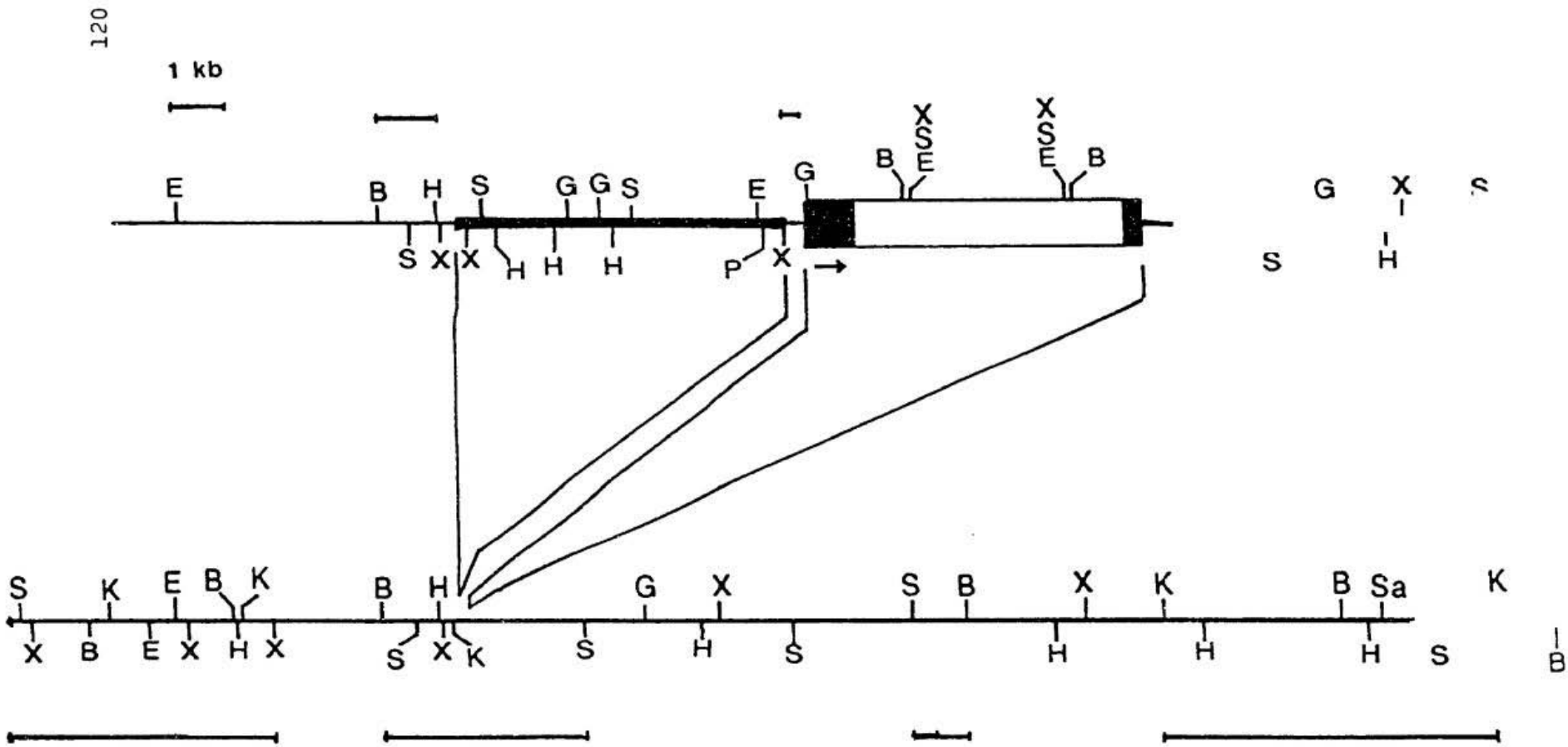
The 052 locus was one of the three loci which contained the ETn genome. Also, it was known that the stem cell specificity of this locus was conferred by the enhancer elements of the ETn. Integration of exogenous DNA sequences into the host chromosomes is believed to be favoured at an actively transcribing locus. Therefore, the available single-copy sequences (a total of 14.5-kb, see Figure III-6) in the cloned 30-kb from this locus was used in the Northern analysis. This locus is described in detail in Chapter IV.

A total of 26-kb of DNA was cloned corresponding to the 060 locus (Figure III-7) using the 0.7-kb single-copy probe, extending from the *Bam*HI-to the enhancer-trap plasmid. This locus contained only a total of 1.5-kb of single-copy sequences (see Figure III-7), which was used in the Northern analysis together with a probe containing some repeated sequences (*Hind*III-*Xba*I 1.6-kb, the dotted line in Figure III-6)

Since the 015 locus contained the ETn genome and the stem cell specificity of this locus was due to this ETn enhancer elements, single-copy mapping of this locus was not done. This locus is discussed in detail in Chapter IV. Also, it should be noted that except for the 052 locus, cellular sequences 3' of Neo-R integration site in cell lines 017, 023, 042, and 060 and 5' of Neo-R integration site in cell

Figure III-6. Cloning, restriction mapping and single-copy mapping of the preinsertion region corresponding to the 052 locus.

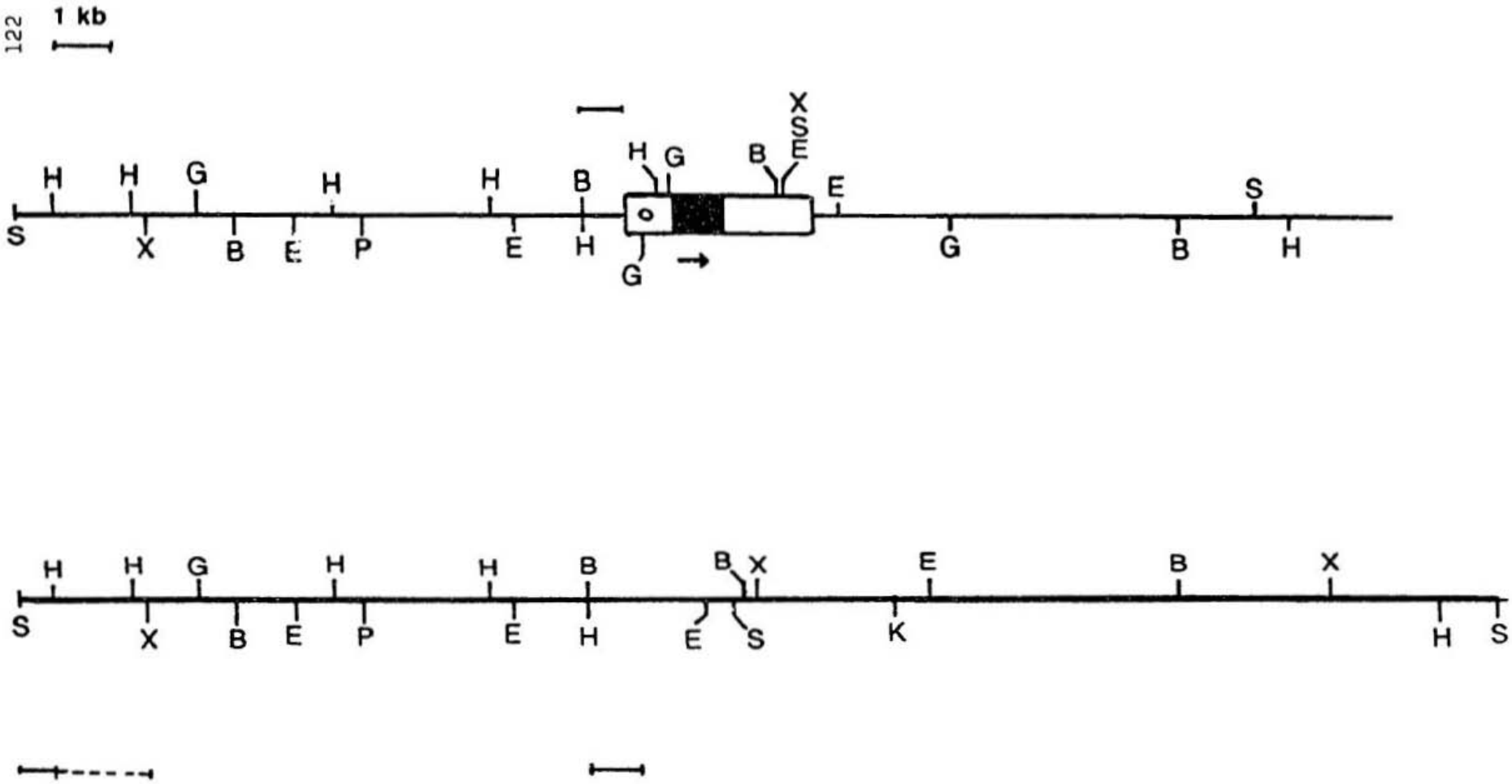
The analysis of this locus shows that an entire copy of the ETn genome (thick line in the upper map) was integrated at the NEO^+-locus. This locus, along with the 015 and 034, loci is discussed in detail in Chapter IV. For symbols and other details, see legend to Figure III-2.



LOCUS 052

Figure III-7. Cloning, restriction mapping and single-copy mapping of the preinsertion region corresponding to the 060 locus.

See legend to Figure III-2 for details.


1 kb
H H G H H B H G B X S E E S
S X B E P E H G G B H
H H G H H B B X E B X
S X B E P E H E S K H S

LOCUS 060

line 034 were not included in the cloned DNA (compare the restriction maps of NEO^+ loci and the respective preinsertion loci in the above figures).

III.3.ii)Northern blot analysis

All of the single-copy sequences available from the seven loci were subcloned in pUC 12 or pUC 18, and subcloned fragments were excised from the plasmid with appropriate enzymes, gel purified and used in the Northern analysis. In addition, considering the possibility that there might be genes in these loci which might belong to families of similar genes, sequences which are repeated a few times in the genome from some of the cloned loci (the dotted lines in the figures) were also used in the Northern analysis. The results of these Northern blots are shown in Figure III-8. There were no detectable messages against cloned DNAs from 017 locus (1st lane) and 052 locus (data not shown).

From the 023 locus, two types of probes were used: i) pooled single-copy sequences, ii) the *Eco*RI-*Bam*HI 2.8 kb fragment containing repeated sequences (the dotted line in Figure III-3). Against the single-copy probes, a band (arrow in Figure III-8) which corresponds to 28S ribosomal RNA (about 5 kb) was detected, along with a smear in the 7.5 kb region. Against the repeated sequences probe, several bands were detected (3rd lane), however, no bands apparently common to probes (i) and (ii) were detected in the blot (compare 2nd and 3rd lane).

Figure III-8. Northern blot analysis of the 017, 023, 034, 042, and 060 loci.

Approximately 10 μg of poly A+ RNA from undifferentiated P19 cells was electrophoresed on a formamide-agarose gel, blotted to nitrocellulose, and hybridized against either pooled single-copy DNA probes or moderately repeated sequence probes from the various loci, as indicated in the figure. The names of the loci are shown at the bottom of each lane. The arrow indicates the approximate position of the 28S ribosomal RNA.

017-Locus
023-Locus
023-Locus
034-Locus
042-Locus
060-Locus
Singel-copy probes
Single-copy probes
Probe containing repeated sequences
Single-copy probes
Probes containing repeated sequebces
Probe containing repeated sequences

Against the 034 locus, the pooled single-copy probes detected a band which corresponded to the size of 28S ribosomal RNA (arrow). There was another, very faint band of about 7 kb size (above the major band in the 4th lane, not reproduced very well in the photograph) (a cDNA screening was done against this locus, data not shown, see Discussion).

To analyse the 042 locus, all the probes, including the repeated-sequences probes, were pooled and used in the Northern analysis. Lane 5 shows that there are several hybridizing messages. This locus was further analyzed by performing individual Northern analysis against each of the probes. However, this time only the probe containing repeated sequences hybridized to a weak smear and none of the single-copy probes detected any discrete messages (data not shown).

The 060 locus probe, which contains repeated sequences, detected a strong discrete band of about 6.5 kb (above the band shown by the arrow in lane 6). The less intense band shown by an arrow corresponded to the size of 28S ribosomal RNA (see Discussion below).

III.4) DISCUSSION

In the Northern blotting analysis, about 10 μg of poly A+ RNA was used; the poly A+ selection was done only once and, therefore, some poly A- and ribosomal RNA contamination can be expected. Furthermore, depending upon the

probe, cross-hybridization of even single-copy probes to ribosomal RNA can be observed. Therefore, it is likely that the bands shown by arrows in the figures are cross-hybridization to 28S ribosomal RNA; especially since the same 5.5 kb band was detected in several loci (023, 034, 042, and 060, see Figure III-8, arrows).

A cDNA library prepared from the undifferentiated P19 stem cells was screened separately with the single-copy probes from the 034 locus and a probe containing some repeated sequences, from the 060 locus. Four cDNAs were isolated (data not shown) with the 060 probe, however, subsequent analysis showed that none of the four clones represented the 060 locus. Therefore, it is likely that the strongly hybridizing 6 kb message was coming from some other locus but cross-hybridizing to repeated sequences in the 060 probe. In the case of 034, no cDNAs were isolated against either the strong or the weak messages. This result and the size of the message suggests that the strong band detected against this locus may possibly be the ribosomal RNA. However, the failure to isolate cDNAs can also be explained by a) poor quality of the cDNA library, b) the fact that when a cDNA library is made from cDNAs, the regions corresponding to the most distal (3') exons of a gene are better represented in the population, particularly if the mRNAs are long. If the probes are from the 5' region of that gene, one would miss the gene unless full length cDNAs of that gene were present. In any case, further chromosomal walking, isolation of single-copy probes, and Northern analysis in the presence of ribosomal DNA as competitor, is necessary to elucidate the identity of these messages.

Against the 023 locus, it is necessary to perform Northern analysis for each of the single-copy probes separately with 28S ribosomal DNA as competitor during the hybridization.

The selection procedure used in our approach involved activation of a Neo-R gene by an endogenous enhancer. Since an enhancer can activate either from 5' or from 3' of a gene, in an orientation independent manner, half of the stem cell specific loci were expected to contain an activating element 5' of the Neo-R gene. Although it is ideal to clone both the 5' and the 3' flanking cellular sequences of the transgene, the availability of fragments of appropriate lengths and having appropriate cohesive ends, cloning was restricted to the 5' sequences in the majority of the cases (see Figure II-6, Chapter II). This reduced the probability of cloning genes from these loci.

In addition, the cloned enhancers (015, and 052, see Chapter II, Results and Discussion on the 034 locus) belonged to a stem cell stage specific transposon. The 5' sequences from the remaining loci, 017, 023, 042, and 060, had no enhancer activity (Chapter II). This lessened the probability of the presence of genes in these loci, unless the enhancer was present in these loci further upstream from the cloned DNA, or activation of the Neo-R gene was brought about by some unknown mechanism similar to the locus activation domains of the β-globin gene locus (Grosveld *et al.* 1987).

Alternatively, if we assume that these loci have enhancers and associated genes at the other side of the Neo-R gene, it is less likely that one will be easily able to clone such enhancers and genes by genome walking using cellular sequence probes from only the 5' or the 3' of the Neo-R gene insertion, as I did here. This is because the integration of the enhancer-trap plasmid is a type II integration event which usually accompanies large chromosomal deletions at the integration sites (Gheysen *et al.* 1987). Therefore, to connect the two sides, it may be necessary to walk hundreds of kilobases along the chromsome. In fact, it is more straight forward and perhaps more worthwhile to clone the other side of Neo-R gene from the NEO^+ cell lines and search for the genes, since this would allow one to walk along the chromosome in all directions from either side of the integration site.

In any case, the question arises as to what is the possibility of having genes in differentially activated chromatin regions. An important clue comes from the work of Frank Grosveld and coworkers (1987). They identified two sequences, located 50 kb upstream and 20 kb downstream of the β-globin gene. These sequences, when together with a minimal β-globin containing DNA segment, brings the β-globin gene expression to quantitatively normal levels in transgenic mice having a single copy of the gene. Furthermore, these sequences conferred autonomy to β-globin expression, i.e. high expression occured independent of chromosomal site of insertion when these flanking sequences were present. These results, therefore, provide the strongest evidence yet for not only the presence of dominant,

differentially active chromosomal regions, but also the association of such active domains with a spatially regulated genetic loci.

Although not quite the same, it is tempting to suggest an association of gene transcription with chromosomal activity, for the chromosomal puffing observed in the polytene chromosomes of *Drosophila* salivary glands (see Serfling, 1982 and references therein) or the heat shock-induced puffs at the Hsp loci (Ashburner and Bonner, 1979). Both accompany the expression of the resident genes.

In addition, there have been several reports indicating that chromosomal position can affect the expression of a newly introduced transgene (see Introduction to Chapter II of this thesis). Also, when genes are introduced into mice by transgenics, correct expression (quantitative) of those genes is never observed. This indicates that transgenes were subjected to position effects (see Grosveld *et al.* 1987, and references therein). However, since we lack information regarding the number of attempts made where a thorough molecular analysis of the regions was performed, and regarding the number of successes in those attempts, it is difficult to predict the chances of cloning genes from differentially activated chromosomal domains.

Finally, I should also point out that the mouse and most other eukaryotes (as well as bacteria) presents a largely overlooked yet potential problem in searches of this kind. That is, the mouse genome contains several families of repeated sequences and retroviral or similar gene families with copy numbers each varying

between 1000-5000. Furthermore, many of these are transcribed in a developmentally regulated manner (temporally and/or spatially, examples include the ETn gene family; intracisternal particle genes; see Brulet *et al.* 1983; Levine *et al.* 1984; Ikuma *et al.* 1986; Callarco and Szollosi, 1973; Lueders and Knuff, 1980; Murphy *et al.* 1983). This large number of loci, therefore, would effectively compete with the small number of differentially active domains that might harbour important regulatory genes. The probability of insertion of an exogenously introduced transgene into a functional regulatory genetic loci, therefore, is extremely low unless ways are designed to exclude or to identify the above mentioned and similar non-specific genetic loci.

CHAPTER IV

STEM CELL SPECIFIC EARLY TRANSPOSON-LIKE GENES ARE MOVABLE IN THE MOUSE GENOME

IV.1) INTRODUCTION

It was Barbara McClintock (1951; 1956), who observed that mobile genetic elements "controlling elements" as she called them, alter the temporal gene expression programs during development. In recent years, several examples have been described in which transposon insertion influenced the expression of adjacent genes. The best studied examples include inactivation of genes involved in *Drosophila* eye colour (Rubin *et al.* 1982), developmental mutants of mouse (Jaenisch *et al.* 1983), activation of c-myc in bursal lymphomas by insertion of proviral copies after infection by avian leukosis virus (Jenkins *et al.* 1981; Westaway *et al.* 1984), and provirus imposed androgen regulation of the sex-limited protein (slp) gene in mouse (Stavenhagen and Robins, 1988).

Most strains of mice contain copies of several types of retroviral sequences in their genomes, capable of replicating in murine cells (Chan *et al.* 1980; Jenkins *et al.* 1982). In addition, the mouse genome contains several families of moderately repeated retrovirus-like structures which are non-infectious and are generally

expressed in early mouse embryos, mouse tumors or EC stem cells but very rarely in normal tissues. The intracisternal A-particle genes (IAP), for instance, is one such group found transcribed and translated in early mouse embryos and mouse tumors (Callarco and Szollosi, 1973; Chase and Piko, 1973; Kuff *et al.* 1972). This IAP genome is known to be mobile in the mouse genome (Kuff *et al.* 1983) and is responsible for the enhanced expression of the Ren-2 gene in mice (Burt *et al.* 1984) and the inactivation of immunoglobin Ck genes in a hybridoma cell line (Hawley *et al.* 1982; Kuff *et al.* 1983).

Akin to IAPs, Brulet *et al.* (1983) described another family of transposon-like or integrated retrovirus-like elements which are expressed as 6 kb mRNA in undifferentiated EC cells but not in the differentiated cell types. Their expression was also found to be restricted to early mouse embryos (Ikuma *et al.* 1986). While cloning chromosomal loci specifically active in the stem cells, I isolated stem cell specific enhancer elements (Chapter II). These enhancers were found to be part of the early transposon-like element described by Brulet *et al.* (1983). Comparison of the loci containing the ETn to wild type pre-insertion loci revealed that the ETn sequences were inserted during the experimental time frame. These observations are of interest because, having a stem cell specific enhancer in their genomes, ETn insertion next to a gene can confer stem cell specificity to its expression. Although a systematic search to demonstrate cohabitation of ETn and a gene in the mouse genome is not very practical, it is important from this viewpoint to demonstrate

that ETn genes are in fact movable in the mouse genome. In this chapter, I describe experiments that provide evidence to show that these ETn genes are transposable in a murine embryonal carcinoma cell line.

IV.2) MATERIAL AND METHODS

Materials and Methods specific to this chapter are described here. Remaining molecular/cell biology procedures (such as cells, plasmid and phage vectors, Southern blotting, molecular cloning procedures etc.) are described in Chapters II and III.

IV.2.i) Cells

In addition to P19 cells, two other EC cell lines, PCC3 and F9, were used for preparing the cell ("Hirt") extract. The cells were maintained as previously described.

The tissue culture work was done by H.Hamada.

IV.2.ii) Preparation of the EC cell extract (Hirt extract)

Cell extracts were prepared from three EC cell lines, P19, F9 and PCC3, from undifferentiated cells, as described by Hirt (1967).

IV.2.iii) Description of probes and Southern blots

IV.2.iii.a) The 015 locus probe

In order to demonstrate rearrangement at the 015 locus, a single-copy sequence, 015 *Hind*III-*Xba*I 0.5 kb (see Figure IV-2) was used as probe in genomic Southern blot analysis of P19 and 015 DNAs.

IV.2.iii.b) The 034 locus probe

The 034 locus rearrangement was demonstrated using the probe, 034 *Sst*I-*Bam*HI 1.0 kb fragment (see Figure IV-1), in the genomic Southern blots of P19 and 034 DNAs.

IV.2.iii.c) The 052 locus probe

Rearrangement at the 052 locus was detected with the single-copy probe 052 *Sst*I-*Hind*III 1.0 kb fragment (see Figure IV-3). The genomic DNAs from P19, 052, testis and various somatic tissues from C3H strain mice were used for the Southern blot analysis.

IV.2.iii.d) A transposon-specific probe

The early transposon-specific probe (Tn-specific) comprising the entire transposon (*Xba*I 5.5 kb fragment from the 052 locus, see Figure IV-3) was subcloned in pUC 12 from the 052 locus. This *Xba*I 5.5 kb fragment was excised, gel purified and used in the Southern blot analysis of the Hirt extracts to examine the presence of extrachromosomal copies of the transposon.

IV.2.iv) Southern blot analysis of the Hirt extract

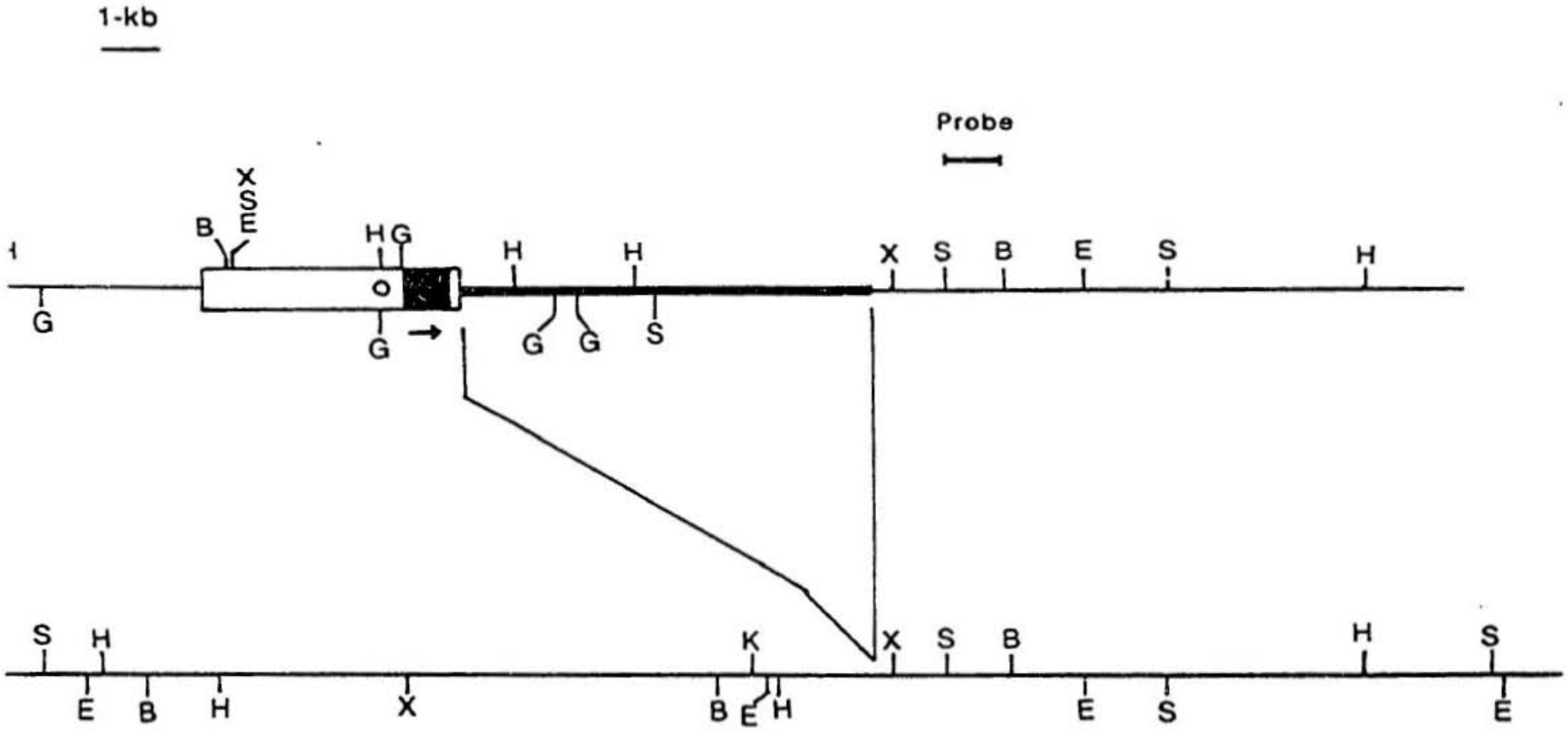
About 10 μg of DNA from the Hirt exctract from P19 cells, PCC3 and F9 cells were electrophoresed on a 0.7% agarose gel and the Southern blot was probed with the above Tn-specific probe. transposon were used as probes. Probe 1 was the *Eco*RI-*Pst* 0.1 kb fragment (right end) and Probe 2 was the *Hind*III-*Sst* 0.1 kb fragment (left end). Both were obtained from the 052 locus, see Figure IV-3.

Figure IV-1. The insertion of a copy of the ETn genome has rearranged the 034 locus.

Using a single-copy DNA probe ("probe", shown above the NEO$^+$-locus in the upper drawing), overlapping genomic clones were isolated from a P19 genomic library (see Chapter III). These clones were analyzed by restriction mapping (lower drawing) and compared to the restriction map of the 034-NEO$^+$ locus isolated previously (Chapter II). The thick line in the upper drawing of the NEO$^+$-locus is the inserted ETn genome. Confirmation of this transposed DNA element as being the ETn genome is based on the sequence analysis of the ETn promoter and also by comparison to the restriction map of the ETn genome reported by Brulet *et al.* 1983. The open box is the pA10neo sequence; closed box, Neo-R coding sequence; open circle, the SV40 early promoter. The location of integration of the ETn genome together with the pA10neo is known by the lines. B, *Bam*HI; E, *Eco*RI; G, *Bgl*II; S, *Sst*I; K, *Kpn*I; Sa, *Sal*I; X, *Xba*I (see text).

IV.2.v)"End" specific probes of the transposon

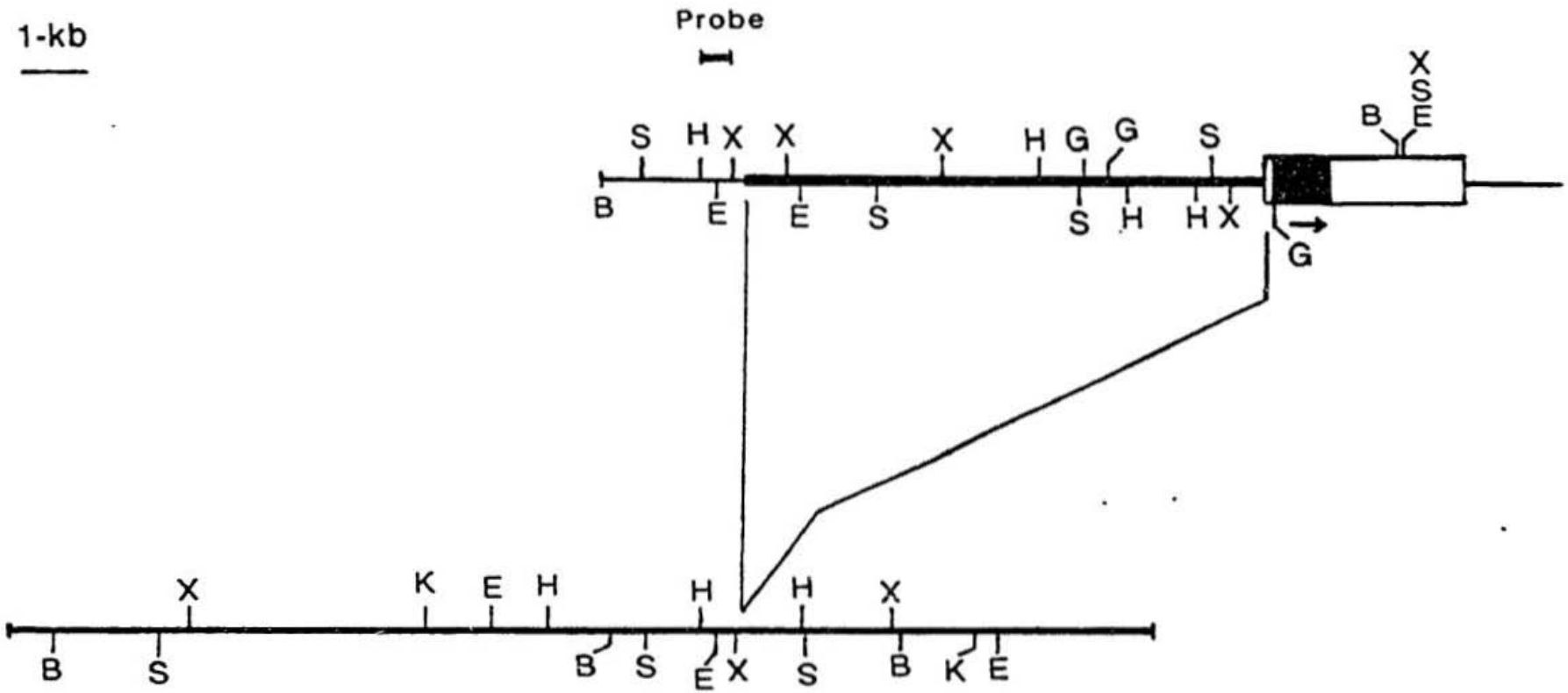
To analyse whether there are differences in the ETn insertion sites between the germ line and the somatic cell lineages, fragments from the two ends of the



LOCUS 034

Figure IV-2. Rearrangement at the 015 NEO^r-locus is due to the ETn genome insertion.

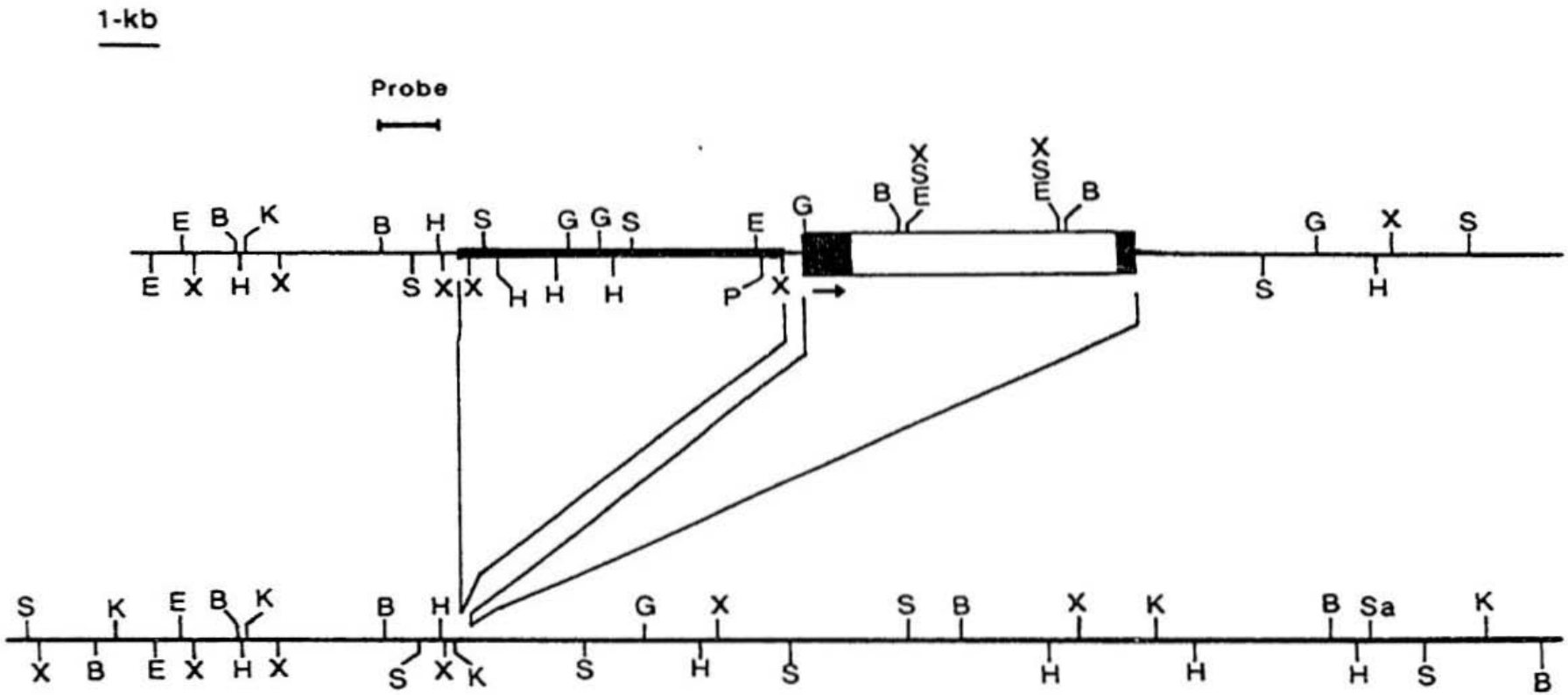
A comparison of the restriction maps of the abnormal 015 NEO^r-locus (upper drawing) and the normal preinsertion locus (lower drawing) shows that a copy of the ETn genome along with the pA10neo has integrated into this locus (location of the insertion is shown by the lines). See legend to Figure IV-1 for symbols and abbreviations of the restriction sites.



LOCUS 015

Figure IV-3. The rearrangement at the 052 NEO$^+$-locus is due to the ETn insertion.

A comparison of the restriction maps of the 052 NEO$^+$-locus (upper drawing) and the corresponding normal preinsertion region (lower drawing) shown in this figure indicates that a copy of the ETn genome has integrated into this locus. The locations of insertion of the ETn and the pA10neo are shown by the lines. Note the small streach of (about 200 bp) mouse sequence between the ETn and the pA10neo (see text). See legend to Figure IV-1 for symbols and abbreviations of the restriction sites.


1-kb
Probe
E B K
B H S G G S E G B X S E X S E B G X S
E X H X S X X H H H P X S H
S K E B K B H G X S B X K B Sa K
X B E X H X S X K S H S H H H S B

LOCUS 052

IV.2.vi) Southern blot of the Germ line and Somatic DNA

DNAs (15 μg each) from C3H mouse strain testis, various somatic tissues, and P19 cells (D- and D+), were digested with *Eco*RI and another batch was digested with *Hind*III and both were electrophoresed in a long gel for two days. The *Eco*RI digested set was hybridized with probe 1 (*Eco*RI-*Pst* 0.1 kb, right end) and the *Hind*III digested set was hybridized with probe 2 (*Hind*III-*Sst* 0.1 kb, left end). Hybridization and washing conditions were as described in Materials and Methods, Chapter II.

IV.3) RESULTS

IV.3.i) Rearrangement in the 034, 015 and 052 loci is the result of ETn insertion

Previously, I isolated several stem cell stage specific loci from P19 EC cells. Further analysis of these loci showed that three of the loci, namely 015, 052 and 034, contained a complete or a partial sequence of the early retrovirus or transposon-like element described by Brulet *et al.* (1983). It was interesting to find out whether these ETn are endogenous to these loci or inserted into them during the experimental time-frame. Therefore, using single-copy probes from these loci, the corresponding pre-insertion regions were cloned from a P19 genomic library

and restriction maps of the 015, 052 and 034 loci were compared with those of corresponding regions in the genome. The results showed that, the ETn insertion was not endogenous to these loci (015, 052, and 034), and thus that the insertion must have occured while isolating the NEO^+ cell lines.

In the 034 locus, an incomplete copy of the ETn genome was inserted 3' to the direction of Neo-R gene transcription (Figure IV-1). The ETn sequence extended from near the *Xba*I site to the pA10neo plasmid (thick bar in the upper drawing in Figure IV-1). The ETn genome was not present in this region of the normal genome.

In the 015 locus as well a comparison of the restriction maps of the 015 locus to the corresponding normal locus revealed that a complete copy of the ETn genome was inserted in the 015 NEO^+ cell line at this locus (Figure IV-2). The ETn extended upstream from the Neo-R gene, replacing the SV40 early promoter of the enhancer-trap plasmid, pA10neo (Figure IV-2, thick bar in the upper drawing).

Similarly, the restriction maps of the 052 locus and the corresponding region in the genome showed that ETn was inserted, extending from near the *Hind*III site to very close to the Neo-R gene (thick bar in the upper line drawing, Figure IV-3).

The 052 cell line contained one full copy and also a partial copy of Neo-R gene (solid box in Figure IV-3) (see Chapter II). From this locus, about 32 kb

of the normal genome was cloned. When the restriction map of the normal region was compared with that of 052 locus, the 32 kb region contained both the right side and the left side of the ETn + Neo-R integration site. This means that the integration in this locus had occured without chromosomal deletion (see Figure IV-3). In the other two cell lines, the ETn side of the integration sites also had no chromosomal deletions. However, the left side of the integration site in the 034 locus (the plasmid side) was different from that of cloned normal regions indicating that integration had been accompanied by a chromosomal deletion (Figures IV-1 and IV-2). Also, in all three loci, the ETn and the pA10neo plasmid were together, and in 015 and 034, the ETn sequences were immediately adjacent to the plasmid sequence. In the 052 locus, however, about 200-bp of single copy mouse sequence was present between the ETn and the plasmid pA10neo (see Figure IV-3). However, it is difficult to decide whether: i) the integration has occurred independently, one followed by the other, ii) at the same time, or iii) the ETn was initially integrated into the plasmid, extrachromosomally, and the combined sequences then inserted into the genome. In the 052 locus, it appears that the ETn integration and the pA10neo plasmid integration occurred independently ("iii" above can be ruled out) since there is an interruption between the ETn and the plasmid by a single copy mouse sequence.

The transposition of the ETn genome was also analysed by genomic Southern blotting. To detect the rearrangement in the 052 locus, high molecular weight DNA

from C3H strain mouse testis, P19 cells, and 052 NEO^+ cells, were digested with *Eco*RI. The Southern blot was probed with a single-copy DNA probe, the *Hind*III-*Bam*HI 1 kb fragment from the 052 locus (see Figure IV-3). In the germ line, a single fragment of about 20-kb corresponding to the normal allele was detected (Figure IV-4A, lane 5). This 20-kb fragment was also detected in P19 cells (lanes 1 and 2). In the 052 cells, in addition to the normal 20-kb fragment, the probe also detected another 10-kb fragment (lanes 3 and 4; the second *Eco*RI site is in the ETn, see Figure IV-3 for the restriction map). This abnormal 10-kb fragment was due to the rearranged allele of this locus.

To detect the rearrangement in the 034 locus, high molecular weight genomic DNA from P19 cells and the 034 cell line were digested with *Hind*III and the Southern blot was probed with the 034 locus-specific probe, the *Sst*I-*Bam*HI 1 kb fragment. In the P19 cells, the probe detected a single 10-kb band (Figure IV-4B, lane 1). However, in the 034 cell line, in addition to the 10-kb fragment, the probe detected a 13-kb fragment (lane 2). In this locus, the ETn insertion had disrupted the *Hind*III site in the genome and as a result, the rearranged allele is detected as a 13-kb fragment (see the restriction map in Figure IV-2).

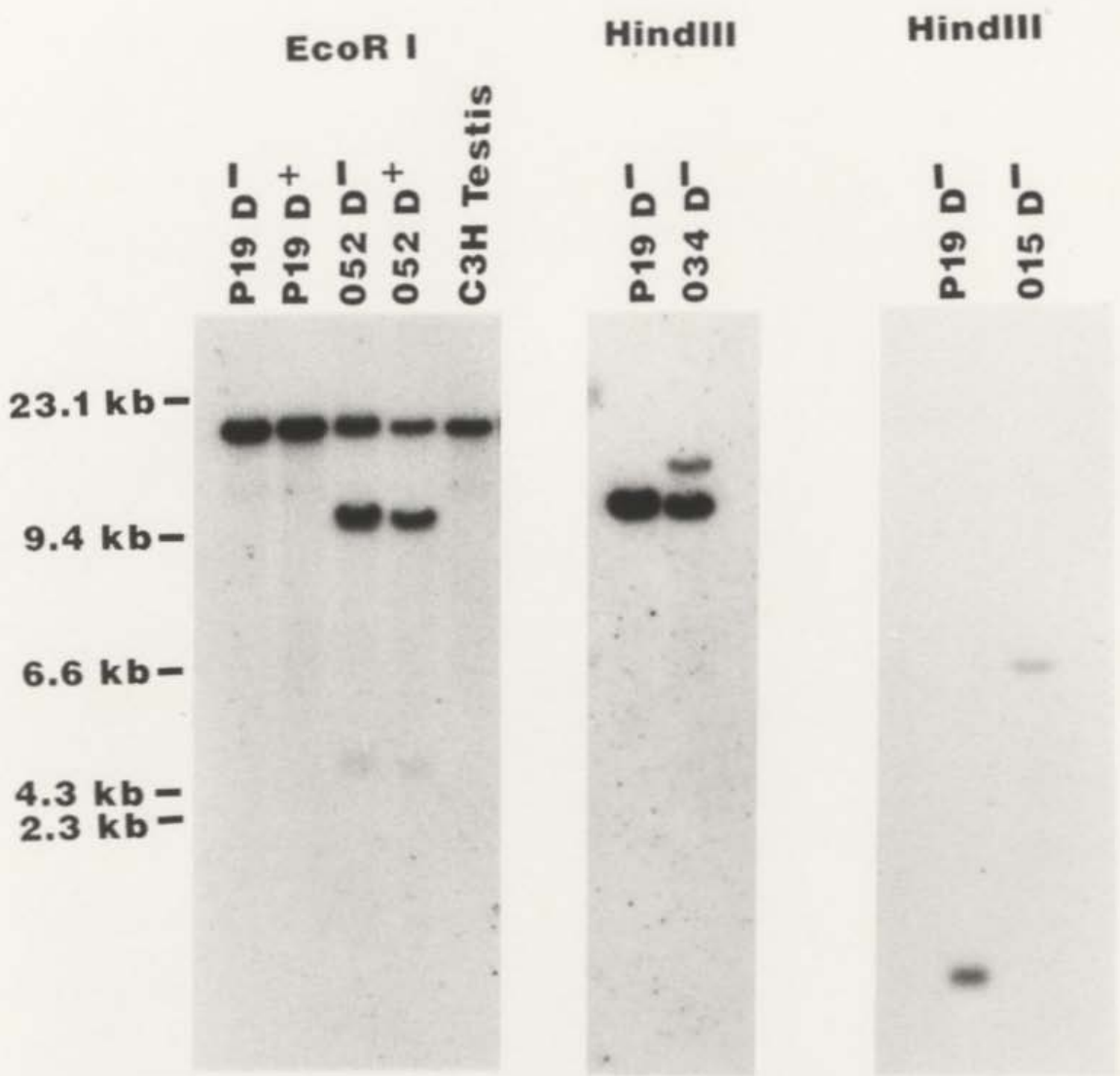
Southern blotting analysis of the 015 locus was done by digesting the high molecular weight DNA from P19 cells and 015 cells with *Hind*III. When the blot was probed with the *Hind*III-*Xba*I 0.5 kb fragment, a single-copy DNA probe

Figure IV-4. Southern blot analysis of the three rearranged loci.

(A) High molecular weight DNAs from undifferentiated (D-) and differentiated (D+) P19 cells, and 052 NEO^r cell line, and C3H testis DNA were analyzed with a single-copy DNA probe specific to the 052 locus. The 10-kb abnormal band in the 052 lanes comes from the rearranged allele (note the *Eco*RI site in the upper restriction map, Figure IV-3).

(B) DNAs from P19 (D-) and 034 NEO^r cell line were analyzed with a 034 locus-specific probe. The upper band of 13-kb size in 034 D- lane arises from the rearranged allele (see restriction map in Figure IV-1, note the *Hind*III site).

(C) DNAs from P19 (D-) and 015 NEO^r-cell line were analyzed with a single-copy DNA probe specific to the 015 locus. The 5.5-kb band in the 015 D-lane arises from the rearranged allele. The normal allele is represented by the 1-kb band in the P19 D- lane. This normal allele was not detected in the 015 NEO^r cell line (see Discussion). The enzymes used for analysis of each locus are shown at the top.



A. 052 locus probe

B. 034 locus probe

C. 015 locus probe

specific to the 015 locus, a 1.2-kb fragment was detected in P19 cells (Figure IV-4C, lane 1). However, in the 015 cell line a single fragment of 5.5-kb size representing the rearranged allele was detected (lane 2) (see Discussion).

IV.3.ii) The ETn genome can be detected extrachromosomally in the P19 stem cells

High frequency transposition of a DNA sequence is generally characterised by a high extrachromosomal copy number within the cell (Krowleski and Rush, 1984). To test whether the ETn genome exists extrachromosomally in the EC stem cells, extrachromosomal DNAs were isolated from the undifferentiated EC cells (P19, PCC3 and F9 cell lines) according to Hirt (1967). The DNAs were resolved by agarose gel electrophoresis and analyzed by Southern blotting using an entire copy of the ETn genome (*Xba*I 5.5 kb from the 052 locus) as the probe. The ETn probe in the blot detected a 6 kb fragment in the P19 cell extract (Figure IV-5, lane 1) but not in the F9 or the PCC3 extracts (lanes 2 and 3). This 6 kb band detected in P19 cells probably represents the linear ETn genome, no other forms (such as circular or supercoiled) were detected in the blot (see Discussion).

Figure IV-5. Extrachromosomal DNA copies of the ETn genome are present in P19 cells.

Hirt extracts from P19, F9 and PCC3 EC cell lines were electrophoresed on an agarose gel, blotted and hybridized against an entire copy of the ETn genome (*Xba*I 5.5-kb from the 052 locus, see restriction map in Figure IV-3). A 6-kb band is seen only in the P19 cell extract. Tn, early transposon-like element (ETn).

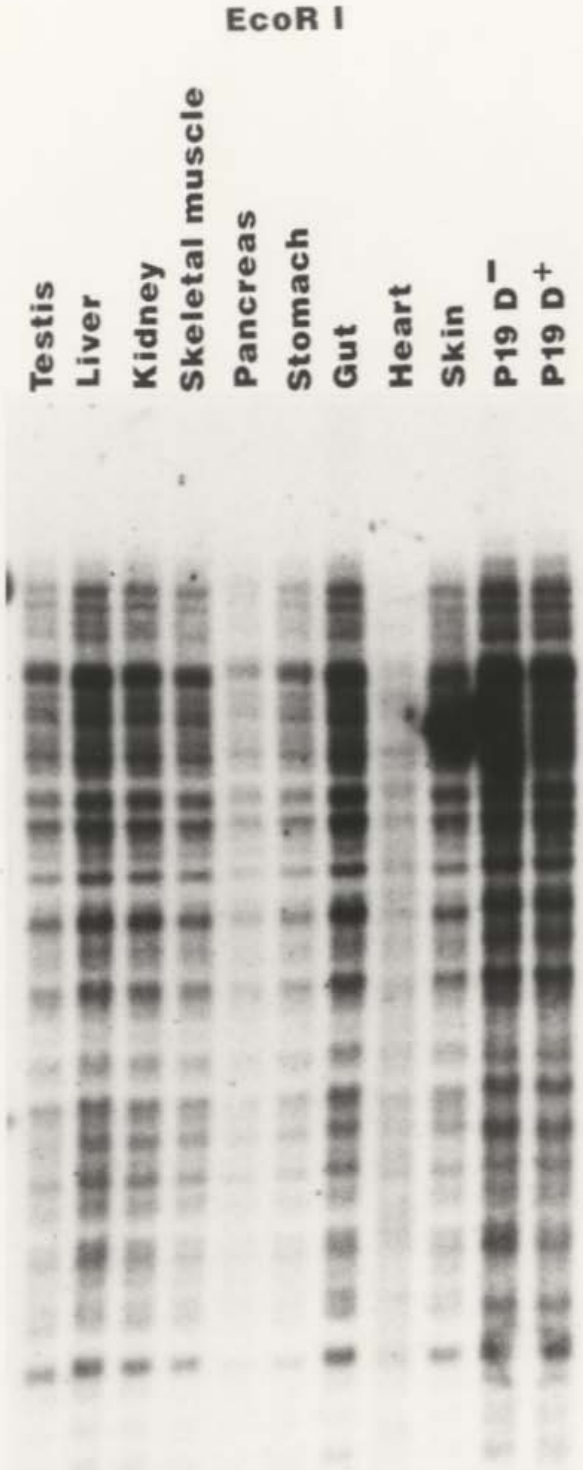
Tn specific probe

IV.3.iii) The ETn transposition event is possibly restricted to very early stages of mouse embryogenesis

During mouse embryogenesis, Soriano and Jaenisch (1986) showed that the germ line is allocated prior to the somatic lineages. If the transposition is active during and/or after the allocation of cells to different lineages, this would result in the unique integration sites in the germ line and soma, and among the various somatic tissues. Alternatively, if the transposition ceases to occur before the allocation of founder cells to somatic lineages then there would be no difference in the ETn integration sites between the germ line and the soma. Therefore, DNA from testis, various somatic tissues, and P19 cells were analysed by Southern hybridization. Two different probes from the two ends of the ETn genome were used (see Materials and Methods for the description of the probes). For probe 1, *Eco*RI-*Pst*I 100 bp, the DNAs were digested with *Eco*RI, which would allow the detection of differences in the right side of the ETn integration site. For probe 2, *Hind*III-*Sst* 100 bp, the DNAs were digested with *Hind*III, allowing the analysis of the flanking cellular sequences at the left side. The results shown in Figure IV-6, A and B, however, indicate that there are no apparent differences in the ETn integration sites between the germ line and the somatic cells, or among the various somatic tissues examined, and P19 cells as well.

Figure IV-6. Analysis of the integration sites of the ETn genomes in the germ cells and various somatic tissues.

(A) Analysis of the DNA at the left side of the ETn integration site. High molecular weight DNA (10 μg) from testis, various somatic lineages of the C3H strain mouse, and D- and D+ P19 cells were digested with *Eco*RI and electrophoresed on a long gel. The DNAs were blotted and hybridized to a left side ETn end-specific probe (*Eco*RI-*Pst*I 100 bp, from the 052 locus, see the restriction map in Figure IV-3, upper drawing).



Probe: Eco-Pst 0·1 kb

Figure IV-6:

(B) Analysis of DNA at the right side of the ETn integration site. The same set of DNAs used in (A) were digested with *Hind*III and the blot was hybridized against a right side ETn end-specific probe (*Hind*III-*Sst* 100 bp, also from 052 locus, see the map in Figure IV-3, upper drawing). *Hind*III digestion was partial in some cases for unknown reasons.

Probe: Hind-Sst 0·1 kb

IV.4) DISCUSSION

The findings in this report provide direct evidence that ETn-like elements can act as a movable genetic unit in the mouse genome. Of importance is the previous finding that the ETn genome contains strong transcriptional enhancers which are stem cell specific (Chapter II), and which had rendered the adjacent Neo-R transgene stem cell specific in the 015, 034 and 052 loci. This demonstrates the possibility that these genetic units can act as powerful stem cell insertion mutagens and regulate and/or alter gene expression in the founder cells of an embryo.

Restriction maps of the normal and the transposed chromosomal regions and the Southern blot analyses of the three loci clearly show that the ETn genomes were inserted into these loci and were not endogenous to them. The restriction maps from these loci indicate a clean transition from the cellular sequences to the ETn sequences with no deletions of cellular sequences at the transposon side (see Figures IV-1, IV-2 and IV-3). There are two general types of DNA insertion into the genome of an eukaryotic cell (see Gheysen *et al.* 1987). In the first mechanism, referred to as a type I event, deletion of target DNA is not common (exemplified by transposons and retroviruses). In the second mechanism, referred to as a type II event, such chromosomal deletions at the integration sites are common (exemplified by introduction of DNA through transfection, electroporation, infection

with viruses such as polyoma, SV40, or adenovirus). Therefore, it is not surprising that there was no deletions at the ETn integration sites in these cases. Although in the 034 locus the ETn is partial, it is unlikely that this partial copy is a result of a deletion during transposon integration (see below).

In the Southern blot analysis of the rearrangement at the 034 locus, lanes 1 and 2), the intensity of the two bands representing the normal and the rearranged fragments was not the same (see Figure IV-4B, lanes 1 and 2). It is possible that the normal 10-kb fragment detected by the probe is repeated in the haploid genome and the transposon insertion had occured only in one location. This would give the different densities for the rearranged and the normal fragments.

In the 015 Southern blot, the normal 1.2-kb fragment was not detected in the 015 cell line (Figure IV-4C, compare lanes 1 and 2). At least two possibilities can be forwarded to explain why the normal allele was not detected in this cell line. First, two copies of the transposon may have integrated into both of the chromosomes by homologous recombination. Second, the *Hind*III-*Xba*I 0.5 kb fragment may have been deleted (or a part of the chromosome or the entire chromosome which harbors this sequence). Judging from the intensities of the two bands (compare lanes 1 and 2), the latter possibility is more likely.

As far as the mechanism of transposition into these three loci is concerned, it may be difficult to distinguish between the transposition of ETn gene copies already present in the genome and the insertion of new proviral forms reverse-

transcribed from an ETn transcript. It is widely believed that circular DNAs are intermediates in DNA-to-DNA transposition (Flavell and Ish-Horowicz, 1981; Temin, 1980; Varmus, 1982). Analysis of the cell culture extract from P19 cells revealed a single fragment of 6-kb (Figure IV-5, lane 1), which is same as the size of the ETn transcript. No other forms were observed. Therefore, cycles of reverse transcription and proviral insertion are more likely than the direct DNA-to-DNA transposition. The presence of a partial copy of the transposon in the 034 locus therefore may be due to the integration of an incompletely reverse transcribed DNA copy of the ETn rather than a deletion during integration. Also, the alternative mechanism of the precise transposition of an integrated retroviral gene copy has not yet been demonstrated.

Although the ETn genome is transcriptionally active in F9 (Ikuma *et al.* 1986) and PCC3 cells (see Brulet *et al.* 1983), extrachromosomal ETn copies were not detected in these EC cell lines (lanes 2 and 3 in Figure IV-5); however, a small but functionally significant number could well have escaped detection.

No apparent differences in the integration sites of the transposon between the germ line and soma, nor among the various somatic tissues was observed. One explanation is that if the transposition event is restricted to early stages before the allocation of embryonic cells to germ cell lineage, the ETn integration sites will not be different in the germ cell and somatic cell ineages. One might wonder, then, why there was no difference between P19 and germ line/somatic tissues, since ETn

transposition was occuring in the P19 cells. Several possibilities should be considered here. Firstly, since the integration occurs generally at random sites in the chromosome, the contribution of individual cells in which the transposition might have created new integration sites to the total cells used for DNA extraction would be too little to be detected by the Southern blotting. Secondly, the transposition frequencies for mobile elements are generally very low, in the order of 10^{-4} to 10^{-6} per copy per generation (Finnegan, 1989). Thus the failure to detect any difference between the P19 cells and the germ line or the somatic tissues may also be due to a very low ETn transposition frequency. It is possible that the transposition of the ETn is activated in the P19 cells for unknown reasons.

It has been proposed that mobile elements, as a part of a regulatory processes, may exise from specific loci and integrate into some other specific loci during development influencing cell differentiation into lineages. The absence of difference for the ETn sites among the various lineages, however, rules out such a possibility, at least for the ETn. Participation of the ETn in the very early determinative events of embryogenesis in the mouse, however, is still an interesting possibility.

There is a possibility that the transposed ETn in the above three loci could be an artefact associated with DNA transfection, i.e., the large numbers of extrachromosomal copies of the ETn got trapped with the DNA-calcium phosphate precipitate and artefactually co-integrated into the mouse genome. The following indirect evidence however support my contention that the ETn is movable in the

mouse genome. They are: firstly, the 6kb-fragment observed in the Southern blot analysis of the cell extract using the ETn probe (Fig. IV-5) is DNA and not RNA (possibly a reverse transcribed product of ETn RNA). Secondly, the ETn has direct repeats at the ends, a commonly observed characteristic of other transposons and retroviruses which integrate into the host genome. These repeats are thought to be involved in the site specific recombination process during transposon integration. Thirdly, there are at least 1000 copies per haploid genome of these ETn-like genes in the mouse genome. A large increase in copy numbers of a gene is generally believed to occur by a) tandem duplication coupled with unequal cross overs and positive selection of such cells under a selective pressure, b) bi-directional replication, or c) extrachromosomal amplification, and d) transposition, particularly through cycles of reverse transcription and copy DNA integration. A likely explanation for the presence of >1000 copies of ETn genes in the mouse genome would be the reverse transcription of ETn RNA into DNA followed by transposition. Our results therefore are in agreement with the contention that these ETn-like genes are movable in the mouse genome.

CHAPTER V

CONCLUSIONS

Using the approach of randomly marking chromosomal loci with a transgene, several chromosomal domains which are active only in the undifferentiated EC stem cells but not in their differentiated derivatives, were cloned. The analysis of the cloned DNA revealed that three of the loci contained stem cell specific enhancer elements of the early transposon-like element genes.

The uncovering of the stem cell specific enhancer of the ETn in this study has provided the basis for further research. The enhancer isolated here is a stem cell specific enhancer. This stem cell specificity could be due to; i) positively regulating stem cell specific trans-acting facotr(s), ii) negatively regulating trans-acting factor(s) repressing the enhancer activity at the onset of differentiation, or iii) the stage specific modification of trans-acting factor(s) that interact with this enhancer either positively or negatively (see review by Ptashne, 1988). In any case, it is possible that there are target genes other than the ETn for these factors. The isolation of factor(s) and the gene(s) encoding them etc., using the enhancer element cloned in this study, will be a significant step in the study of embryogenesis.

The preinsertion chromosomal regions corresponding to the above stem cell specific loci were subsequently cloned and searched for genes. It might be

possible that the hybridizing messages found in some of the loci are from functional genes (possibly the 034 locus and the 023 locus), in which case the time consuming and risky part of their study has been accomplished.

Finally, evidence has been provided that the stem cell-specific ETn genes are movable in the mouse genome. This is significant because it indicates that the transposon can act as a powerful stem cell stage mutagen and can alter gene expression programs in the founding cells of the embryo. This has obvious implications for the study of evolutionary and developmental biology.

LITERATURE CITED

Adamson, E.D. (1987) Oncogenes in development. Development 99: 449-471.

Allen, N.D., Cran, D.G., Barton, S.C., Hettle, S., Reik, S. and Surani, M.A.H. (1988) Transgenes as probes for active chromosomal domains in mouse development. Nature 833: 852-855.

Anderson, K.V. (1987) Dorsal-ventral embryonic pattern genes of *Drosophila*. Trends. Genet. 3: 91-97.

Ashburner, M. and Bonner, J.J. (1979) The induction of gene activity in *Drosophila* by heat shock. Cell 17: 241-254.

Barklis, E., Mulligan, R.C. and Jaenisch, R. (1986) Chromsomal position or virus mutation permits retrovirus expression in embryonal carcinoma cells. Cell 47: 391-399.

Barton, S.C., Surani, M.A.H. and Noris, M.L. (1984) Role of paternal and maternal genomes in mouse development. Nature 311: 374-376.

Bensaude, O. and Morange, M. (1983) Spontaneous high expression of heat shock proteins in mouse embryonal carcinoma cells and ectoderm from day 8 mouse embryo. EMBO J 2: 173-177.

Bhat, K., McBurney, M.W. and Hamada, H. (1988). Functional cloning of mouse chromosomal loci specifically active in embryonal carcinoma stem cells. Mol. Cell. Biol. 8: 3251-3259.

Braude, P., Pelham, H., Flach, G. and Lobbato, R. (1979) Post-transcriptional control in the early mouse embryo. Nature 282: 102-105.

Brinster, R.L. (1974) The effect of cells transferred into the mouse blastocyst on subsequent development. J. Exp. Med. 140: 1049-1055.

Browder, L.W. (1984) *Developmental Biology*. Saunders College Publishing (Holt, Rinehart and Winston: The Dryden Press).

Brulet, P., Kaghad, M., Xu, Y.S., Croissant, O. and Jacob, F. (1983) Early differential tissue expression of transposon-like repetitive DNA sequences of the mouse. Proc. Natl. Acad. Sci. USA 80: 5641-5645.

Burt, D.W., Reith, A.D. and Brammar, W.J. (1984) A retroviral provirus closely associated with the Ren-2 gene of DBA/2 mice. Nucl. Acids Res. 12: 8579-8593.

Calarco, P.G. and McLaren, A. (1976) Ultrastructural observations of the preimplantation stages of the sheep. J. Embryol. Exp. Morph. 36: 609-622.

Calarco, P.G. and Szollosi, D. (1973) Intracisternal A-particles in ova and preimplantation stages of mouse. Nature New Biol. 243: 91-93.

Calos, M.P. and Miller, J.H. (1980) Transposable elements. Cell 20: 579-595.

Chabot, B., Stephenson, D.A., Chapman, V.M., Besmer, P. and Bernstein, A. (1988) The proto-oncogene c-kit encoding a transmembrane tyrosine kinase receptor maps to the mouse W locus. Nature 335: 88-89.

Chan, H.W., Bryan, T., Moore, J.L., Stall, S.P., Rowe, W.P. and martin, M.A. (1980) Proviral sequence in inbred mouse strains with a cloned subgenomic DNA fragment. Proc. Natl. Acad. Sci. USA 77: 5779-5783.

Chasan, R. and Anderson, K.V. (1989) The role of easter, an apparent sovine protease, in organizing the dorsal-ventral pattern of the *Drosophila* embryo. Cell 56: 391-400.

Chase, D.G. and Pilo, L. (1973) Expression of A- and C-type particles in early mouse embryos. J. Natl. Cancer Inst. 67: 677-685.

Cooper, A.R., Taylor, A. and Hogan, B.L.M. (1983) Changes in the ratio of lamin and entactin synthesis in F9 embryonal carcinoma cells treated with retinoic acid and cyclic AMP. Dev. Biol. 99: 510-514.

Copeland, N.G., Hutchinson, K.W. and Jenkins, N.A. (1983) Excision of the DBA ecotropic provirus in dilute coat-color revertants of mice occurs by homologous recombination involving the viral LTRs. Cell 33: 379-387.

Coppola, J.A. and Cole, M.D. (1986) Constitutive c-Myc oncogene expression blocks murine erythroluekemia cell differentiation but not commitment. Nature 320: 760-765.

Crosby, L.M., Bandolfi, F. and Moor, R.M. (1988) Control of protein synthesis during early cleavage of sheep embryos. J. Reprod. Fertil. 82: 769-775.

Cuttanach, B.M. (1986) Parental origin effects in mice. J. Embryol. Exp. Morph. 97 (Suppl): 137-150.

Cuttanach, B.M. and Kirk, M. (1985) Differential activity of maternally and paternally derived chromsome regions in mice. Nature 315: 496-498.

Damjanov, I., Damjanov, A. and Solter, D. (1987) In: *Teratocarcinomas and embryonic stem cells: A practical approach*, E.J. Robertson, ed. (Oxford-Washington: IRL Press), pp. 1-17.

Darnell, J.E., Lodish, H. and Baltimore, D. (1986) *Molecular cell biology*. (New York: Scientific Americal Books, Inc.), pp. 988-1009.

Davidson, E. (1976) *Gene activity in early development*. (New York, San Francisco, London: Academic Press).

Davis, R.L., Weintraub, H. and Lassar, A.B. (1987) Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell 51: 987-1000.

DeLotto, R. and Spierer, P. (1986) A gene required for the specification of dorsal-ventral pattern in *Drosophila* appears to encode a serine protease. Nature 32: 688-692.

Distel, R.J., Ro, H-S., Rose, B.S., Broves, D.L. and Spiegelman, B.M. (1987) Nucleoprotein complex that regulate gene expression in adipocyte differentiation: Direct participation of c-fos. Cell 49: 835-844.

Diwan, S.B. and Stevens, L.C. (1976) Development of teratomas from the ectoderm of mouse egg cylinders. J. Natl. Cancer Inst. 57: 937-939.

Dziadek, M. and Adamson, E.D. (1978) Localization and synthesis of alpha fetoprotein in post-implantation mouse embryos. J. Embryol. Exp. Morphol. 43: 289-295.

Edwards, M.K.S., Harris, J.F. and McBurney, M.W. (1983) Induced muscle differentiation in an embryonal carcinoma cell line. Mol. Cell. Biol. 3: 2280-2286.

Edwards, M.K.S. and McBurney, M.W. (1983) The concentration of retinoic acid determines the differentiated cell types formed by a teratocarcinoma cell line. Dev. Biol. 98: 187-191.

Ek[illegible], R., Mossige, J., Johannessen, J.V. and Aars, H. (1981) Heriditary renal adenomas and adenocarcinomas in rats. Diag. Histopathol. 4: 99-110.

Engels, W.R. (1979) Extrachromosomal control of mutability in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 76: 4011-4015.

Federoff, N.V. (1983) In: *Mobile genetic elements*, J.A. Shapiro, ed. (New York: Acadmic Press), pp. 1-57.

Feinberg, A.P. and Vogelstein, B. (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 132: 9-13.

Finnegan, D.J. (1989) Eukaryotic transposable elements and genome evolution. Trends. Genet. 5 (4): 103-107.

Flavell, A.J. and Ish-Horowicz, D. (1981) Extrachromosomal circular copies of the eukaryotic transposable element copia in cultured *Drosophila* cells. Nature 292: 591-585.

Frischauf, A-M., Lehrach, H., Poustka, A. and Murray, N. (1983) Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol. 170: 827-842.

Gardner, H.A., Gallie, B.L., Knight, L.A. and Philips, R.A. (1982) Multiple karyotypic changes in retinoblastoma tumor cells: Presence of normal chromosome No. 13 in most tumors. Cancer Genet. Cytogenet. 6: 201-211.

Gardner, R.L. and Rossant, J. (1979) Investigation of the fate of 4.5 day post-coitum mouse inner cell mass cells by blastocyst injection. J. Embryol. Exp. Morphol. 52: 141-152.

Gardner, R.L. (1978) In: *Genetic Mosaics and Cell Differentiation*, W.J. Gehring, ed. (Berlin: Springer-Verlag), pp. 205-241.

Gardner, R.L. and Papaioannou, V.E. (1975) In: *The early development of mammals*, M. Balls and A.E. Wild, eds. (Cambridge, England: Cambridge University Press), pp. 107.

Gaunt, S.J., Miller, J.R., Powell, D.J. and Duboule, D. (1986) Homeobox gene expression in mouse embryos varies with position by the primitive streak stage. Nature 324: 141-152.

Gehring, W.J. (1987) Homeoboxes in the study of development. Science 236: 1245-1252.

Gehring, W.J. (1985) The homeo box: A key to the understanding of development? Cell 40: 3-5.

Gheyesen, G., Montagu, M.V. and Zambryski, P. (1987) Integration of *Agrobacterium tumefaciens* transfer DNA (T-DNA) involves rearrangements of target plant DNA sequences. Proc. Natl. Acad. Sci. USA 84: 6169-6173.

Giguere, V., Ong, E.S., Segui, P. and Evans, R.M. (1987) Identificaton of a receptor for the morphogen retinoic acid. Nature 330: 624-629.

Gorman, C.M., Moffat, L.F. and Howard, B. (1982) Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. Mol. Cell. Biol. 2: 1044-1051.

Griep, A.E., Westphal, H. (1988) Antisense Myc sequences induce differentiation of F9 cells. Proc. Natl. Acad. Sci. USA 85: 6806-6810.

Grosveld, F., Assendelft, G.B., Greaves, D.R. and Kollias, G. (1987) Position-independent, high-level expression of the human β-globin gene in transgenic mice. Cell 51: 975-985.

Groudine, M. and Weintraub, H. (1982) Propagation of globin DNAase I-hypersensitive sites in absence of factors required for induction: A possible mechanism for determination. Cell 30: 131-139.

Guads, L.J., Singh, J.P. and Stiles, C.D. (1983) Secretion of growth regulatory molecules by teratocarcinoma stem cells. Cold Spring Harbor Conference, Cell Proliferation 10: 229-236.

Gurdon, J. (1976) *The control of gene expression in animal development.* (Clarendon Press), pp. 3-36.

Hamada, H. (1986a) Activation of an enhancerless gene by chromosomal integration. Mol. Cell. Biol. 6: 4179-4184.

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

Hart, C.P., Awgulewitsch, A., Fainsod, A., McGinnis, W. and Ruddle, F.H. (1985) Homeobox gene complex on chromosome 11: Molecular cloning, expression in embryogenesis and homology to a human homeobox locus. Cell 43: 9-18.

Hawley, R.G., Shulman, M.J., Murialdo, H., Gibson, D.M. and Hozumi, N. (1982) Mutant immunoglobulin genes have repetitive DNA elements inserted into their intervening sequences. Proc. Natl. Acad. Sci. USA 79: 7425-7429.

Hen, R., Borelli, E., Fromental, C., Sassone-Corsi, P. and Chambon, P. (1986) A mutated polyoma virus enhancer which is active in undifferentiated embryonal carcinoma cells is not repressed by adenovirus-2 Ela products. Nature 321: 249-251.

Hilman, N. and Tasca, R.J. (1969) Ultrastructural and autoradiographic studies of mouse cleavage stages. Am. J. Anat. 126: 151-174.

Hiromi, Y. and Gehring, W.J. (1987) Regulation and function of the *Drosophila* segmentation gene fushi tarazu. Cell 50: 963-974.

Hirt, B. (1967) Selective extraction of polyoma DNA from infected mouse cell cultures. J. Mol. Biol. 26: 365-369.

Hogan, B.L.M., Constantini, F. and Lacy E. (1986) *Manipulating the embryo: A lab manual.* (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 18.

Hogan, B.L.M., Taylor, A. and Adamson, E. (1981) Cell interactions modulate embryonal carcinomal cell differentiation into pariental or visceral endoderm. Nature 291: 235-237.

Holland, P.W. and Hogan, B.L.M. (1988) Expression of homeobox genes during mouse development: A review. Genes Dev. 2: 773-782.

Hozumi, N. and Tonegawa, S. (1976) Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions. Proc. Natl. Acad. Sci. USA 73: 3628-3632.

Ikuma, S., Kiyota, M., Setoyama, C. and Shimada, K. (1986) Isolation and characterization of the cDNAs corresponding to mRNAs abundant in undifferentiated mouse embryonal carcinoma stem cells, but not in differentiated mouse parietal endoderm cells. J. Biochem. 100: 1185-1192.

Imperiale, M.J., Kao, H., Feldman, L.T., Nevins, J.R. and Strickland, S. (1984) Common control of the heat shock gene and early adenovirus genes: evidence for a cellular E1a-like activity. Mol. Cell. Biol. 4: 867-874.

Ingham, P.W. (1988) The molecular genetics of embryonic pattern formation in *Drosophila*. Nature 335: 25-34.

Jacobovits, A., Schwab, M., Bishop, M. and Martin, G.R. (1985) Expression of N-Myc in teratocarcinoma stem cells and mouse embryo. Nature 318: 188-190.

Jacobovits, A., Shackleford, G.M., Varmus, H.E. and Martin, G.R. (1986) Two proto-oncogenes implicated in mammary carcingenesis, int-1 and int-2 are independently regulated during mouse development. Proc. Natl. Acad. Sci. USA 83: 7806-7810.

Jaenisch, R. and Berns, A. (1987) In: *Concepts in mammalian embryogenesis*, M. Sherman, ed. (Cambridge, Massachusetts: The MIT Press), pp. 131-162.

Jaenisch, R., Harbers, S., Schnicke, A., Lohler, J., Chumokov, I., Jahner, D., Grofkopp, D. and Hoffman, E. (1983) Germline integration of Moloney murine leukemia virus at the Mov 13 locus leads to recessive lethal mutation and early embryonic death. Cell 32: 209-216.

Jaenisch, R., Jahner, D., Nobis, P., Simon, I., Lohler, J., Harbers, K. and Grotkopp, D. (1981) Chromosomal position and activation of retroviral genomes inserted into the germ line of mice. Cell 24: 519-529.

Jaynes, J.B. and O'Farrell, P.H. (1988) Activation and repression of transcription by homeodomain-containing proteins that bind a common site. Nature 336: 744-749.

Jenkins, N., Copeland, N., Taylor, B. and Lee, B. (1981) Dilute (d) coat colour mutation of DBA/2J mice is associated with the site of integration of an ecotropic MuLV genome. Nature 293: 370-374.

Jenkins, N., Copeland, N., Taylor, B. and Lee, B. (1982) Organization, distribution and stability of endogenous ecotropic murine luekemia virus DNA in chromosomes of *Mus musculus*. J. Virol. 43: 26-36.

Jetten, A.M. and Jetten, M.E.R. (1979) Possible role of retinoic acid binding protien in retinoid stimulation of embryonal carcinoma cell differentiation. Nature 278: 180-182.

Johnson, D.R. (1974) Hairpin tail: a case of post-reductional gene activity in the mouse eggs? Genetics 76: 795-805.

Johnson, M.H., Handyside, A.H. and Braude, P.M. (1977) In: *Development of mammals 2*, M.H. Johnson, ed. (Amsterdam: Elsevier/North-Holland), pp. 67-97.

Jones-Villeneuve, E.M.V., Rudnicki, M.A., Harris, J.F. and McBurney, M.W. (1983) Retinoic acid-induced neural differentiation of embryonal carcinoma cells. Mol. Cell. Biol. 3: 2271-2279.

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

Kleinsmith, L.J. and Pierce Jr, G.B. (1964) Multipotentiality of single embryonal carcinoma cells. Cancer Res. 24: 1544-1552.

Knowland, J. and Graham, C. (1972) RNA synthesis at the two-cell stage of mouse development. J. Embryol. Exp. Morphol. 27: 167-176.

Knudson Jr, A.G. (1986) Genetics of human cancer. Ann. Rev. Genet. 20: 231-251.

Ko, H.-S., Fast, P., McBride, W. and Staudf, L.M. (1988) A human protein specific for the immunoglobulin octomer DNA motif contains a functional homeobox domain. Cell. 55: 135-144.

Konieczny, S.F., Baldwin, A.S. and Emerson, C.P. Jr. (1986) In: *Molecular and Cellular Biology*, new series, vol 29, C. Emerson, D.A. Fischman, B. Nadal-Ginard and M.A.Q. Siddiqui, eds. (New York: Alan R. Liss), pp. 21-34.

Krolewski, J.J. and Rush, M.G. (1984) Some extrachromosomal circular DNAs containing the Alu family of dispersed repetitive sequences may be reverse transcripts. J. Mol. Biol. 174: 31-40.

Kubli, E. (1986) Molecular mechanisms of suppression in *Drosophila*. Trends. Genet. 2: 204-209.

Kuff, E.L., Feenstra, A., Lueders, K., Smith, L., Hawley, R., Hozumi, N. and Shulman, M. (1983) Intracisternal A-particle genes as movable elements in the mouse genome. Proc. Natl. Acad. Sci USA 80: 1992-1996.

Kuff, E.L., Lueders, K.K., Ozer, H.L. and Wivel, W.A. (1972) Some structural and antigenic properties of the intracisternal A-particles occuring in mouse tumors. Proc. Natl. Acad. Sci. USA 69: 218-222.

Kuziora, M.A. and McGinnis, W. (1988) Autoregulation of a *Drosophila* homeotic selector gene. Cell 55: 477-485.

Lachman, H.M. and Skoultchi, A.I. (1984) Expression of c-Myc changes during differentiation of mouse erythroleukaemia cells. Nature 310: 592-594.

Lacy, E.S., Roberts, E.P., Evans, D., Burtenshaw, D. and Constantini, F.D. (1983) A foreign β-globin gene in transgenic mice: Integration at abnormal chromosomal positions and expression in inappropriate tissues. Cell 34: 343-358.

Laimins, L.A., Gruss, P., Pozzatti, R. and Khoury, G. (1984) Characterization of enhancer elements in the long terminal repeat of Moloney murine sarcoma virus. J. Virol. 49: 183-189.

Lassar, A.B., Paterson, B.M. and Weintraub, H. (1986) Transfection of a DNA locus that mediates the conversion of 10T½ fibroblasts to myoblasts. Cell 47: 649-656.

Laughon, A. and Scott, M.P. (1984) Sequence of a *Drosophila* segmentation gene: Protein homology with DNA-binding proteins. Nature 310: 25-31.

Lee, S., Gilula, N.B. and Warner, A.E. (1987) Gap junctional communication and compaction during preimplantation stages of mouse development. Cell 51: 851-860.

Levine, M. and Hoey, T. (1988) Homeobox proteins as sequence-specific transcription factors. Cell 55: 537-540.

Levine, R.A., LaRosa, G.J. and Gudas, L.J. (1984) Isolation of cDNA clones for genes exhibiting reduced expression after differentiation of murine teratocarcinoma stem cells. Mol. Cell. Biol. 4: 2142-2150.

Levy, I.L., Stull, G.B. and Brinster, R.L. (1978) Poly(A) and synthesis of polyadenylated RNA in the preimplantation mouse embryo. Dev. Biol. 64: 140-148.

Littlefield, J.W. and Felix, J.S. (1982) Rescue of terminally differentiated teratocarcinoma cells by fusing to undifferentiated parental cells. Somatic Cell Genet. 8: 743-757.

Lueders, K.K. and Kuff, E.L. (1980) Intracisternal A-particle genes: Identification in the genome of *Mus musculus* and comparison of multiple isoaltes froma mouse gene library. Proc. Natl. Acad. Sci. USA 77: 3751-3575.

Mackay, T.F.C. (1985) Transposable element-induced response to artificial selection in *Drosophila melanogaster*. Genetics 111: 351-374.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular cloning: A laboratory manual*. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory).

Martin, G.R. (1975) Teratocarcinomas as a model system for the study of embryogenesis and neoplasia. Cell 5: 229-253.

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

Martin, G.R., Smith, S. and Epstein, C.J. (1978) Protein synthesis patterns in teratocarcinoma stem cells and mouse embryos at early stages of development. Dev. Biol. 66: 8-16.

Martin, G.R., Willey, L.M. and Damjavno, I. (1977) The development of cystic embryoid bodies *in vitro* from clonal teratocarcinoma stem cells. Dev. Biol. 61: 230-244.

Martin, G.R. and Evans, M.J. (1975) Differentiation of clonal lines of teratocarcinoma cells: Formation of embryoid bodies *in vitro*. Proc. Natl. acad. Sci. USA 72: 1441-1445.

McBurney, M.W., Jones-Villeneuve, E.M.V., Edwards, M.K.S. and Anderson, P.J. (1982) Control of muscle and neuronal differentiation in a cultured embryonal carcinoma cell line. Nature 299: 165-167.

McBurney, M.W. and Rogers, B.J. (1982) Isolation of male embryonal carcinoma cells and their chromosome replication patterns. Dev. Biol. 89: 503-508.

McClintock, B. (1951) Controlling elements and the gene. Cold Spring Harbor Symp. Quant. Biol. 16: 13-47.

McClintock, B. (1952) Chromosome organization and genic expression. Cold Spring Harbor Symp. Quant. Biol. 16: 13-18.

McClintock, B. (1956) Intracellular systems controlling gene action and mutation. Brookhaven Symp. Biol. 8: 58-62.

McGinnis, W., Garber, R.L., Wirz, J., Kuroiwa, A. and Gehring, W.J. (1984) A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans. Cell 37: 403-408.

McGrath, J. and Solter, D. (1986) Nucleocytoplasmic interactions in the mouse embryo. J. Embryol. Exp. Morph. 97 (Suppl): 277-289.

McGrath, J. and Solter, D. (1984a) Completion of mouse embryogenesis requires both the maternal and paternal genomes. Cell 37: 179-183.

McGrath, J. and Solter, D. (1984b) Maternal Thp lethality in the mouse is a nuclear, not cytoplasmic, defect. Nature 308: 550-551.

McLaren, A. (1979) In: *Maternal effects in development*, D.R. Newth and M. Balls, eds. (England: Cambridge University Press), pp. 287-320.

Mintz, B. and Illmensee, K. (1975) Normal genetically mosaic mice produced from malignant teratocarcinoma cells. Proc. Natl. Acad. Sci. USA 72: 3585-3589.

Muller, R., Slamon, D.J., Tremblay, J.M., Cline, M.J. and Verma, I. (1982) Differential expression of cellular oncogenes during pre- and postnatal development of the mouse. Nature 299: 640-643.

Muller, R., Slamon, D.J., Adamson, E.D., Tremblay, J.M., Muller, D., Cline, M.J. and Verma, I. (1983) Transcription of c-onc genes, c-ras Ki and c-fms during mouse development. Mol. Cell. Biol. 3: 1062-1069.

Muller, R., Verma, I. and Adamson, E.D. (1983) Expression of c-oncogenes: C-fos transcripts accumulates to high levels during development of mouse placenta, yolk sac and amnion. EMBO J. 2: 679-684.

Murphy, R., Brickell, P.M., Latchman, D.S., Willison, K. and Rigby, P.W.J. (1983) Transcripts regulated during normal embryonic development and oncogenic transformation share a repetitive element. Cell 35: 865-871.

Natt, E. and Scherer, G. (1986) EMBL 12, a new lambda replacement vector with sites for *Sal*I, *Xba*I, *Bam*HI, *Sst*I, and *Eco*RI. Nucleic Acids Res. 14: 7128.

Nusse, R. and Varmus, H.E. (1982) Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. Cell 31: 99-109.

O'Hare, K., Levis, R. and Rubin, G.M. (1983) Transcription of the white locus in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 80: 6917-6921.

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.

Ono, M., Cole, M.D., White, A.T. and Huang, R.C.C. (1980) Sequence organization of cloned intracisternal A-particle genes. Cell 21: 465-473.

Ouweneel, W.J. (1976) Developmental genetics of homeosis. Adv. Genet. 18: 179-248.

Palmiter, R.D. and Brinster, R.L. (1986) Germ line transformation of mice. Ann. Rev. Genet. 20: 465-499.

Parkhurst, S.M., Harrison, D.A., Remington, M.P., Spano, C., Kelly, R., Coyne, R.S. and Corces, V.G. (1988) The *Drosophila* su(Hw) gene, which controls the phenotypic effect of the gypsy transposable element, encodes a putative DNA-binding protein. Genes Dev. 2: 1205-1215.

Parkhurst, S.M. and Corces, V.G. (1986) Mutation at the suppressor of *forked* locus increases the accumulation of gypsy-encoded transcripts in *Drosophila melanogaster*. Mol. Cell. Biol. 6: 2271-2274.

Paulson, K.E., Deka, N., Schmid, C.W., Misra, R., Schindler, C.W., Rush, M.G., Kadyk, L. and Leinwand, L. (1985) A transposon-like element in human DNA. Nature 316: 359-361.

Pinnay, D.F., Pearson-White, S.F. Konieczny, S.F., Latham, K.E. and Peteron, C.P. Jr. (1988) Myogenic lineage determination and differentiation: Evidence for a regulatory gene pathway. Cell 53: 781-793.

Prochownick, E.V. and Kukowska, J. (1986) Deregulated expression of c-Myc by murine erythroleukaemia cells prevent differentiation. Nature 322: 848-850.

Ptashne, M. (1988) How eukaryotic transcriptional activators work. Nature 335: 683-689.

Reik, W., Collick, A., Norris, M.L., Barton, S.C. and Surani, M.A. (1987) Genomic imprinting determines methylation of parental alleles in transgenic mice. Nature 328: 248-251.

Riabowol, K.T., Fink, J.S., Bilman, M.Z., Walsh, D.A., Goodman, R.H. and Feramisco, J.R. (1988) The catalytic subunit of cAMP-dependent protein kinase induces expression of genes containing cAMP-responsive enhancer elements. Nature 336: 83-86.

Rigby, P.W.J., Diekmann, M., Rhodes, S. and Berg, P. (1977) Labeling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. 13: 237-251.

Rijsewijk, F., Schuermann, M., Wagenaar, E., Parren, p., Weigel, D. and Nusse, R. (1987) The *Drosophila* homolog of the mouse mammary oncogene int-1 is identical to the segment polarity gene wingless. Cell 50: 649-657.

Robl, J.M., Gilligan, B., Critser, E.S. and First, M.L. (1986) Nuclear transplantation in mouse embryos: Assessment of recipient cell stage. Biol. Reprod. 34: 733-739.

Rodnicki, M.A. and McBurney, M.W. (1987) In: *Teratocarcinomas and embryonic stem cells: A practical approach*, E.J. Robertson, ed. (Oxford, England: IRL Press), pp. 19-49.

Rogel, A., Popliker, M., Webb, C.G. and Oren, M. (1985) p53 cellular tumor antigens: Analysis of mRNA levels in normal adult tissues, embryos and tumors. Mol. Cell. Biol. 5: 2851-2855.

Rosenstraus, M.J. and Levine, A.J. (1979) Alternations in the developmental potential of embryonal carcinoma cells in mixed aggregates of nullipotent and pluripotent cells. Cell 17: 337-346.

Rubin, M.R., Toth, L.E., Patel, M.D., D'Eustachio, P. and Nguyen-Huu, M.C. (1986) A mouse homeo box gene is expressed in spermatocytes and embryos. Science 233: 663-667.

Rubin, M.R., Hazelrigg, T., Karess, R.E., Laski, F.A., Laverty, T., Levis, R., Rio, D.E., Spencer, F.A. and Zuker, C.S. (1985) Germline specificity of P-element transposition and some novel patterns of expression of transduced copies of the white gene. Cold Spring Harbor Symp. Quant. Biol. 50: 329-335.

Rubin, G.M., Kidwell, M.G. and Bingham, P.M. (1982) The molecular basis of P-M hybrid dysgenesis: The nature of induced mutations. Cell 29: 987-994.

Sapienza, C., Peterson, A.C., Rossant, J. and Balling, R. (1987) Degree of methylation of transgenes is dependent on gamete of origin. Nature 328: 251-254.

Schwarz-Sommer, Z., Gierl, A., Cuypers, H., Peterson, P.A. and Saedler, H. (1985) Plant transposable elements generate the DNA sequence diversity needed in evolution. EMBO J 4: 591-597.

Scott, M.P. and Weiner, A.J. (1984) Structural relationships among genes that control development: Sequence homology bewteen the antennapedia, ultrabithorax and fushi tarazu loci of *Drosophila*. Cell 35: 763-776.

Searle, A.G. and Beechy, C.V. (1978) Complementation studies with mouse translocations. Cytogenet. Cell. Genet. 20: 282-303.

Serfling, E. (1982) in: *Cell differentiation: Molecular basis and problems*, L. Nover, M. Luckner and B. Parthier, eds. (Berlin, Heidelberg, New York: Springer-Verlag), pp. 349-374.

Shapiro, J.A. (1979) Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. Proc. Natl. Acad. Sci. USA 76: 1933-1937.

Sherman, M.I. and Miller, R.I. (1978) F9 embryonal carcinoma cells can differentiate into endoderm-like cells. Dev. Biol. 63: 27-34.

Slamon, D.J. and Cline, M.J. (1984) Expression of cellular oncogenes during embryonic and fetal development of the mouse. Proc. Natl. Acad. Sci USA 81: 7141-7145.

Snow, M.H.L. and Monk, M. (1983) In: *Current problems in germ cell differentiation*, A. McLaren and C.C. Wylie, eds. (Cambridge: Cambridge University Press), pp. 115-135.

Snow, M.H.L. (1977) In: *Development in mammals*, 3, M. Johnson ed. (Amsterdam: Elsevier), pp. 337-362.

Solter, D. (1986) In: *Molecular genetics in developmental neurobiology*, Y. Tsukada, ed. (Japan Scientific Societies Press), pp. 91-100.

Solter, D. (1987) Inertia of the embryonic genome in mammals. Trends Genet. 3: 23-27.

Solter, D., Shevinsky, L., Knowles, B.B. and Strickland, S. (1979) The induction of antigenic changes in a teratocarcinoma stem cell line (F9) by retinoic acid. Dev. Biol. 70: 515-521.

Sorger, P.K. and Pelham, H.R.B. (1988) Yeast heat shock factor is an essential DNA-binding protein that exhibits temperature-dependent phosphorylation. Cell 53: 855-864.

Sorger, P.K., Lewis, M.J. and Pelham, H.R.B. (1987) Heat shock factor is regulated differently in yeast and HeLa cells. Nature 329: 81-84.

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

Southern P.J. and Berg, P. (1982) Transformation of mammalian cells to antobiotic resistance with a bacterial gene under control of the SV40 early promoter. J. Mol. Appl. Genet. 1: 327-341.

Spradling, A.C. and Rubin, G.M. (1982a) Transposition of cloned P elements into *Drosophila* germ line chromosomes. Science 218: 341-347.

Spradling, A.C. and Rubin, G.M. (1982b) Genetic transformation of *Drosophila* with transposable element vectors. Science 218: 348-353.

Spradling, A.C. and Mahowald, A.P. (1980) Amplification of genes for chorion proteins during oogenesis in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 77: 1096-1100.

Stacey, A.J. and Evans, M.J. (1984) A gene sequence expressed only in undifferentiated EC, EK cells and testes. EMBO J. 3: 2279-2285.

Stavenhagen, J.B. and Robins, D.M. (1988) An ancient provirus has imposed andogen regulation on the adjacent mouse sex-limited protein gene. Cell 55: 247-254.

Stevens, L.C. (1975) In: *Teratomas and differentiation*, M.I. Sherman and D. Solter, eds. (New York: Academic Press), pp. 17-33.

Stevens, L.C. (1982) In: *Genetic approaches to developmental neurobiology*, Y. Tsukada, ed. (Tokyo: University of Tokyo Press), pp. 87-93.

Steward, C., Stuhlmann, H., Jahner, D. and Jaenisch, R. (1982) De novo methylation, expression and infectivity to retroviral genomes introduced into embryonal carcinoma cells. Proc. Natl. Acad. Sci. USA 79: 4098-4102.

Steward, R. (1987) Dorsal, an embryonic polarity gene in *Drosophila*, is homologous to the vertebrate proto-oncogene, c-rel. Science 238: 692-694.

Stoy, J.P., Fenner, S., Greenoak, G.E., Moran, C. and Coffin, J.M. (1988) Role of endogenous retroviruses as mutagens: The *hairless* mutation of mice. Cell 54: 383-391.

Strickland, S. and Mahdavi, V. (1978) The induction of differentiation in teratocarcinoma stem cells by retinoic acid. Cell 15: 393-403.

Surani, M.A.H. (1985) In: *Genetic manipulation of the early mammalian embryo*, G. Constantini and R. Jaenisch, eds. (Cold Spring Harbor, New York: Cold Spring Harbor Laboratory), pp. 43-55.

Surani, M.A.H., Barton, S.C. and Noris, M.L. (1986) Nuclear transfer in the mouse: Heritable differences between parental genomes after activation of the embryonic genome. Cell 45: 127-136.

Swain, J.L., Steward, T.A. and Leder, P. (1987) Parental legacy determines methylation and expression of an autosomal transgene: A molecular mechanism for Parental imprinting. Cell 50: 719-727.

Swift, F.V., Bhat, K., Younghusband, H.B. and Hamada, H. (1987) Characterization of a cell-type specific enhancer found in human papilloma virus type 18 genome. EMBO J. 5: 1339-1344.

Taketo, M. and Tanaka, M. (1987) A cellular enhancer of retrovirus gene expression in embryonal carcinoma cells. Proc. Natl. Acad. Sci. USA 84: 3748-3752.

Talbot, D., Collis, P., Antoniou, M., Vidal, M., Grosved, F. and Greaves, D.R. (1989) A dominant control region from the human β-globin locus conferring integration site-independent gene expression. Nature 338: 352-355.

Taylor, S.M. and Jones, P.A. (1979) Multiple new phenotypes induced in 10T½ and 3T3 cells treated with 5-azacytidine. Cell 17: 771-779.

Temin, H.M. (1980) Origin of retroviruses from cellular moveable genetic elements. Cell 21: 599-600.

Ulrich, R., Garber, C., Komitowski, D., Muller, R. and Wagner, E.F. (1987) Deregulated c-fos expression interferes with normal bone development in transgenic mice. Nature 325: 412-416.

Umesono, K., Giguere, V., Glass, C.K., Rosenfeld, M.G. and Evans, R.M. (1988) Retinoic acid and thyroid hormone induce gene expression through a common responsive element. Nature 336: 262-264.

Varmus, H.E. (1982) Form and function of retroviral proviruses. Science 216: 812-820.

Wasserman, P.M. (1987) Early events in mammalian fertilization. Annu. Rev. Cell Biol. 3: 109-142.

Westaway, D., Payne, G. and Varmus, H.E. (1984) Proviral deletions and oncogene base-substitutions in insertionally mutagenized c-Myc alleles may contribute to the progression of avian bursal tumors. Proc. Natl. Acad. Sci. USA 81: 843-847.

Westin, E.H., Wong-Staal, F., Gelmann, E.P., Dalla Favera, R., Pappas, T.K., Lautenberger, S.A., Eval, A., Reddy, E.P., Tronick, S.R., Aaronson, S.A. and Gallo, R.C. (1982) Expression of cellular homologues of retroviral oncgenes in human hematopoietic cells. Proc. Natl. Acad. Sci. USA 79: 2490-2494.

Willadsen, S.M. (1986) Nuclear transplantation in sheep embryos. Nature 320: 63-65.

Wolgemuth, D.J., Behringer, R.R., Mostoller, M.P., Brinster, R.L. and Palmiter, R.D. (1989) Transgenic mice overexpressing the mouse homeobox-containing gene Hox-1.4 exhibit abnormal gut development. Nature 337: 464-467.

Wolgemuth, D.J., Viviano, C.M., Ginsberg, E.G., Frohman, M.A., Joyner, A.L. and Martin, G.R. (1987) Differential expression of the mouse homeobox-containing gene Hox 1.4 during male germ cell differentiation and embryonic development. Proc. Natl. Acad. Sci. USA 84: 5813-5817.

Wolgemuth, D.J., Engelmyer, D., Duggal, R.N., Gizang-Ginsberg, E., Mutter, G.L., Ponzetto, C., Viviano, C. and Zakevi, Z.F. (1986) Isolation of a mouse cDNA coding for a developmentally regulated, testis-specific transcript containing homeobox homology. EMBO J 5: 1229-1235.

Wray, L.K. and Jaeckle, H.K. (1983) Protein synthesis by rat hepatoma X mouse teratocarcinoma somatic cell hybrids. Somatic Cell Genet. 9: 111-119.

Wudl, L. and Chapman, V. (1976) The expression of β-glucuronidase during preimplantation development of mouse embryos. Dev. Biol. 48: 104-109.